

Seventh Annual

Carl B. Menges Colloquium

**“War and the West:
Strategic Issues;
Past, Present, and Future”**

April 3-5, 2014



*The Alexander Hamilton Institute
for the Study of Western Civilization*

at The Alexander Hamilton Inn
21 W. Park Row
Clinton, NY 13323
USA

web: www.theahi.org
email: ahi@theahi.org

Seventh Annual Carl B. Menges Colloquium, April 3-5, 2014

**“War and the West: Strategic Issues Past, Present, and Future”
Turning Stone Resort**

Location: Turning Stone Resort, 5218 Patrick Road, Verona, NY, 13478
<http://www.turningstone.com/>

Tentative Schedule:

Thursday, 3 April 2014:

3:00 - 5:00 p.m.: Hospitality Suite (3rd Floor VIP Room)

4:00 – 5:15 p.m.: Cocktail Reception (Cypress Room)

5:15 - 5:30 p.m.: Welcome and Introductory Remarks

Richard Erlanger, President, AHI

5:30-6:30 : Past, Present, and Future of the AHI,

Robert Paquette, Co-Founder and Charter Fellow, AHI

6:30 – 7:30 p.m.: Banquet (Cypress Room)

7:30 – 8:30 p.m.: Keynote Address

Seventh Annual Carl B. Menges Lecture

Keynote Address, Dr. Michael D. Swaine, Senior Associate, Carnegie Endowment for International Peace

9:30 p.m.: Hospitality Suite (3rd Floor VIP Room)

Friday, 4, April 2014:

7:30 – 8: 45 a.m.: Continental Breakfast (Cypress Room)

8:45-10:30 a.m.--Session I: “War before Western Civilization”

8:45 – 10:00 a.m.: Panel Discussion (Cypress Room)

10:00- 10:30 a.m.: Questions from audience

10:30 – 11:00 a.m. Break

11:00 a.m.-12:45 p.m. Session II: “The West and the Near East – Past and Present”

11:00 – 12:15 p.m.: Panel Discussion (Cypress Room)

12:15 - 12:45 p.m.: Questions from audience

12:45 – 1:45 p.m.: Lunch (Cypress Foyer)

1:45 – 3:30 p.m.: Free time

2:00-4:00 Meeting of AHI Board of Directors (Cedar Room)

3:30--4:00 p.m.: Refreshments (Cypress Room)

4:00 -5:45 p.m.--Session III: “The West and the Far East – Past and Present”

4:00 – 5:15 p.m.: Panel Discussion (Cypress Room)

5:15 - 5:45 p.m.: Questions from audience



The Alexander Hamilton Institute
for the Study of Western Civilization

6:10 p.m.: Conferees gather at Front Desk for bus transport to AHI
6:30-7:00 p.m.: Reception at AHI
7:00 – 9:30 p.m.: Directors' Dinner **at the headquarters of the AHI** in Clinton, NY
9:30 p.m.: Return by bus to Turning Stone
9:30p.m. – midnight: Hospitality Suite (3rd Floor VIP Room)

Saturday, 5 April 2014:

7:30 – 8: 30 a.m.: Continental Breakfast (Cypress Room)
8:45-10:30 a.m. Session IV: “The West, Women, and War”
8:45 – 10:00 a.m.: Panel Discussion (Cypress Room)
10:00- 10:30 a.m.: Questions from audience
10:30 – 11:00 a.m.: Break (Refreshments)
11:00-12:45 p.m. Session V: “Technology, Ethics, and War”
11:00 – 12:15 p.m.: Panel Discussion (Cypress Room)
12:15- 12:45 p.m.: Questions from participating students
12:45 – 1:45 p.m.: Lunch (Event Center Atrium)
1:45 – 3:30 p.m.: free time
3:30-:4:00 p.m. Refreshments (Cypress Room)
4:00 p.m.-5:45 p.m. Session VI: “War after Western Civilization”
3:00 – 4:15 p.m.: Panel Discussion (Cypress Room)
4:15 - 4:45 p.m.: Questions from audience
6:15 p.m. – Conference concludes with Barbecue Buffet
7:30 – midnight: Hospitality Suite (3rd Floor VIP Room)

Sunday, 6 April 2014:

Breakfast and departure at your convenience.

List of Participants

Keynote:

Michael D. Swaine, Senior Associate at the Carnegie Endowment for International Peace

Conferees:

1. Alfred Kelly, Edgar B. Graves Professor of History, Hamilton College
2. David Frisk, Theodore J. Eismeier Fellow, Alexander Hamilton Institute
3. Lauren Hall, Assistant Professor of Political Science, Rochester Institute of Technology



The Alexander Hamilton Institute
for the Study of Western Civilization

4. Paul Gottfried Professor of Humanities and Raffensperger Chair Emeritus, Elizabethtown College
5. Mackubin Owens, Professor, National Security Affairs, U. S. Naval War College
6. Edward Barrett, Director of Strategy and Research Stockdale Center for Ethical Leadership, U.S. Naval Academy
7. Matthew Zeller, Adjunct Fellow, American Security Project
8. Eric Hannis, Senior Fellow in Defense Studies, American Foreign Policy Council
9. Chris Hill (discussion leader), Resident Fellow, Alexander Hamilton Institute
10. Alex Crowther, Adjunct Research Professor, Strategic Studies Institute
11. Doug Macdonald, Associate Professor of Political Science, Colgate University
12. John Kelsay, Distinguished Research Professor, Florida State University
13. Miri Eisin, Security Studies, Haifa University
14. Mary Habeck, Associate Professor of Strategic Studies, Johns Hopkins University
15. Joseph Capizzi, Associate Professor of Moral Theology, Catholic University

Session 1: War Before Western Civilization

War Before Civilization: The Myth of the Peaceful Savage. Lawrence H. Keeley, Oxford University Press, New York, 1996.

“What Our Primate Relatives Say About War.” Dominic Johnson, Bradley Thayer, *The National Interest*, January 29, 2013

“The Elements of Augustine’s Just War Theory.” John Langan, *The Journal of Religious Ethics*, Vol 12, No.1 (Spring 1984).

Violence and Social Orders. Douglass C. North, John Joseph Wallis, Barry R. Weingast, Cambridge University Press, 2009.

Session 2: The West and the Near East – Past and Present

“Politics, Ethics and War in Premodern Islam.” John Kelsay, *Arguing the Just War in Islam*, Harvard University Press, 2009.

“Why Is There so Much Conflict in the Middle East?” Mirjam E. Sørli, Nils Petter Gleditsch and Håvard Strand, *The Journal of Conflict Resolution*, Vol. 49, No. 1 (Feb., 2005).

“Primitivization of War and Prospects for Peace.” Mehar Omar Khan, *Small Wars Journal*, September 14, 2012.

“Tehran's Take: Understanding Iran's U.S. Policy.” Mohsen M. Milani, *Foreign Affairs*, Vol. 88, No. 4 (July/August 2009).

“Just War and Extraterritoriality: The Popular Geopolitics of the United States' War on Iraq as Reflected in Newspapers of the Arab World.” Ghazi-Walid Falah, Colin Flint and Virginie Mamadouh, *Annals of the Association of American Geographers* Vol. 96, No. 1 (Mar., 2006)

Session 3: The West and the Far East – Past and Present

“Just War: Chinese and Western Perspectives.” Edmund Ryden, S.J., http://www.crvp.org/book/Series03/III-17/chapter_xiii.htm

“China’s Assertive Behavior—Part One: On “Core Interests.” Michael Swaine *China Leadership Monitor* » 2011 no. 34 » foreign policy.

“Regime Insecurity and International Cooperation: Explaining China's Compromises in Territorial Disputes.” M. Taylor Fravel, *International Security*, Vol. 30, No. 2 (Fall, 2005).

“A Salutation To Arms: Asia’s Military Buildup, Its Reasons, and Its Implications.” Felix Chang, *Foreign Policy Research Institute E-Notes*, September 2013.

“What We’ve Learned from China’s Air Defense Zone (so Far)” Robert Haddick, Commentary and Analysis, warontherocks.com, December 9, 2013.

Session 4 – The West, Women and War

“Integrating Women into the Infantry,” Adam Nojack, *Military Review*, November-December 2002.

“Get Over It! We Are Not All Created Equal.” Katie Petronio, *Marine Corps Gazette*, July 2012.

“Let Women Fight.” Megan MacKenzie, *Foreign Affairs*, November-December 2012.

“Why Can’t Anything Be Done? Measuring Physical Readiness of Women for Military Occupations.” William Granger. Paper Presented at the 2011 International Biennial Conference of the Inter-University Seminar on Armed Forces and Society.

Session 5 - Technology, Ethics and War

“Enhanced Warfighters: Risk, Ethics and Policy.” Patrick Lin, Maxwell Mehlman, Kieth Abney. Report prepared for The Greenwall Foundation, January 2013.

“Compromised by Design? Securing the Defense Electronics Supply Chain.” John Villasenor. Brookings Center for Technology Innovation, Center for 21st Century Security and Intelligence, November 2013.

“Nuclear Blindness: AN Overview of the Biological Weapons Programs of the Former Soviet Union and Iraq.” Christopher Davis, *Emerging Infectious Diseases*, Vol. 5, No. 4, July-August 1999.

“War and Technology.” Alex Roland, *Foreign Policy Research Institute, Footnotes*, Vol. 14, No. 2. February 2009.

“Stuxnet’s Secret Twin.” Ralph Langer, *Foreign Policy*, November 21, 2013.

Session 6 – War After Western Civilization

“The Evolution of Law of War.” Sheng Hongsheng, *Chinese Journal of International Politics*, Vol. 1, 2006.

“The Wars of the 21st Century.” Herfried Munkler, *International Review of the Red Cross*, Vol. 85, No. 849. March, 2003.

“Conflict in the 21st Century: The Rise of Hybrid Wars.” Frank Hoffman, Potomac Institute for Policy Studies, Arlington, VA, December 2007.

Session 1: War Before Western Civilization

should be able to provide the documentation. The evaluation of these ideas (and, of course, any ideas contrary to them) requires careful surveillance of both ethnographic and archaeological data, with special attention to questions of how recent tribal and ancient prehistoric warfare was actually conducted and what the direct results of such conflicts were. Since implicit in any discussion of primitive warfare is a contrast with the corresponding forms of civilized conflict, it is also vital to make direct comparisons between the two in equivalent terms. Only then is it possible to achieve a realistic view of all warfare and to determine whether anthropology has anything to offer us in our attempts to understand and eventually eliminate the awful scourge of war. The purpose of this book is to provide just such a survey and evaluation.

TWO

The Dogs of War

The Prevalence and Importance of War

As we have seen, many recent popular and academic views of precivilized warfare agree that it was a trivial and insubstantial activity. Proponents of primitive war and the pacified past claim or imply that peaceful societies were common, fighting was infrequent, and active participation in combat was limited among nonstate peoples until they either evolved into or made contact with states and civilizations.

If these views are correct, they should be supported by broad surveys of ethnographic and archaeological evidence. Ethnographic data should indicate that nonstate societies were commonly pacifistic, resorted to combat much less frequently than did ancient or modern states, and mobilized little of their potential manpower for the warfare they did conduct. In the more thoroughly studied regions, archaeology should recover very little evidence of violent conflicts before the development of indigenous states or the intrusion of foreign states. As we shall see, on the contrary, the available evidence shows that peaceful societies have been very

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

rare, that warfare was extremely frequent in nonstate societies, and that tribal societies often mobilized for combat very high percentages of their total manpower.

LEVELS OF SOCIAL COMPLEXITY

Before proceeding with any ethnographic survey, we must review some terms that are used by anthropologists in roughly classifying the size and complexity of societies. These terms include *bands*, *tribes*, *chiefdoms*, *states*, and *civilized* or *urban states*. They loosely describe the population size and the economic and political complexity of various societies.

Bands are small, politically autonomous groups of twenty to fifty people with an informal headman. They usually consist of a few related extended families who reside or move together. Typically, bands are hunter-gatherers or foragers. Several such *micro-bands* usually congregate for a few weeks each year into a *macro-band* of several hundred people for ceremonies, festivities, courting and marriage arrangements, and the exchange of goods. Such macro-bands usually speak a distinct dialect and are sometimes referred to as *dialect tribes*. The classic examples of societies with band organization are the Eskimo of the central Arctic, the Paiutes of the American Great Basin, and the Aborigines of central Australia.

The term *tribe* covers a multitude of social and political organizations. Tribes generally incorporate a few thousand people into a single social organization via pan-tribal associations. These associations are usually kin groups that trace descent to a common hypothetical or mythological ancestor. But nonkin associations, such as *age-grades* (groups of young men who were initiated together) and *sodalities* (voluntary nonkin associations such as dance societies, clubs, etc.), can also integrate a tribe. Tribes are collections of such associations or kin groups that unite for war. While tribal leaders may be called big men or chiefs, they are not formal full-time political officials, and they usually exercise influence rather than what we would call power. In most cases, there is no central political organization except informal councils of "elders" or local chiefs. Foraging, pastoral, and agricultural economies are all found among tribes. Tribes are so various in their features that it is difficult to list classic cases, but the Indian tribes of the Plains, the southwestern Pueblos, and the Masai of East Africa are familiar examples.

Chiefdoms are organizations that unite many thousands or tens of thousands of people under formal, full-time political leadership. The populace of a chiefdom is usually divided into hereditary *ranks* or incipient social classes, often consisting of no more than a small chiefly or noble class and a large body of

commoners. Both the means of production and economic surpluses are concentrated under the control of the chief, who redistributes them. A central political structure integrates many local communities. This central body may consist of a council of chiefs, but in most cases a single head chief controls a hierarchy of lesser chiefs. Accession to chiefship is hereditary, permanent, and justified on religious or magical grounds. But a chief, unlike a king, does not have the power to coerce people into obedience physically; instead, he must rely on magical and economic powers to enforce his dictates. Some typical examples, ranging from weak to strong chiefdoms, include some Pacific Northwest Coast tribes, many Polynesian societies, early medieval Scottish clans, and some traditional petty kingdoms in central Africa.

States are also political organizations that incorporate many tens or hundreds of thousands of people from numerous communities into a single territorial unit. They have a central government empowered to collect taxes, draft labor for public works or war, decree laws, and physically enforce those laws. Essentially, states are class-stratified political units that maintain a "monopoly of deadly force"—a monopoly institutionalized as permanent police and military forces. *Civilized* states are simply those with cities and some form of record keeping (usually writing). Since few people in the world today are not citizens of some state, examples are unnecessary.

The term *primitive*, when used in its usual sense in anthropology, merely refers to a technological condition—that of using preindustrial or preliterate technology. In social terms, primitive refers to societies that are not urban or literate. Precisely such societies are the traditional subject matter of anthropology. But because the word has negative connotations in everyday speech, *primitive* has fallen out of favor. It has been erratically replaced by a number of inelegant neologisms such as *preliterate* or *nonliterate*, *prestate* or *nonstate*, *preindustrial* and *small-scale*. The term *tribal societies* usually encompasses bands, tribes, and weak chiefdoms but excludes strong chiefdoms and states. In the broadest sense, all these terms refer to societies that are simpler in technology and some aspects of social organization—and usually smaller in size—than societies that have produced historical records. Primarily for stylistic variety, all these terms are used interchangeably here.

IS WARFARE UNIVERSAL?

According to the most extreme views, war is an inherent feature of human existence, a constant curse of all social life, or (in the guise of real war) a perversion of human sociability created by the centralized political structures of states and civilizations. In fact, cross-cultural research on warfare has estab-

lished that although some societies that did not engage in war or did so extremely rarely, the overwhelming majority of known societies (90 to 95 percent) have been involved in this activity.

Three independent cross-cultural surveys of representative samples of recent tribal and state societies from around the world have tabulated data on armed conflict, all giving very consistent results. In one sample of fifty societies, only five were found to have engaged "infrequently or never" in any type of offensive or defensive warfare.¹ Four of these groups had recently been driven by warfare into isolated refuges, and this isolation protected them from further conflict. Such groups might more accurately be classified as defeated refugees than as pacifists. One California Indian tribe, the Monachi of the Sierra Nevada, apparently did occasionally go to war, but only very rarely. The results of this particular survey indicate that 90 percent of the cultures in the sample unequivocally engaged in warfare and that the remaining 10 percent were not total strangers to violent conflict.

In another larger cross-cultural study of politics and conflict, twelve of a sample of ninety societies (13 percent) were found to engage in warfare "rarely or never."² Six of these twelve were tribal or ethnic minorities that had long been subject to the peaceful administration of modern nation-states—for example, the Gonds of India and the Lapps of Scandinavia. Three were agricultural tribes living in geographically isolated circumstances, such as the Tikopia islanders of Polynesia (who were defeated refugees) and the Cayapa tribe of Ecuador.³ The final three were nomadic hunter-gatherers of the equatorial jungles and arctic tundra: the Mbuti Pygmies of Zaire, the Semang of Malaysia, and the Copper Eskimo of arctic Canada. Most of these peaceful societies were recently defeated refugees living in isolation, lived under a "king's peace" enforced by a modern state, or both. The real exceptions, representing only 5 percent of the sample, were some small bands of nomadic hunter-gatherers and a few isolated horticultural tribes.

In a study of western North American Indian tribes and bands, again only 13 percent of the 157 groups surveyed were recorded as "never or rarely" raiding or having been raided—meaning, in this case, *more than once a year*.⁴ Of these 21 relatively peaceable groups, 14 gave other evidence of having conducted or resisted occasional raids, presumably only once every few years. This leaves only 7 truly peaceful societies (4.5 percent of the sample) that apparently did not participate in any type of warfare or raiding. All these were very small nomadic bands residing in the driest, most isolated regions of the Columbia Plateau and the Great Basin.⁵ Again, we find the most peaceful groups living in areas with extremely low population densities, isolated by distance and hard country from other groups.

Even highly nomadic, geographically isolated hunter-gatherers with low

population densities are not universally peaceable. For example, many Australian Aboriginal foragers, including those living in deserts, were inveterate raiders.⁶ The seeming peacefulness of such small hunter-gatherer groups may therefore be more a consequence of the tiny size of their social units and the large scale implied by our normal definition of warfare than of any real pacifism on their part. Under circumstances where the sovereign social and political unit is a nuclear or slightly extended family band of from four to twenty-five people, even with a sex ratio unbalanced in favor of males, no more than a handful of adult males (the only potential "warriors") are available. When such a small group of men commits violence against another band or family, even if faced in open combat by all the men of the other group, this activity is not called war but is usually referred to as feuding, vendetta, or just murder.

Thus many small-band societies that are regarded by ethnologists as not engaging in warfare instead evidence very high homicide rates.⁷ For example, the Kung San (or Bushmen) of the Kalahari Desert are viewed as a very peaceful society; indeed, one popular ethnography on them was titled *The Harmless People*. However, their homicide rate from 1920 to 1955 was four times that of the United States and twenty to eighty times that of major industrial nations during the 1950s and 1960s. Before local establishment of the Bechuanaland/Botswana police, the Kung also conducted small-scale raids and prolonged feuds between bands and against Tswana herders intruding from the east. The Copper Eskimo, who appear as a peaceful society in the cross-cultural surveys just discussed, also experienced a high level of feuding and homicide before the Royal Canadian Mounted Police suppressed it. Moreover, in one Copper Eskimo camp of fifteen families first contacted early in this century, every adult male had been involved in a homicide. Other Eskimo of the high arctic who were organized into small bands also fit this pattern. Based on figures from different sources, the murder rate for the Netsilik Eskimo, *even after the Mounties had suppressed interband feuding*, exceeds that of the United States by four times and that of modern European states by some fifteen to forty times. At the other end of the New World, the isolated Yaghan "canoe nomads" of Tierra del Fuego, whose only sovereign political unit was the "biological family," had a murder rate in the late nineteenth century "10 times as high as that of the United States."⁸ Thus armed conflict between social units does not necessarily disappear at the lowest levels of social integration; often it is just terminologically disguised as feuding or homicide.

Both Richard Lee and Marvin Harris, defending the pacifistic nature of Kung and other simple societies compared with our own, decry the "semantic deception" that disguises the "true" homicide rates of modern states by ignoring the murders inflicted during wars.⁹ Let us undertake such a comparison for one simple society, the Gebusi of New Guinea. Calculations show that the

United States military would have had to kill nearly the *whole* population of South Vietnam during its nine-year involvement there, in addition to its internal homicide rate, to equal the homicide rate of the Gebusi.¹⁰ As their ethnographer Bruce Knauff notes, "Only the most extreme instances of modern mass slaughter would equal or surpass the Gebusi homicide rate over a period of several decades."¹¹ There is, then, an equal semantic deception involved in manufacturing peaceful societies out of violent ones by refusing to characterize as war their only possible form of intergroup violence, merely because of the small size of the contending social units.

If many of the "peaceful" hunter-gatherer bands did in reality engage in armed conflict, were any of them genuine pacifists? Perhaps the most striking case of peaceful hunters involves the Polar Eskimo of northwestern Greenland.¹² In the early nineteenth century, they consisted of a small band of some 200 people whose circumstances seemed ideally suited to a postapocalyptic science-fiction plot or perhaps a heartless social science experiment. Their icy isolation had been so complete for so long that they were unaware that any other people existed in the world until they were contacted in 1819 by a European explorer. This tiny society, whose members eked a precarious livelihood from a frozen desert, not surprisingly avoided all feuds and armed conflicts, although murder was not unknown.¹³ When other Eskimo from Canada and southwestern Greenland reached them after hearing of their existence from Europeans, relations with these strangers and with the Europeans they encountered were always quite amicable. The Polar Eskimo thus provide a counterexample to the recent theory that contact with Western civilization and its material goods inevitably turns peaceful tribesmen into Hobbesian berserkers.

There are a few other examples of peaceful hunter-gatherers.¹⁴ The Mbuti Pygmies and Semang of the tropical forests of central Africa and Malaysia seem to have completely eschewed any form of violent conflict and can legitimately be regarded as pacifistic. However, the Pygmy foragers were in fact politically subordinate to and economically dependent on the farmers who surrounded them (Chapter 9). Although they frequently engaged in nonlethal violence involving weapons, the last small "wild" band of Aborigines in the western Australian desert, the Mardudjara, never (at least while ethnographers were present) permitted such fighting to escalate into killing. Although they possessed shields and specialized fighting weapons, the Mardudjara had no words in their language for feuds or warfare. The Great Basin Shoshone and Paiute bands mentioned earlier apparently never attacked others and were themselves attacked only very rarely; most just fled rather than trying to defend themselves. But these few peaceful groups are exceptional. The cross-cultural samples indicate that the vast majority of other hunter-gatherer groups did engage in

warfare and that there is nothing inherently peaceful about hunting-gathering or band society.¹⁵

Pacifistic societies also occur (if uncommonly) at every level of social and economic complexity. Truly peaceful agriculturalists appear to be somewhat less common than pacifistic hunter-gatherers. In the cross-cultural samples discussed earlier, almost all the peaceful agricultural groups could be characterized as defeated refugees, ethnic minorities long administered by states, or tribes previously pacified by the police or by paramilitary organs of colonial or national states.¹⁶ Low-density, nomadic hunter-gatherers, with their few (and portable) possessions, large territories, and few fixed resources or constructed facilities, had the option of fleeing conflict and raiding parties. At best, the only thing they would lose by such flight was their composure. But with their small territories, relatively numerous possessions, immobile and labor-expensive houses, food stores, and fields, sedentary farmers or hunter-gatherers who attempted to flee trouble could lose everything and thereupon risk starvation. Farmers and sedentary hunter-gatherers had little alternative but to meet force with force or, after injury, to discourage further depredations by taking revenge. Groups that depended on very localized, essential resources—such as desert springs, patches of fertile soil, good pastures, or fishing stations—had to defend these or face severe privation. Even nomadic pastoralists in extensive grasslands had to defend their herds, wherever they might be. For obvious reasons, then, agriculturalists, pastoralists, and less nomadic foragers have seldom been entirely peaceful. But such pacifistic farmers have occasionally appeared.

The best-known peaceful agriculturalists are the Semai of Malaysia, who strictly tabooed any form of violence (although their homicide rate was numerically significant).¹⁷ Their reaction to any use of force involved "passivity or flight." Interestingly, they were recruited as counterinsurgency scout troops by the British during the Communist insurgency in Malaya in the 1950s. The Semai recruits were profoundly shocked to discover that as soldiers they were expected to kill other men. But after the guerrillas killed some of their kinsmen, they became very enthusiastic warriors. One Semai veteran recalled, "We killed, killed, killed. The Malays would stop and go through people's pockets and take their watches and money. We did not think of watches or money. We thought only of killing. Wah, truly we were drunk with blood." However, when the Semai scouts were demobilized and returned to their villages, they quietly resumed their nonviolent life-style. The low density of population, shifting settlement, and abundances of unused land probably allowed the Semai, unlike many other farmers, the option of flight from violent threats.¹⁸ But their strong moral distaste for violence was undoubtedly important in maintaining their peacefulness.

Peaceful societies even exist among industrial states. For example, neither Sweden nor Switzerland has engaged in warfare for nearly two centuries; their homicide rates are among the lowest in the world. Like many peaceful tribal societies, Switzerland is to some degree geographically isolated behind its mountains. Sweden was once home to the legendarily belligerent Vikings and remained one of the most warlike societies in Europe until the eighteenth century. Nevertheless, Sweden has fought no wars since 1815. Both nations traditionally maintain modern military forces; indeed, every Swiss male between the ages of twenty and fifty is a military reservist, while Sweden is one of the world's leading arms exporters. But they and a few other nations in Asia and South America offer testimony that there is nothing inherently warlike about states.

Thus pacifistic societies seem to have existed at every level of social organization, but they are extremely rare and seem to require special circumstances. The examples of Sweden and the Semai demonstrate that societies can change from pacifistic to warlike, or vice versa, within a few generations or, (as with the Semai) within the lifetime of an individual. As these examples and the case of the Polar Eskimo establish, the idea that violent conflicts between groups is an inevitable consequence of being human or of social life itself is simply wrong. Still, the overwhelming majority of known societies have made war. Therefore, while it is not inevitable, war is universally common and usual.

THE FREQUENCY OF WARFARE IN STATE AND NONSTATE SOCIETIES

How frequent are primitive wars, and do nonstate societies engage in warfare less frequently than states or civilized societies? These questions are related to the question of how intense primitive warfare is. Again turning to cross-cultural research, we find that many of the myths about primitive war are untrue.

The three cross-cultural surveys mentioned earlier also include data on the frequency of warfare. All these studies show that warfare has been extremely frequent among primitive societies.¹⁹ In the sample of fifty societies, 66 percent of the nonstates were continuously (meaning every year) at war, whereas only 40 percent of the states were at war this frequently. In this survey, warfare was therefore found to be less frequent in state societies. The larger sample of ninety societies, however, indicated that the frequency of war increased somewhat with greater political complexity; 77 percent of the states were at war once a year, whereas 62 percent of tribes and chiefdoms were this war prone. Nevertheless, 70 to 90 percent of bands, tribes, and chiefdoms went to war at least once every five years, as did 86 percent of the states. All these figures support yet another survey, which found that about 75 percent of all prestate societies went to war "at least once every two years before they were pacified or incorporated

by more dominant societies" and that warfare was no more frequent "in complex societies than in simple band or tribal societies." In the sample of U.S. western Indian tribes, which consisted wholly of nonstate societies, 86 percent were raiding or resisting raids undertaken more than once each year. And such high frequencies of fighting were not peculiar to North America.²⁰ For example, during a five-and-a-half-month period, the Dugum Dani tribesmen of New Guinea were observed to participate in seven full battles and nine raids. One Yanomamo village in South America was raided twenty-five times over a fifteen-month period. These independent surveys show that the great majority of nonstate societies were at war at least once every few years and many times each generation. Obviously, frequent, even continuous, warfare is as characteristic of tribal societies as of states.

The high frequencies of prestate warfare contrast with those of even the most aggressive ancient and modern civilized states. The early Roman Republic (510-121 B.C.) initiated a war or was attacked only about once every twenty years. During the late Republic and early Empire (118 B.C.-A.D. 211), wars started about once every six or seven years, most being civil wars and provincial revolts.²¹ Only a few of these later Roman wars involved any general mobilization of resources, and all were fought by the state's small (relative to the size of the population), long-service, professional forces supported by normal taxation, localized food levies, and plunder. In other words, most inhabitants of the Roman Empire were rarely directly involved in warfare and most experienced the Pax Romana unmolested over many generations.

Historic data on the period from 1800 to 1945 suggest that the average modern nation-state goes to war approximately once in a generation.²² Taking into account the duration of these wars, the average modern nation-state was at war only about one year in every five during the nineteenth and early twentieth centuries. Even the most bellicose, such as Great Britain, Spain, and Russia, were never at war every year or continuously (although nineteenth-century Britain comes close). Compare these with the figures from the ethnographic samples of nonstate societies discussed earlier: 65 percent at war continuously; 77 percent at war once every five years and 55 percent at war every year; 87 percent fighting more than once a year; 75 percent at war once every two years. The primitive world was certainly not more peaceful than the modern one. The only reasonable conclusion is that wars are actually more frequent in nonstate societies than they are in state societies—especially modern nations.

MOBILIZATION

The informal and voluntary mobilization for war supposedly characteristic of tribal societies is often cited as evidence of the lack of importance and effective-

of manpower mobilization, then, war is no less important to tribes than to nations.

PREHISTORIC WARFARE

With regard to prehistory, nothing comparable to the surveys of historical and ethnographic societies cited earlier exists as yet. Any attempts to survey 2 million years of human prehistory for evidence of violence and armed conflict face several daunting difficulties. The first is that most regions of the world are poorly known archaeologically—the rare exceptions being Europe (especially the west), the Near East, and parts of the United States. The most unequivocal evidence of armed conflict consists of human skeletons with weapon traumas (especially, embedded bone or stone projectile points) and fortifications. However, humans have buried their dead for only the past 150,000 years or so; before this, the human remains that have been found were often disturbed and fragmented by scavengers and natural forces. Even during the past 150,000 years, many prehistoric peoples disposed of their dead in ways—for example, cremation and exposure—that left no remains for anthropologists to study. Only among some peoples—those for whom the use of stone- and bone-tipped weapons (which can survive embedded in or closely associated with human skeletons) was commonplace—is it easy to distinguish accidental traumas from those inflicted by humans. The use of these weapons occurred only during the past 40,000 years, and in many regions perishable wooden and bamboo spears and projectiles continued to be used until modern times. Until humans began living in permanent villages, fortifications would have not repaid the labor required to construct them (Chapter 3). But humans seem to have become sufficiently sedentary only during the past 14,000 years, and permanent villages are common in most regions only after the adoption of farming (8000 B.C. at the earliest). Thus it is possible to document prehistoric warfare reliably only within the past 20,000 to 30,000 years and in only a few areas of the world. Granting these limitations, what does the archaeological evidence say about the peacefulness of prehistoric peoples?

Some authors have claimed that the evidence of homicide is as old as humanity—or at least as old as the genus *Homo* (that is, over 1 million years).²⁵ But many of the traumas found on early hominid skeletons have been proved by subsequent investigation to have had nonhomicidal causes or cannot be distinguished from accidental traumas of a similar character.²⁶ For instance, the paired “spear wounds” found on some South African Australopithecine skulls are now recognized as punctures created by leopard canines as the predator carried these luckless ancestors of ours, gripping their heads in its teeth. As another example, Neanderthals seem to have been especially accident prone,

compared with the modern humans who followed them. Neanderthals’ bones evidence many injuries and breakages (one study determined that 40 percent of them had suffered head injuries). Which, if any, of these injuries were caused by human violence cannot be determined. Since the heavy musculature and robust bones of Neanderthals imply that their way of life was much more strenuous and physically demanding than that of more recent humans, it seems probable that most of the traumas in question were accidental. Why they so often “forgot to duck” remains a mystery, however.

Whenever modern humans appear on the scene, definitive evidence of homicidal violence becomes more common, given a sufficient sample of burials.²⁷ Several of the rare burials of earliest modern humans in central and western Europe, dating from 34,000 to 24,000 years ago, show evidence of violent death. At Grimaldi in Italy, a projectile point was embedded in the spinal column of a child’s skeleton dating to the Aurignacian (the culture of the earliest modern humans in Europe, ca. 36,000 to 27,000 years ago). One Aurignacian skull from southern France may have been scalped; it has cut-marks on its frontal (forehead). Evidence from the celebrated Upper Palaeolithic cemeteries of Czechoslovakia, dating between 35,000 and 24,000 years ago, implies—either by direct evidence of weapons traumas, especially cranial fractures on adult males, or by the improbability of alternative explanations for mass burials of men, women, and children—that violent conflicts and deaths were common. In the Nile Valley of Egypt, the earliest evidence of death by homicide is a male burial, dated to about 20,000 years ago, with stone projectile points in the skeleton’s abdominal region and another point embedded in its upper arm (a wound that had partially healed before his death). The one earlier human skeleton found in Egypt bears no evidence of violence, but the next more recent human remains there are rife with evidence of homicide.

The human skeletons found in a Late Palaeolithic cemetery at Gebel Sahaba in Egyptian Nubia, dating about 12,000 to 14,000 years ago, show that warfare there was very common and particularly brutal.²⁸ Over 40 percent of the fifty-nine men, women, and children buried in this cemetery had stone projectile points intimately associated with or embedded in their skeletons. Several adults had multiple wounds (as many as twenty), and the wounds found on children were all in the head or neck—that is, execution shots. The excavator, Fred Wendorf, estimates that more than half the people buried there had died violently. He also notes that homicidal violence at Gebel Sahaba was not a once-in-a-lifetime event, since many of the adults showed healed parry fractures of their forearm bones—a common trauma on victims of violence—and because the cemetery had obviously been used over several generations. The Gebel Sahaba burials offer graphic testimony that prehistoric hunter-gatherers could be as

ruthlessly violent as any of their more recent counterparts and that prehistoric warfare continued for long periods of time.

In western Europe (and more poorly known North Africa), ample evidence of violent death has been found among the remains of the final hunter-gatherers of the Mesolithic period (ca. 10,000 to 5,000 years ago).²⁹ One of the most gruesome instances is provided by Ofnet Cave in Germany, where two caches of "trophy" skulls were found, arranged "like eggs in a basket," comprising the disembodied heads of thirty-four men, women, and children, most with multiple holes knocked through their skulls by stone axes. Indeed, some archaeologists, impressed by the abundant evidence of homicide in the European Mesolithic, date the beginnings of "real" war to this period.

Indications of conflict, as reflected by violent death and the earliest fortifications, become especially pervasive in western Europe during the ensuing Neolithic period (the era of the first farmers, ca. 7,000 to 4,000 years ago, depending on the region).³⁰ Some archaeologists have argued that real warfare begins only when hunters become farmers. This mistaken point of view does have some especially grim support in the remains of Neolithic mass killings at Talheim in Germany (ca. 5000 B.C.) and Roaix in southeastern France (ca. 2000 B.C.). At Talheim, the bodies of eighteen adults and sixteen children had been thrown into a large pit; the intact skulls show that the victims had been killed by blows from at least six different axes.³¹ More than 100 persons of all ages and both sexes, often with arrowpoints embedded in their bones, received a hasty and simultaneous burial at Roaix. The villages of the first farmers in many regions of western Europe were fortified with ditches and palisades. Several of these early enclosures in Britain, after being extensively excavated, yielded clear evidence of having been attacked, stormed, and burned by bow-wielding enemies. The early agricultural tribes and petty chiefdoms of Neolithic Europe were anything but peaceful.

Interestingly, the historically blood-soaked Near East has yielded little evidence of violent conflict during the Early Neolithic.³² Although extensive and elaborate fortifications were erected during this period at Jericho, they became common in the Near East only in the later Neolithic and in the Bronze Age.

When we turn to the United States—specifically to those areas that have been subject to intensive archaeological scrutiny and where large samples of human burials have been excavated, such as the Southwest, California, the Pacific Northwest Coast, and the Mississippi drainage—violent deaths are at least in evidence and, in some periods, were extremely common.³³ Fortifications were constructed at various times and in various regions by prehistoric farmers in the Mississippi drainage and in the Southwest, as well as by the prehistoric sedentary hunter-gatherers of the Northwest Coast.³⁴ As with the best-studied re-

gions of the prehistoric Old World, the prehistoric New World was also a place where the dogs of war were seldom on a leash.

In each of these regions, the indications are that warfare was relatively rare during some periods; nothing suggests, however, that prehistoric nonstatic societies were significantly and universally more peaceful than those described ethnographically. The archaeological evidence indicates instead that homicide has been practiced since the appearance of modern humankind and that warfare is documented in the archaeological record of the past 10,000 years in every well-studied region. In the chapters that follow, it will become clear that archaeological evidence strongly supports ethnographic accounts concerning the conduct, consequences, and causes of prestate warfare.

There is simply no proof that warfare in small-scale societies was a rarer or less serious undertaking than among civilized societies. In general, warfare in prestate societies was both frequent and important. If anything, peace was a scarcer commodity for members of bands, tribes, and chiefdoms than for the average citizen of a civilized state.

THE NATIONAL INTEREST

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

Published on *The National Interest* (<http://nationalinterest.org>)

Source URL (retrieved on Dec 10, 2013): <http://nationalinterest.org/commentary/what-our-primate-relatives-say-about-war-7996>

What Our Primate Relatives Say About War

| [More](#) [1]

| [Dominic Johnson](#) [2], [Bradley A. Thayer](#) [3]

| January 29, 2013

| [Dominic Johnson](#) [2], [Bradley A. Thayer](#) [3]



In the early 1930s, with political unrest in Europe and war on the horizon, Albert Einstein wrote to Sigmund Freud asking: “Why war?” Einstein sought an answer to a simple and fundamental question that has been on the minds of scholars and practitioners throughout history. Freud’s response was that war was the result of an impulse, a destructive instinct found in many humans.

For many students of conflict, little has changed in the eighty years since—whatever the political or historical context, war seems to have something to do with human nature. Indeed, this notion has deep and illustrious roots. Most explanations of the causes of war are rooted in two philosophical camps. For followers of Swiss philosopher Jean-Jacques Rousseau, war is a tragic result of misunderstandings and the negative influences of our

society. For English philosopher Thomas Hobbes, war is likewise tragic, but rooted in a metaphysical spirit that causes people to endlessly attempt to dominate others. With the end of the Cold War, there were high hopes that both of these pessimistic outlooks could be consigned to history and the millennium would bring forth a radically different world—a new world order. Yet, despite signs of many types of violence falling in western societies (a point recently championed by Steven Pinker), war continues unabated in the new century.

The 21st century does, however, promise new hope—at least in our ability to understand it. Terrific advances in the life sciences equip us with novel tools and insights to help us get to the bottom of Einstein's question. While these insights draw on a range of scientific disciplines including genetics, physiology, psychology and neuroscience, the unifying framework to understand human nature as a whole is evolution. With a clearer picture of where we came from, we may find a better understanding of who we are and where we are going. As a primate, key to this objective is understanding human evolutionary history and the behavior of our closest relatives. The results are surprising and significant for Einstein's query.

Primatologists who have studied our closest relative of all, chimpanzees, have depressing news. Chimpanzee behavior is "bad." They form coalitions to wage lethal violence against rival groups, ambushing individuals of other groups to maim and kill them. Jane Goodall, the foremost student of their behavior, wrote in her classic *The Chimpanzees of Gombe*: "as a result of a unique combination of strong affiliative bonds between adult males on the one hand and an unusually hostile and violently aggressive attitude toward nongroup individuals on the other," the chimpanzee "has clearly reached a stage where he stands at the very threshold of human achievement in destruction, cruelty, and planned intergroup conflict." Other primatologists since have verified these conclusions, in different locales and settings, with detailed descriptions of systematic and lethal intergroup violence. Hobbes seems to have described chimpanzee behavior very accurately, reinvigorating the question of how much it applies to humans as well.

The example of the bonobos (pygmy chimpanzees in central Africa) is often raised in contrast to chimpanzee behavior, and by implication, an alternative perspective on human behavior as well. Bonobos are seen as the "good" ape because they seem to be the antithesis of the chimpanzee. Although not as well studied as chimpanzees, enough is known about them to recognize that bonobos are different, including exhibiting less aggression than chimpanzees. They still experience significant conflict, at least within groups, but it is manifested and resolved in different ways. Indeed, sex appears to replace war as their distinguishing characteristic. Their sexual activity, both heterosexual and homosexual, appears to be used for stress reduction, conciliatory purposes and resource competition. Given what seems to be a lifestyle straight out of San Francisco's bohemian Haight-Ashbury district, people could be forgiven for wishing humans were related to bonobos, inferring that they prove Rousseau right rather than Hobbes, and that they serve as a better model for humans than chimpanzees.

However, there are three very important qualifications in drawing inferences from our primate ancestry. First, the two species of chimpanzee are much more closely related to one another than either is to us. Chimpanzees and bonobos separated perhaps 0.86–0.89 million years ago (mya) and perhaps as long ago as between 1.5–2.5 mya. By contrast, we all shared a common ancestor long before, about 5–6 mya. This common ancestor is likely to have been different from all of us—different from humans, chimps, and bonobos. Understanding this phylogeny is important because it means there is little reason to believe that bonobos are the best model of human behavior. Chimpanzees are not a perfect model either, but they are likely to be a better one because they share more social and ecological similarities with humans than bonobos do.

This leads us to the second qualification. A key reason why humans, bonobos and chimpanzees differ at all is the environment in which they evolved, and that environment's selective pressures shaping the evolution of their social behavior. Here again, bonobos are the odd man out. Humans and chimpanzees have occupied diverse habitats across a very large range in the last several million years, stretching across the breadth of Africa and often comprising of patchy food resources that give rise to competition. In contrast, the ecology of the bonobos has been very different. Confined to a range below the Congo river in central Africa, bonobos occupied a more stable and benign habitat with dense, lowland forest and more evenly distributed and high-quality foliage, and the absence of competition from gorillas, all of which reduced between-species and within-species competition, especially among males. Accordingly, ecology makes bonobos a less representative model than chimpanzees for understanding human behavior.

Third, human behavior must be understood as a consequence of evolutionary history both *long before* and *long after* any common ancestor with primates. Let's consider those older roots first. Humans share many traits that are common not just to primates, but to all mammals and even all vertebrates extending tens or hundreds of million of years into the past. For example, evolutionary older parts of the human brain elicit responses such as fear, aggression and the fight or flight response, which are shared among all mammals. Forget chimps or bonobos—much of human behavior can be understood as *mammal-typical*. But by the same logic many human traits are primate- or great ape-typical as well, making gorilla and orangutan behavior also important contexts for a full understanding of human behavior.

Now let us consider human evolution since our great ape common ancestor. Although, as argued above, much of human behavior carries the baggage of our distant evolutionary heritage, it has been fine tuned as the result of adaptations to the unique physical and social environment in which we—our species—evolved over the last 2 million years or so. Over and above whatever chimps or bonobos do, therefore, humans display adaptations to deal with problems that were uniquely ours. While there are many things unique to humans, perhaps the most significant has been collective action in large groups of non-kin, and defense and offense against rival groups. Inter-group conflict may be important among chimpanzees, but *Homo sapiens* turned it into an art. Although the details remain highly controversial, a series of new studies in archaeology and

anthropology have debunked Rousseau's myth of the peaceful savage. Death rates (as the percentage of adult males killed in intergroup conflict) among indigenous and prehistoric societies make the wars of the 20th century seem like skirmishes. Although humans were not always at war, human societies were always organized around its ever-present threat.

A sensitivity to human evolution and the behavior of chimpanzees and bonobos forces students of modern students of conflict to face two hard truths. First, we evolved in conditions of resource competition where fear of others, aggression and violence offered adaptive solutions to protect and provide for ourselves and our kin. We therefore need to amend Clausewitz. Humans do indeed wage war for political purposes, but long before war for *raison d'état* there was war for resources. International politics is therefore not the root of war but merely an example of it—the continuation of seeking access to valuable resources by other means. Accordingly, when we consider “Why war?,” we have an answer: war is one of Mother Nature's solutions to compete successfully for resources.

Second, the human traits of egoism, dominance, and in-group/out-group bias are at least partly adaptations to the ecological conditions prevalent in *human* evolution. It is not assumed that we simply inherited these wholesale from a common ancestor, or the common ancestor we share with the chimpanzee. Clearly, we have undergone many physiological and behavioral changes since then and ecology has been as or more important. But although humans and chimpanzees appear to have travelled much of the road to war together, we have gone far further. The particular socioecological setting in which humans evolved meant that aggression and war were significant behavioral adaptations. These same settings led to remarkable levels of cooperation as well, but note that this cooperation is selectively directed towards in-group members, the better to avoid exploitation by rival groups and organize for war. These adaptations, lamentably, remain with us today and influence our behavior, politics and society.

The true source of the tragedy of the human condition is that we evolved in conditions of intense resource competition. Like Russell Crowe organizing the outnumbered combatants in *Gladiator*, natural selection favored individuals who cooperated to avoid being killed—and, if necessary, to kill. Regrettably, Einstein's concise question, “Why war?,” is answered by nature equally succinctly: because we are human.

Dominic Johnson is Professor in the Department of Politics and International Relations, University of Oxford. Bradley A. Thayer is Professor and Head of the Department of Political Science, Utah State University.

Image: Wikimedia Commons/Caelio [4]. CC BY-SA 3.0.

More by

[Dominic Johnson](#) [2], [Bradley A. Thayer](#) [3]

Topics:

- [Society](#) [5]

Tags:

- [Albert Einstein](#) [6]
- [Jane Goodall](#) [7]
- [Jean-Jacques Rousseau](#) [8]
- [Sigmund Freud](#) [9]
- [Steven Pinker](#) [10]
- [Thomas Hobbes](#) [11]

- [Apes](#) [12]
- [Biology](#) [13]
- [Bonobo](#) [14]
- [Chimpanzee](#) [15]
- [Ethologists](#) [16]
- [Human](#) [17]
- [Primate](#) [18]
- [Primatologists](#) [19]
- [Primateology](#) [20]
- [Zoology](#) [21]

Source URL (retrieved on Dec 10, 2013): <http://nationalinterest.org/commentary/what-our-primate-relatives-say-about-war-7996>

Links:

- [1] <http://www.addthis.com/bookmark.php?v=250&username=nationalinterest>
- [2] <http://nationalinterest.org/profile/dominic-johnson>
- [3] <http://nationalinterest.org/profile/bradley-a-thayer>
- [4] http://commons.wikimedia.org/wiki/File:Adult_male_chimps_in_mahale.jpg
- [5] <http://nationalinterest.org/topic/society>
- [6] <http://nationalinterest.org/tag/person/albert-einstein>
- [7] <http://nationalinterest.org/tag/person/jane-goodall>
- [8] <http://nationalinterest.org/tag/person/jean-jacques-rousseau>
- [9] <http://nationalinterest.org/tag/person/sigmund-freud>
- [10] <http://nationalinterest.org/tag/person/steven-pinker>
- [11] <http://nationalinterest.org/tag/person/thomas-hobbes>
- [12] <http://nationalinterest.org/tag/apes>
- [13] <http://nationalinterest.org/tag/biology>
- [14] <http://nationalinterest.org/tag/bonobo>
- [15] <http://nationalinterest.org/tag/chimpanzee>
- [16] <http://nationalinterest.org/tag/ethologists>
- [17] <http://nationalinterest.org/tag/human>
- [18] <http://nationalinterest.org/tag/primate>
- [19] <http://nationalinterest.org/tag/primatologists>
- [20] <http://nationalinterest.org/tag/primatology>
- [21] <http://nationalinterest.org/tag/zoology>

WILEY

Journal of Religious Ethics, Inc

The Elements of St. Augustine's Just War Theory

Author(s): John Langan

Source: *The Journal of Religious Ethics*, Vol. 12, No. 1 (Spring, 1984), pp. 19-38

Published by: on behalf of Journal of Religious Ethics, Inc

Stable URL: <http://www.jstor.org/stable/40014967>

Accessed: 11/12/2013 09:07

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)



Wiley and *Journal of Religious Ethics, Inc* are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Religious Ethics*.

<http://www.jstor.org>

THE ELEMENTS OF ST. AUGUSTINE'S JUST WAR THEORY

John Langan, S.J.

ABSTRACT

St. Augustine's just war theory involves eight principal elements: *a*) a punitive conception of war, *b*) assessment of the evil of war in terms of the moral evil of attitudes and desires, *c*) a search for authorization for the use of violence, *d*) a dualistic epistemology which gives priority to spiritual goods, *e*) interpretation of evangelical norms in terms of inner attitudes, *f*) passive attitude to authority and social change, *g*) use of Biblical texts to legitimate participation in war, and *h*) an analogical conception of peace. It does not include non-combatant immunity or conscientious objection. A contemporary assessment of the elements is offered.

Christian thought about justifying reasons for war and limits on the conduct of war can be described either as a presentation of just war theory or as an integral part of just war tradition. A special part of the interest of just war theory is that it is built on a long and rich historical development, just as a special part of the interest of the just war tradition is that it is the history of what is still a very important instrument for dealing with some of the most difficult problems that we confront as human beings and as members of a free political community. Both the tradition and the theory have been identified over time as the Christian approach to the problem of violence, and in a very rough way this is true.

But such a claim should not be taken in any exclusive sense. For there is a long and important tradition of Christian pacifism both as a theological position and as a form of Christian witness in a world full of the sorrows and crimes so often found in war. On the other hand, there have been episodes of violence for which a special religious blessing has been sought or given. There has also been a recurrent interweaving of religious and military values in ways which often exceed without repudiating the rather austere limits of just war theorizing on the use of violence. The Christian tradition, if we conceive it in historical rather than normative terms, includes crusaders and conquistadors, kings and knights, rebels and sheriffs. Even in the case of Roman Catholicism, the church which has had the longest and closest official identification with just war doctrine, it would be a mistake to equate the church's teaching and experience simply with this doctrine or approach.

Recent developments in American Catholicism from the time of Dorothy Day onwards should make this clear. On the other hand, a just war approach to the problem of violence is not confined to Christian believers. The approach has roots in the teachings of Cicero and in the legal and moral theory of natural law and Greek philosophy. It continues to flower in the contemporary work of Michael Walzer. We are then confronted with a wealth of material in just war tradition and theory. This wealth can be a resource for reflection or a cause of embarrassment, for it makes it very difficult to identify any particular formulation as simply *the* just war theory and it means that any presentation of or appeal to the just war tradition will have to be highly selective. The task of the present paper is to take one of the crucial figures in the development of the just war tradition, to present the elements of his position as constituting the bases of just war theory, and to suggest some possibilities for the use of these elements in current just war debates.

I. RELIGIOUS AUTHORIZATION OF WAR

Around the year 400 Augustine published a polemical work *Against Faustus*, a proponent of Manicheanism who figures in Book V of the *Confessions*. A central theme of the work is the continuity and compatibility of the Old and New Testaments, a point which the Manicheans had denied in setting up their fundamental opposition between the defective and inferior deity of the Hebrew Bible who is the creator of matter and evil and the perfectly good Father of Jesus. This polemical concern leads Augustine to defend Moses for the spoliation of the Egyptians and the wars of the Israelites. The general strategy which he adopts has considerable affinity with divine command ethics and with Aquinas's treatment of divinely authorized exceptions to the precepts of natural law. He says (Milhaven, 1968: 161):

According to the eternal law, which requires the preservation of natural order, and forbids the transgression of it, some actions have an indifferent character, so that men are blamed for presumption if they do them without being called upon, while they are deservedly praised for them when required. The act, the agent, and the authority for the action are all of great importance in the order of nature.

The translation here is a bit misleading, since we are not dealing with a class of indifferent actions, which a moral agent is free to perform or not without making himself or herself liable to moral criticism or blame, but rather with actions which are *prima facie* wrong, at least to the extent that there is some burden of justification to be carried by the person who would perform them. In fact, the example that Augustine offers emphasizes this point since it is the sacrifice of Isaac by Abraham, an action that, without

divine authorization, would universally be held to be wrong. As Augustine observes (1887:300), "For Abraham to sacrifice his son of his own accord is shocking madness. His doing so at the command of God proves him faithful and submissive." The example also indicates that the three key elements, "the act, the agent, and the authority for the action," are not adequately distinct. It is the divine authorization that makes the action or attempted or contemplated action praiseworthy rather than indifferent or in need of justification and that makes Abraham "faithful and submissive" (*fidelis et devotus*). Actually one is a bit surprised to see Augustine hanging so much on the divine authorization in this case. No matter what one's theological views, there must be very few theologians or ecclesiastics today who would point to God's demand for the sacrifice of Isaac as a perspicuous manifestation of the moral character of the deity. The appeal to this case does, however, have a certain *ad hominem* value, since Augustine affirms that Faustus is not willing to raise moral objections to the sacrifice of Isaac as he did to the despoiling of the Egyptians. With regard to the latter case, Augustine urges (1887:300), "Your feeling of disapproval for the mere human action should be restrained by a regard for the divine sanction (*auctoritas*)."

When he actually gets to considering the question of war, Augustine points to the divine authorization for the wars of Moses and says of Moses that "in wars carried on by divine command, he showed not ferocity but obedience" (Augustine, 1887: 301). The difference here is not a matter of altering the way in which Moses conducts wars, or the limits that he observes in the use of violence, but a difference in motivation and justification. Augustine's views here are connected with his remarkable position on what constitutes the evil of war. He asks:

What is the evil in War? Is it the death of some who will soon die in any case, that others may live in peaceful subjection? This is mere cowardly dislike, not any religious feeling. The real evils in war are love of violence (*nocendi cupiditas*), revengeful cruelty (*ulciscendi crudelitas*), fierce and implacable enmity, wild resistance, and the lust of power (*libido dominandi*) and such like. (Augustine, 887:301)

For a modern reader, the language here prefigures the famous description of the state of nature as a state of war given by Thomas Hobbes in the celebrated Chapter 13 of *Leviathan*. Both Hobbes and Augustine would agree that such a condition is menacing and unattractive to reasonable human beings. Hobbesian egotists who are living in such a situation without moral constraints will see it as a threat to their self-preservation and will, if they are reasonable, agree on articles of peace. This move rests on a clear agreement about the importance of earthly values. Thus Hobbes writes (1947: 84): "The passions that incline men to peace, are fear of death; desire of such things as are necessary to commodious living; and a hope by their industry

to obtain them." But what Augustine finds repellent in the state of war is its expression of human selfishness and of the disordered desires of human beings. As Augustine observes (1887: 303):

Thus, in all the things which appear shocking and terrible to human feebleness, the real evil is the injustice: the rest is only the result of natural properties or of moral demerit. This injustice is seen in every case where a man loves for their own sake things which are desirable only as means to an end, and seeks for the sake of something else things which ought to be loved for themselves.

A Hobbesian peace which left unchallenged the distorted values of people who fail to "use things only for the end for which God appointed them, and to enjoy God as the end of all" (Augustine, 1887: 303) would not count as a just or satisfactory peace for Augustine, even though it might prevent various forms of violent behavior and resulting harm. In fact, the response that Augustine proposes with regard to the evils of war is not to strike a rational agreement but to punish them. He writes (1887: 301): "It is generally to punish these things, when force is required to inflict the punishment, that, in obedience to God or some lawful authority, good men undertake wars." The resort to violence that is inherent in war is undertaken, not as a means of self-defense, but as a punitive effort initiated by lawful authority.

Augustine interprets war along lines inspired by the Old Testament as both an element in religious pedagogy and an exercise of divine power and judgment:

When war is undertaken in obedience to God, who would rebuke, or humble, or crush the pride of man, it must be allowed to be a righteous war; for even the wars which arise from human passion cannot harm the eternal well-being of God, nor even hurt His saints; for in the trial of their patience, and the chastening of their spirit, and in bearing fatherly correction, they are rather benefited than injured. (Augustine, 1887: 301)

This gives us an approach to the whole problem of war which sees it in primarily spiritual and attitudinal terms rather than as a threat to human interests and survival or as the doing of actions which are evil. It is this approach which lies behind Augustine's astonishing remark (1887: 303) about Moses that "there would have been less harm in making war of his own accord, than in not doing it when God commanded him." Augustine's insistence on the power and the mystery of God's providence leads to a kind of agnosticism about the value of what human beings do and suffer in the course of war:

As the judgments of God and the movements of man's will contain the hidden reason why the same prosperous circumstances which some make a right use of are the ruin of others, and the same afflictions under which some give way are profitable to others, and since the whole mortal life of man upon earth is

a trial, who can tell whether it may be good or bad in any particular case—in time of peace, to reign or to serve, or to be at ease or to die—or in time of war, to command or to fight, or to conquer or to be killed. (Augustine, 1887: 303)

This level of agnosticism about human values and of abandonment to divine providence takes the whole task of making moral decisions about war out of the hands of individual moral agents in two ways. First, because it questions our ability to judge what is really for our good, it leaves us fundamentally passive in the face of the workings of divine providence. Second, it turns the question of determining the justice of war and the right use of violence into a search for an appropriate authorization. Thus Augustine concludes (1887: 301):

A great deal depends on the causes for which men undertake wars, and on the authority they have for doing so: for the natural order which seeks the peace of mankind ordains that the monarch should have the power of undertaking war if he thinks it advisable, and that the soldiers should perform their military duties in behalf of the peace and safety of the community.

Authorization to take part in hostilities, even when this comes from “an ungodly king” or involves an “unrighteous command” leaves the soldier innocent “because his position makes obedience a duty” (Augustine, 1887: 301).

It is sometimes observed that Augustine spiritualized the command to turn the other cheek and not to resist aggression (Mt. 5.39). In one sense, this is clearly true. As Augustine says quite openly in the present work (1887: 301): “What is here required is not a bodily action, but an inward disposition.” But we would be mistaken if we took this spiritualization as a mere evacuation of the demanding content of the Gospel or as a veil behind which Christians would be free to assert themselves or to pursue their own selfish desires. Rather, it is part of Augustine’s effort to seek our happiness in another life and another kingdom. The Old Testament does not primarily provide us with a collection of examples authorizing violence on religious grounds, although it does do that. Rather, it serves as a stage in a progressive revelation of what is truly good. Augustine writes (1887: 302):

The patriarchs and prophets, then, have a kingdom in this world, to show that these kingdoms, too, are given and taken away by God: the apostles and martyrs had no kingdom here, to show the superior desirableness of the kingdom of heaven.

Not merely is the kingdom of God ultimately what we are to aspire to in our desires and attitudes; but the authority of God is our ultimate guarantee of the righteousness of what we do. So Augustine can affirm (1887: 301) that the man must “be blameless who carries on war on the authority of God, of whom every one who serves him knows that He can never require what is wrong.”

II. THE ELEMENTS OF AUGUSTINE'S POSITION

As we reflect on the complex of ideas presented in this important work of Augustine, we can discern some principal themes: *a*) a conception of war as punitive rather than defensive, *b*) an assessment of the evil of war in terms of the moral evil of certain attitudes and desires rather than in terms of damage to premoral interests and values or in terms of actions wrong in themselves or by reason of their consequences, *c*) a search for appropriate authorization, either divine or human, for the use of violence, *d*) a divided epistemological stance which includes certainty with regard to the superiority of spiritual goods and uncertainty about the ultimate desirability of other events and experiences and their connection with the higher spiritual goods, *e*) a willingness to interpret evangelical norms in terms of inner attitudes rather than overt actions, *f*) an assumption of general social passivity and quiescence in the decisions and moral judgments of authority. To these, two other important elements have to be added, one of which is already present in the text we are analyzing from the *Contra Faustum*. This is *g*) the appeal to specific New Testament texts to legitimate military service and participation in war. Thus Augustine here and elsewhere refers to Luke 3: 14, the words of John the Baptist to the soldiers, and cites the favorable reference to the centurion in Matthew 8. The final element to be brought in is *h*) Augustine's notion of peace, which does not figure in the text we have been examining.

We can see most of these elements at work in two important letters of Augustine to Roman officials in Africa. These are Letter 138, written in 412 to the tribune Marcellinus, and Letter 189 to Boniface, the governor of Africa, in 418. Marcellinus had reported a pagan arguing that Christian "preaching and doctrine were not adaptable to the customs of the state," and instancing the precepts about not returning evil for evil and about turning the other cheek (Augustine, 1953: 41). Augustine characterizes these precepts (1953: 4) as "not worked out by human arguments but written by divine authority." He interprets them not in merely negative and passive terms but along the lines of Paul's admonition in Romans 12.21 to overcome evil with good. The patient bearing of evil without retaliation may produce a change of heart in the evildoer. Augustine comments (1953: 43-44):

Evil is overcome by good in the evil man, and the man is set free, not from an exterior foreign evil, but from an interior, personal one, by which he is more grievously and ruinously laid waste than he would be by the inhumanity of any enemy from without. Therefore, he overcomes evil by good who suffers the loss of temporal goods with patience, in order to show how far these goods are to be despised for the sake of faith and justice.

Here we can see Augustine's consistent affirmation of the primacy of spiritual goods (element *d* above). The patient relinquishing of temporal goods

has a pedagogical value, and it helps to liberate the other persons from concern about such goods, which a struggle to protect one's possession of such goods would not do. Augustine, however, does not absolutize the value of non-resistance. Its value is instrumental, and Augustine (1953: 44) treats it as a tactic when he observes that "the right time for this to be done is when it seems likely to benefit the one for whose sake it is done, in order to bring about correction and a return to agreement."

Augustine's primary concern is with a conversion of mind and heart, and his preference is for achieving this by the example of patience. But as the term "correction" may indicate, there is also reliance on punitive methods. Augustine puts it thus (1953: 4): "We often have to act with a sort of kindly harshness, when we are trying to make unwilling souls yield, because we have to consider their welfare rather than their inclination." The complex epistemological stance mentioned as element *d* above includes both the affirmation of the priority of spiritual goods and a strong paternalistic tendency, in which one is willing to take action overriding others' conception of what constitutes their good. Thus Augustine writes (1953: 4):

He whose freedom to do wrong is taken away suffers a useful form of restraint, since nothing is more unfortunate than the good fortune of sinners, who grow bold by not being punished – a penalty in itself – and whose evil will is strengthened by the enemy within.

Part of the attraction of the punitive model of war for Augustine is that it contributes to the restoration of a moral order in which the various goods are properly estimated and in which human passions are restrained. Thus it is a sign of God's mercy that "wars should be waged by the good, in order to curb licentious passions by destroying those vices which should have been rooted out and suppressed by the rightful government" (Augustine, 1953: 4). It is, I would argue, crucial to our understanding of Augustine's approach to the just war that he is really interested in the preservation of a moral order which is fundamentally a right internal order of dispositions and desires and in which the question of whether action is violent or not is not fundamental. The restoration of that order constitutes a sufficient justification for resort to violence. In this sense, the problem that confronts Augustine in justifying the use of violence is closer in its fundamental epistemological and ethical character to the New Testament than it is to arguments for pacifism and against coercion on the basis of a liberal permissiveness and agnosticism or to arguments for justified violence on the basis of protecting the freedom and the interests of the individual or the community, arguments which work with the categories of modern political theory and moral philosophy and which address a pluralistic audience which has been subject to strong sceptical and relativistic influences.

The preservation of order and the avoidance of disordered interior states

also figure as important themes in Letter 189 to Boniface in 418. After an eloquent invocation of charity and a depiction of our progress to the promised kingdom, Augustine (1955:26) tells Boniface: "Do not imagine that no one can please God while he is engaged in military service." Augustine is willing to draw a parallel between the preservation of order through prayer and through warfare: "Thus some fight for you against invisible enemies by prayer, while you strive for them against visible barbarians by fighting" (Augustine, 1955:29). Augustine reminds Boniface that war is waged for the sake of peace and that he is to wage war as a peacemaker. Violence is appropriate in dealing with rebels who reject peace. War is "the result of necessity," and therefore "let it be necessity, not choice, that kills your warring enemy" (Augustine, 1955:26). Here we can see a recurring tendency in Augustine's language about the justified use of violence to make this less than a full object of responsible choice. This cuts two ways: on the one hand, it reminds us that violence is justified in the Christian tradition as a lesser evil in a situation of restricted choice among unattractive options; on the other hand it suggests an alienated attitude to violent action which may ultimately lead to a denial of one's responsibility and freedom. We should, I suggest, connect this with Augustine's search for authorization for violence so that the action is not fully the result of one's own freedom. The presence of some of these problems about violence in Augustine's thought from a very early period can be attested if we look back at an early anti-Manichean dialogue of Augustine, *De libero arbitrio*, of 388. Here Augustine considers the problem of self-defense, but not so much in terms of a conflict of rights as in terms of the internal dispositions one brings to actions in one's own interest. Evodius, Augustine's friend and interlocutor in the dialogue, asks: "How am I to judge that these men are free of passion who take up the sword in defense of things that can be lost against their will?" (Augustine, 1967:81). But he then offers the reassuring observation that a law which is enacted for the protection of people and which authorizes violence "cannot be charged with passion" (Augustine, 1967:81). Augustine then concludes that "one can, therefore, without passion, obey a law enacted for the protection of its citizens when it commands that an enemy force be met by the same kind of force" (Augustine, 1967:82). He draws a similar conclusion for public servants in general, who serve under an established order. But though the law leaves private persons free to use violence in self-defense, Evodius argues that, since the law does not compel them, they are wrong to use force to protect things "which can be lost against their will and which, on this account ought not to be loved at all" (Augustine, 1967:82).

Actions to protect such things may be permitted by human laws, but are to be punished by divine providence. Augustine here anticipates a point made later by Aquinas (1966:31) to the effect that not all evils are appropriately forbidden by human laws. The fact that a good can be taken from us involuntarily shows that it is an inferior good, for the defense of which violence is

not justifiable. This rests in part on a standard Stoic contrast between what is within our control and what is not and is a key move in what Sir Isaiah Berlin (1969: 135) has referred to as "the retreat to the inner citadel." The goods within our control are the virtues and attributes of the soul (e.g., chastity), which are among the spiritual goods which are more valuable in Augustine's system generally. A consequence of this move is a radical depreciation of the right of self-defense, which Augustine expresses thus (1967: 82):

Whatever the slain attacker was going to snatch from us is something not entirely within our power and, consequently, I fail to see how we can call it our own. Accordingly, I certainly am not blaming the law which permits such assailants to be slain, yet I can find no way to defend those who kill them.

This denial of legitimate self-defense for the protection of goods which can be lost is the underlying reason for Augustine's adoption of a punitive model of war (element a above) in preference to a defensive model. Augustine here is in the sharpest possible opposition to Hobbes's conception of the right of nature which he defines as:

the liberty each man hath, to use his own power, as he will himself, for the preservation of his own nature; that is to say, of his own life; and consequently of doing anything, which in his own judgment and reason, he shall conceive to be the aptest means thereunto. (Hobbes, 1947: 84)

This conception of a natural right of self-defense which is to be given the widest scope in the state of war and which is inherent in the individual is probably not held in all its amoral starkness by many today either among political theorists or among ordinary citizens. But I hazard the guess that most contemporary people, if forced to a choice between Hobbes and Augustine on this point, would side with Hobbes.

III. PEACE

The texts we have been looking at come from different phases of Augustine's life, from soon after his conversion to well into his episcopal career, and they represent different forms in his literary output: letters, polemical treatise, dialogue. But I would argue that they form a set of ideas which show a reasonable constancy and coherence in Augustine's view of war and which manifest interesting connections with larger themes in his theology and his moral theory. Some of these connections and some idea of the place of peace in relation to war can be seen if we look briefly at some passages from Book XIX of Augustine's most ambitious work, *The City of God*.

When he contemplates the unity of the Roman empire, Augustine laments the wars that were necessary to produce this unity, the wars that arose from the complexity and divisions of the empire, and the wars necessary to protect

this unity against external empires. Even if these wars are just, they are for Augustine a cause for lamenting:

For it is the wrongdoing of the opposing party which compels the wise man to wage just wars; and this wrongdoing, even though it gave rise to no war, would still be matter of grief to man because it is man's wrongdoing. Let every one, then, who thinks with pain on all these great evils, so horrible, so ruthless, acknowledge that this is misery and if any one either endures or thinks of them without mental pain, this is a more miserable plight still, for he thinks himself happy because he has lost human feeling. (Augustine, 1950:683)

Here we again encounter Augustine's tendency to focus on inner attitudes rather than on external harms as well as his predominantly negative attitudes to war. War is an evil, but its fundamental status is instrumental. Augustine argues a bit later (1950:687):

It is therefore with the desire for peace that wars are waged even by those who take pleasure in exercising their warlike nature in command and battle. And hence it is obvious that peace is the end sought for by war. For every man seeks peace by waging war, but no man seeks war by making peace.

The priority of peace to war in the order of final causes or goals is a point that had already been seen by Aristotle (1925: 1177b9-12). But Augustine sees more explicitly the ways in which the type of peace is correlated with the society that pursues it, e.g., a society of robbers, a family, a monarchy, just or unjust. Peace is a natural goal, even if corrupted nature seeks an unjust peace (Augustine, 1950:687-690).

But this claim has a double edge to it, for while it maintains that peace is universally desired it partly undercuts the claim that all forms of peace are good. Thus Augustine observes (1950:695): "The families which do not live by faith seek their peace in the earthly advantages of this life." These advantages are not closed or forbidden to people of faith, for Augustine argues (1950:695):

Thus the things necessary for this mortal life are used by both kinds of men and families alike, but each has its own peculiar and widely different aim in using them. The earthly city, which does not live by faith seeks an earthly peace, and the end it proposes, in the well-ordered concord of civic obedience and rule, is the combination of men's wills to attain the things which are helpful to this life. The heavenly city, or rather the part of it which sojourns on earth and lives by faith makes use of this peace only because it must, until this mortal condition which necessitates it shall pass away.

The peace which we enjoy under the conditions of this life is, in Augustine's terms (1950:707-08) "rather the solace of our misery than the positive enjoyment of felicity." Augustine then points to the final misery of those separated from God and not belonging to the city of God as a state of war. In the present order of things, both the justice or righteousness of which we are capable

and the peace we can attain are flawed and imperfect. But in the final peace, where we shall be freed from both sin and death, Augustine says (1950: 708), "it will not be necessary that reason should rule vices which no longer exist but God shall rule the man, and the soul shall rule the body, with a sweetness and facility suitable to the felicity of a life which is done with bondage." It is this peace which Augustine regards as the supreme good, but it is a peace which takes us beyond the circumstances of justice, the limited resources and limited benevolence of our present earthly condition (Rawls, 1971: 126–130). The peace of the earthly city is appropriate to the circumstances of justice and to the need of a sinful and divided humanity to maintain a "common agreement" about the "necessaries of life" (Augustine, 1950: 696).

The common element in all the various forms of peace that Augustine recognizes in nature, in the body, the soul, the individual, the society, and the great cities, is the "tranquillity of order," that order which "allots things, equal and unequal, each to its own place" and which in the heavenly city is "the perfectly ordered and harmonious enjoyment of God, and of one another in God" (Augustine, 1950: 690). We may feel tempted to visualize this ultimate vision of peace on the lines, say, of Van Eyck's *Adoration of the Lamb*. But we should remember that it includes a very important punitive element. The damned share in the final peace, since they are bound with each and separated from the blessed by "the law of order" and since their continued existence requires some minimal harmony with the natural order of things. Augustine explains this (1950: 690): "There cannot be war without some kind of peace, because war supposes the existence of some natures to wage it, and these natures cannot exist without peace of one kind or another." This position has the attraction of making peace more fundamental than war, but of broadening the concept so that it includes war. What is of more immediate interest for us is that it shows the compatibility of a punitive conception of war with his overall conception of peace.

This gives us the last essential element in Augustine's thinking about war that we mentioned earlier, namely, his conception of peace. This conception is marked by a fundamental distinction between the ultimate peace of the heavenly city, which has an absolute value but is not directly attainable, and the partial, temporary, and imperfect peace available to sinful humanity. These two forms of peace are analogically related, which is to say that they are alike in some ways and different in others. For anyone who, like Augustine, accepts the absolute value of heavenly peace, a great deal will depend on whether one stresses the similarities between the two forms of peace or the differences, for instance whether one sees earthly peace as preparation for, prefigurement of, and participation in heavenly peace or whether one sees it as marred by egoism and disfigured by structures of oppression. Augustine himself, one may safely say, puts more stress on the similarity between earthly peace and heavenly peace and sees earthly peace as making an important instrumental contribution to the well-being and progress of those called to heavenly peace.

A great deal also depends on the gradations of similarity to the ultimate peace one is prepared to recognize among the various forms of earthly peace that may be available as alternatives. The morally significant differentiating factor in various forms of earthly peace is justice, since the order which Augustine sees as the fundamental element in peace is an order of justice which assigns to each thing or person its due. A question that naturally occurs to us as heirs and critics of the revolutionary impulse in modern societies is the extent to which it is possible to fashion a new order of society, a new city, that will be appreciably more just and hence closer to the order of heavenly peace. Augustine himself was certainly aware of significant differences among political regimes, but he tended to present these in terms of personal moral differences and to accept even negative situations with equanimity. Thus in Book IV of *The City of God*, after extolling the benefits of the rule of good men who worship God rightly, he says (1950: 112): "But the dominion of bad men is hurtful chiefly to themselves who rule, for they destroy their own souls by greater license in wickedness, while those who are put under them in service are not hurt except by their own iniquity. For to the just all the evils imposed on them by unjust rulers are not the punishment of crime, but the test of virtue." It would be hard to imagine a line of thought which, while affirming the existence and the extent of injustices in society, does more to cut the nerve of revolutionary change. Augustine's pastoral concern is not to urge Christians to join in fashioning a more just order here with a correspondingly better peace, except insofar as the republic to which they belong rightly worships the true God and so possesses true virtues. (Augustine, 1950: 706). Rather, he wants to keep Christians moving on to the peace of the heavenly city and to prevent them from placing their felicity in this life. But the peace of the heavenly city is attained not by a series of social experiments and approximations, but by the eschatological events of death, resurrection, and judgment. The possibility of different forms of order and of peace which might differ in moral quality and which might then provide justification for wars of revolutionary change and for violent movements aiming at a more just society is not one that Augustine himself wishes to explore. To do so, he would need to adopt a much broader conception of the range and moral significance of political choices open to Christians and citizens than would have seemed practicable to one contemplating the pressures of barbarian invasion and the attractions of religious vocation.

IV. CONSCIENTIOUS OBJECTION AND NONCOMBATANT IMMUNITY

As the search for human and especially for divine authorization makes clear, Augustine does not envisage the citizen as confronted with the neces-

sity of judging the justice of actual or possible social systems, or of judging the justice of particular wars undertaken by the rulers. The Christian, as, for instance, the martyr, can be confronted with a demand to do what is clearly evil, a demand which he or she must not obey; but this confrontation still leaves the Christian passive in the face of authority. Herbert Deane has put the matter thus in his valuable study of *The Political and Social Ideas of St. Augustine* (1963: 149):

Even when we refuse to obey the orders of the temporal ruler because they clearly conflict with God's commands (e.g., his prohibition against the worship of idols), we have no right to *resist* the state's commands or to *rebel* against the constituted authority. Our only recourse is to follow the example of the holy martyrs, that is, to refuse to obey the ruler's sacrilegious commands and to accept quietly, without resistance, even joyously, whatever punishment he may impose upon us for our failure to obey. Even death is to be accepted without any effort to resist or subvert the political authority.

Augustine preaches a doctrine of passive obedience, which exceeds the requirements of Hobbes for obedience to the sovereign, since it rules out the right of self-defense. In the context of contemporary controversy, it is interesting that Augustine himself in the *Contra Faustum* (1887: 22.74) applies the famous text about the tribute money (Mt 22.21) to providing support for the military. When we look at Augustine's views on the just war in the context of his social and political thought as a whole, we find that the anthropological, political, and moral bases necessary to sustain a doctrine of civil disobedience and of selective conscientious objection are either absent or inadequate.

Another element that we expect to find in just war theory that is not prominent in Augustine is the doctrine of noncombatant immunity. In fact, Richard Hartigan has argued (1968: 203) that "Augustine presented no clear-cut argument for the protection of the innocent, especially for the civilian innocent or noncombatant, in time of war." Hartigan believes that this is not merely a gap but is an absence that indicates an "inherent inconsistency" in Augustine's position. He points this out in the context of what I term the punitive model of war (1968: 202):

If there are innocents, as it is obvious there are, how can war be justly waged against them, since in the case of soldiers, who are merely doing their duty, and of noncombatants, who are not actively engaged in prosecuting the unjust war, there has certainly been no volitional commitment to an injustice which alone would qualify them as guilty and as such proper objects of retribution.

Hartigan offers two possible rationales for Augustine's neglect of noncombatant immunity. One of these, that for Augustine it is better to suffer death, a mere physical evil, than to be guilty of moral evil or vice is certainly right in its claim about Augustine's hierarchy of values. But while this may show

that it is better for the victims to suffer death than to do morally evil things, it does not show why the executioners can be justified in depriving the victims even of the lesser good of life. The other explanation is that Augustine's belief in the "extremely intimate relationship between individual and social morality means that an unjust nation will not be characterized by the presence of a just citizenry" (Hartigan, 1968: 202). This claim unfortunately flies in the face of Augustine's general account of the coexistence and overlapping of the city of God and the earthly city during the present order of things. It has some interest in that it suggests a quite different way of dealing with enemy populations, which may in some cases have many guilty noncombatants who share in the benefits of and offer approval to a morally evil regime (e.g., judges or party officials in Nazi Germany). The killing of guilty noncombatants rather than interiorly innocent combatants fits with a conception of war that is punitive rather than defensive in nature.

I would like to offer two alternative explanations for Augustine's neglect of noncombatant immunity. The first comes from his argument with Faustus over the wars of Moses. If one accepts Augustine's belief in the divine authorization for the initiation and conduct of the wars of Moses and Joshua, it is very difficult to build in a notion of noncombatant immunity as an essential element in the carrying on of a just war. (See Deuteronomy 20.16; Joshua 6.17.) If the justice of God does not require sparing the lives of noncombatants, why should Augustine be more stringent? Second, it is possible to see Augustine's concern to minimize the evils of war (which he clearly and emphatically denounces and laments) as finding expression in an ethics of virtues rather than in an ethics of rules or principles (of which noncombatant immunity would be one). Augustine's repeated warnings against being motivated by revenge and his desire that war not be based on human passion point the way to a limitation of the horrors of war by focusing on the virtues and attitudes of warriors. This does not eliminate the need for a rule of noncombatant immunity, but it is reasonable to think that care to develop a regard for certain moral virtues and values among soldiers would contribute in an important and perhaps decisive way to the preservation of many of the values which the rules of *jus in bello* are designed to protect. Even in an ethics of principles, care for the cultivation of military virtue can be seen to be particularly important when we consider the numerous ways in which combat situations challenge the personal qualities of warriors. Still the gap remains, and we should not be surprised to find that one of the important ways forward from Augustine's form of just war theory is to develop principles about classes of people against whom violence should not be employed.

Reference to the elements of Augustine's position may be helpful in giving a very brief indication of Aquinas's position. Aquinas's main treatment of war is found in Question 40 of the *Secunda Secundae* in the *Summa Theologiae*, where it is one of a number of vices opposed to charity. In the first

article, Aquinas lays down three conditions for a war to be just. They are the authority of the sovereign; a just cause which involves *culpa* or fault in the opposing side; and a right intention in those waging war, namely to promote good and avoid evil. In this first article, Augustine is quoted no less than eight times. Of the elements of Augustine's theory *a*, *b*, *c*, *e*, *f*, *g*, and *h* are all present in some measure. The second article, which deals with the question of clerical participation in warfare, argues that fighting is incompatible with contemplation of divine things and "with the duties of a bishop or a cleric" but that in a complex society fighting is an appropriate and necessary function which ranks below contemplation but above business (Aquinas, 1972: 87). Aquinas goes on to observe (1972: 89) that fighting in a just war is not a sin and that "physical wars should be considered by Christian people as directed towards a divine spiritual good as their end." In this second article, Aquinas is moving to make closer connections between the activities of our social world and the attainment of divine and spiritual goods and is also developing a more differentiated and active concept of order in society than we have seen in Augustine. Aquinas, in accordance with his general systematic position, modifies elements *d* and *h* in Augustine's position. This can also be seen in the fourth article of this question, on whether wars may be fought on feast days. Here Aquinas speaks of our need to preserve the safety of the state (*salus reipublicae*), which is of more worth than the bodily well-being of the individual person, since by it "many more people are saved from slaughter and numberless temporal and spiritual evils are prevented" (Aquinas, 1972: 93).

V. CONTEMPORARY REFLECTIONS

What general conclusions of contemporary interest can we draw from this overview of the major elements of Augustine's just war theory? To what extent is the theory available for our use and enlightening for our choices? To some extent, of course, this will depend on our general assessment of Augustine's theology. Here I can do no more than offer some suggestions bearing on the eight basic elements that I have listed.

A. The punitive conception of war as the restoration of moral order rather than the defense of vital national interest seems too broad and difficult to apply, both because there are so many instances of moral disorder in the world and because the forms of interior evil with which Augustine is preoccupied cannot be satisfactorily detected or effectively corrected by the ordinary measures of military policy. These difficulties are much stronger when one accepts element *b*, the account of the evil of war primarily in terms of interior attitudes. The punitive conception of war can be proposed without the general Augustinian framework, and also without denying the defensive conception. In fact, there have been, in a number of public justifications of the

British position in the Falklands dispute, strong signs of a punitive conception. After all, what is being defended in this case is peripheral to the survival of the national community that is making war. The focus is rather on protecting and vindicating certain rules of international order and the rights they affirm. It is also interesting to recall that the encyclical of John XXIII, *Pacem in terris*, challenges the punitive conception of war, at least nuclear war (1963:30): "Therefore, in this age of ours, which prides itself on its atomic power, it is irrational to think that war is a proper way to obtain justice for violated rights."

B. Augustine's insistence on seeing the evil of war primarily in terms of certain attitudes and desires rather than in terms of damage to premoral interests and values is an expression of the concern of his ethical thought with intention and virtue, but it is not a persuasive emphasis today. The special concern aroused by modern weapons of mass destruction is over the vast increase in the amount of harm that they can do to human societies, their members, and their environment. There is also a particular concern over the impersonal character of warfare carried on by advanced technological means the use of which may be compatible with detached and even indifferent attitudes on the part of the warrior-technicians who actually direct and use the weapons. Theirs may well be an evil frame of mind (perhaps of a banal sort), but it is clearly different from the lust for revenge and the craving for power that so troubled Augustine.

Item *c* should be considered with *f*, since they both involve a denial of the active role of the responsible citizen in shaping defense policy and in making decisions about the use of force. Some modern pacifists might see in this denial evidence to support the claim that Christians should avoid the use of violence. Without rejecting this line of reflection altogether (since I think an aversion to violence is a humane and Christian attitude), I would point out that Augustine's position attempts to export responsibility for the violence onto the accounts of higher powers (human or divine) and that he does this in a way that does not conform to an awareness of "man come of age" and that is at odds with the conceptions of human agency and moral responsibility that are found in contemporary western cultures and in various liberation movements. This is one issue where one could line up Jefferson, Kant, Freud, and Marx along with Bonhoeffer, the Niebuhrs, Vatican II, and Gutierrez in united opposition to Augustine. The turn to authority and the expectation of individual passivity can be used both for and against violence, for and against a particular order of things. But one of their main uses in military contexts is to create an overwhelming presumption against selective conscientious objection, a possibility which has to be kept open by any form of just war theory that is to be credible and applicable in an open society.

D. The divided epistemological stance which both affirms the clear superiority of spiritual goods and professes a certain agnosticism about whether

certain worldly goods are really worthwhile or worth fighting for has figured prominently in traditional appeals to accept the providential ordering of things even when they go badly for us in worldly terms. "We have not a lasting city." "All things work together unto good for those who love God." Taken by itself, this line of reflection can form part of a pacifist acceptance of earthly evils, or it can serve (as in Augustine's position) to discredit the use of violence to defend one's interests and earthly values (including one's life). This is a familiar and important part of the Christian's dealing with the world in a spirit of trust and is an important aspect of conversion from the pursuit of worldly goods to the spiritual values of charity and union with God. But it encounters two major difficulties in contemporary discussion. The first is political and in one sense peripheral from a Christian point of view. It is that national decisions in a country such as ours cannot for a number of reasons, both practical and theoretical, be based on Christian conversion or on appeals to transcendent religious values. The second difficulty has to do with the sharp dualism between spiritual and earthly in being, in knowing, and in valuing which this approach expresses. This approach is widely regarded as wrongheaded in a number of different areas of life (the meaning of sexuality, the effects of poverty and material deprivation, personality development, and cognitive theory, to name a few). In a more explicitly theological response, one can argue that this spiritual line of reflection causes us to misunderstand our nature as incarnate beings that have to realize spiritual values in a material environment of which we ourselves are part and that Augustine's dualism is unfaithful to the Biblical anthropology and relies on Greek categories of soul and body. The turn away from dualism is sufficiently pervasive in our culture and in our theology that dualistic appeals to spiritual values are likely to encounter resistance from intellectual criticism as well as from weakness of will.

E. Augustine's interpretation of evangelical norms in terms of interior attitudes rather than overt actions has the attraction of avoiding a constricting literalism but threatens to sever the connection between virtuous attitude and right action. While no important virtuous disposition can be identified with a single type of action or even with a list of types of actions, virtues and interior dispositions, if really possessed and effective, must make some difference in the conduct of one's life and in one's public behavior. Giving up this connection is both a misunderstanding of our psychology and an invitation to self-deception. Part of Augustine's difficulty here is that he is reluctant to say that turning the other cheek may in some cases be a wrong or inappropriate action, even when this would be the clearer and simpler statement of his view. Another part of the problem is the ambivalence that is inherent in any Christian acceptance of justified violence. Even when a Christian judges in good conscience that recourse to violence is necessary and justifiable, the presence of conflicting attitudes for and against the violence

is appropriate. Such conflicts can be quite painful, and we have a variety of ways of handling or evading them. But to act on one attitude (whether it leads toward or away from violence) does not destroy the contrary attitude. This is a valuable point in Augustine's contrast between attitude and action. Just how close a concord there can be between the norms of the Sermon on the Mount and the actions of Christians remains problematic, even when we agree that there should be conformity between evangelical norms and our inner attitudes as well as between our inner attitudes and our outer actions. It is characteristic of many contemporary Christian pacifists to desire that this conformity be both close and transparent; this desire is manifest in their concern for the symbolic and witness value of what they do and in the content of their appeals to their fellow Christians.

Item *f* has already been considered in connection with *c*. It is a point that divides Augustine's position from activist and participatory forms of political theology currently dominant in the Third World and broadly influential in the First World.

G. Augustine's use of New Testament texts to legitimate military service shows a certain literalist ingenuity and a tendency to rely on inferences from particular texts (such as Jesus' praise of the centurion) rather than to argue the issues in terms of a Biblical perspective which, in the strategy of modern scholarship, is historically recoverable in part, demonstrably beyond recovery in part, and inconclusively but suggestively normative for us. The biblical issue on which Augustine may have the most to offer us is the defense of the Hebrew Bible against Manichean discrediting. A crucial point with regard to war (and a number of other social institutions and practices) is the question of the weight to be given to normative positions assumed or defended in the Old Testament.

H. Augustine's analogical conception of peace as order, which we have already considered earlier, presents both possibilities for justifying the use of force as a means of maintaining order and preventing chaos (recognized by Augustine) and as an instrument for bringing about new forms of order (generally not recognized or approved by Augustine, except when he touches on the topic of religious coercion).

This contemporary overview of the elements that constitute Augustine's position on the just war may well appear sketchy and inconclusive. This partly reflects the limitations of this paper, but it is also a consequence of the ambivalence and complexity of Augustine's own position. There are a number of different ways of combining and developing these different elements. Thus, to put the matter very roughly and schematically, the various elements could be put together in support of a just war position or a holy war position or a pacifist position. For instance, emphasizing *a*, *c*, and *d* could well produce a very ambitious conception of war as divinely ordained punishment for violations of moral order. Similarly, *b*, *d*, and *f* could be combined to produce a passive and other-worldly form of pacifism.

The contemporary point of focusing on the discrete elements in Augustine's theory is not to resolve the dispute between pacifism and just war theory or to suggest new conclusions, but rather to make more explicit the range of considerations (theological, anthropological, and ethical) that Augustine and other theological and philosophical theorists of just war bring to these issues, to make more clear to ourselves some of the reasons why we find their positions satisfactory or unsatisfactory, and to suggest to ourselves that there may be less than perfect harmony among the various elements of our own position. For instance, we may be arguing for pacifism on the basis of an assessment of the evil of war which is much less spiritual than Augustine's, and of a reading of human nature which is less dualistic and more optimistic than his, and of a conception of society which is more individualistic and more participatory than his. Such a set of views will give us a different set of problems and opportunities in the argument. But our sense of these problems and opportunities and of the complex ways in which a variety of human religions and intellectual concerns come together to shape the positions that we adopt on the momentous questions of peace and war can only be deepened and enriched by the study of the history of Christian thought and practice in the making of war and the making of peace.

REFERENCES

- Aquinas, Thomas
1966 *Summa Theologiae* I-II, 90-97, tr. Thomas Gilby, O.P. New York: McGraw-Hill.
1972 *Summa Theologiae* II-II, 34-46, tr. Thomas R. Heath, O.P. New York: McGraw-Hill.
- Aristotle
1925 *Nicomachean Ethics*, tr. W.D. Ross. London: Oxford University Press.
- Augustine, St.
1887 *Contra Faustum Manichaeum*, tr. Richard Stathert. Pp. 151-365 in Philip Schaff (ed.) *The Nicene and Post-Nicene Fathers*, vol. IV. Buffalo: Christian Literature.
1950 *The City of God*, tr. Marcus Dods. New York: Modern Library.
1953 *Letters*, volume 3, tr. Sr. Wilfrid Parsons, S.N.D. New York: Fathers of the Church.
1955 *Letters*, volume 4, tr. St. Wildrid Parsons, S.N.D. New York: Fathers of the Church.
1967 *The Free Choice of the Will*, tr. Robert P. Russell, O.S.A. Washington, D.C.: Catholic University of America Press.
- Berlin, Isaiah
1969 *Four Essays on Liberty*. London: Oxford University Press.
- Deane, Herbert
1963 *The Political and Social Ideas of St. Augustine*. New York: Columbia University Press.

Hartigan, Richard Shelly

1966 "St. Augustine on War and Killing: The Problem of the Innocent." *Journal of the History of Ideas* 27 (April-June): 195-204.

Hobbes, Thomas

1947 *Leviathan*. Michael Oakeshott (ed.). Oxford: Blackwell.

John XXIII

1963 *Pacem in terris*. Washington, D.C.: National Catholic Welfare Conference.

Milhaven, John Giles

1968 "Moral Absolutes and Thomas Aquinas." Pp. 154-185 in Charles Curran (ed.), *Absolutes in Moral Theology?* Washington, D.C.: Corpus Books.

Ramsey, Paul

1961 *War and the Christian Conscience*. Durham, N.C.: Duke University Press.

Rawls, John

1971 *A Theory of Justice*. Cambridge, MA: Harvard University Press.

four hundred years moved back and forth along the dimensions of social organizations that define basic and mature natural states.

2.4 Privileges, Rights, and Elite Dynamics

The example of Republican and Imperial Rome provides clear evidence that no teleology pushes societies to move along the progression from simpler to more complex natural states. No forces inevitably move societies along the continuum from fragile to basic to mature natural states. Societies appear as capable of regression as progression.²⁰

The logic of the natural state offers some insight into why the forces operating on limited access societies do not produce an inevitable progression from simpler to more complex forms of social organization. The forces of double balance are an important element of development within the natural state. In order for the public organizations of the state to become more sophisticated, private-sector organizations must develop. The development of public and private organizations must proceed together, including the degree of specialization and division of labor within and among organizations, the ability to own and transfer real estate and movable wealth, the ability to monitor and record the activities of the organization, the degree to which external third-party enforcement of agreements is used and, in large enough organizations, the ability to provide internal third-party enforcement.

As basic natural states emerge from fragile natural states, support for organizations is vested in powerful members of the dominant coalition. Elites have strong interests to promote trade and specialization and division of labor. They also have strong interests to define their privileges with respect to one another as elites. Initially these privileges are fluid, the result of the dynamics of the coalition. However, if a coalition stabilizes, the creation of public law institutions can in part be a mechanism for resolving elite conflicts. Our third proposition or prediction is that:

The origin of legal systems lies in the definition of elite privileges.

Legal systems initially develop to enforce unique and personal elite privileges, including the privilege to form organizations. The organizations

²⁰ Over the last decade, for example, Bolivia, Venezuela, and Russia all seem to be regressing as they nationalize, control, or outlaw once-independent organizations. In the 1930s, Nazi Germany regressed from a mature to a basic natural state as it forced previously independent organizations into the orbit of the state.

formed by elites often contain elements of both public and private organizations; for example, the manor in medieval England or the town in medieval Germany. Early legal forms can be credible among elites because they individually possess the power and incentives to discipline each other, through extralegal means if necessary. The existence of a court or laws for elites does not imply rule of law or unbiased enforcement. Powerful individuals will probably receive more favorable treatment (as reflected in the Salic law). If, however, a privilege is held by all elites, then not only do the incentives of all elites align to maintain that privilege but also the shared nature of a privilege makes it possible to transform the privilege into a right. For example, in medieval England, as we will see in Chapter 3, major lords all possessed the right to hold their own manorial courts. Their rights created a shared interest in maintaining those courts (sometimes in opposition to or in competition with the king's courts).

As elites form more powerful organizations, it becomes more credible for elites to believe that arrangements defined by the internal rules of the dominant coalition will be followed. More powerful elite organizations provide elites with both the ability to discipline the coalition and incentives to support the shared institutions that support their organizations. An important consequence follows when private elite organizations have some degree of independence from the state and freedom of action, even if it is limited. When elite organizations have an enhanced ability to discipline (or reward) the state for violating (honoring) its commitments, then natural states are better able to credibly commit to more sophisticated public organizations as well. This is the virtuous side of the logic of the natural state. If and when limited elite rights to form organizations emerge, basic natural states move closer to mature natural states. Mature natural states cannot develop without more sophisticated private organizations.

The process of converting elite privileges into elite rights secured by a balance of political and economic interests does not happen automatically, however, because there is an equally powerful and persistent reason for elites to convert rights into privileges. The dominant coalition holds together only if the balance of economic and political interests can be maintained. If circumstances change, and they always do, there may be incentives to shift resources or privileges to elements within the coalition in order to maintain balance. It may be difficult, if not impossible, however, to make small marginal changes to arrangements. If a faction within the coalition becomes more powerful, it maneuvers to get more rents or it threatens to fight or leaves (and perhaps returns to fight). Intra-elite relationships are always more or less dynamic, and are never static. As natural states

NOTICE:
This material may be protected by copyright law (Title 17 U.S. Code)

develop, they do not reduce the uncertainty or dynamism of the dominant coalition. Instead, they manage to secure more sophisticated public and private elite organizations that sustain a wider range of adjustments to changing circumstances that do not require transforming elite rights back into elite privileges. When such dramatic adjustments are required, natural states often suffer partial or complete breakdowns in the dominant coalition, and civil war, rather than legal adjustments, can be the result. In those circumstances, mature natural states can quickly move back toward social arrangements typical of a basic or even a fragile natural state.

2.5 Origins: The Problem Scale and Violence

The discovery of agriculture and the domestication of plants and animals began the Neolithic revolution ten thousand years ago. The growth of cities, new production technologies, and new forms of social organization transformed human societies over the following five thousand years. Archeological evidence reveals the emergence of groups of larger than several hundred people for the first time in human history. Regardless of the causes of the Neolithic revolution—climate change, genetic change, the discovery of agriculture, or a new social technology—we need to understand how societies managed to become substantially larger beginning five to ten thousand years ago.

Larger societies required new ways to manage and control violence. Basic social units in foraging orders (bands or family groups) were typically groups of twenty-five to fifty individuals. Larger social units (tribes or local groups or big-men collectivities) range up to five hundred people. For each of these ways of organizing social interaction, increasing size produces increasing intragroup conflict and what Rappaport called the “irritation coefficient”: “sources of irritation . . . increase at a rate greater than population size. If population increase were taken to be linear, the increase of some kinds of dispute . . . might be taken to be roughly geometric” (1968, p. 116). The increasing violence and disorder with group size, *ceteris paribus*, conform to the generally observed positive correlation between the level of violence and size of population in modern societies.²¹

The numbers twenty-five and two hundred recur in anthropological research. The modal size of the basic foraging social group appears to be

²¹ Johnson’s (1982) notion of scalar stress suggests that human cognitive limits begin to be reached when six or seven items have to be attended to simultaneously, so that six groups of six people, 36 people, or six groups of six, 216 people, might be natural sizes for groups.

around twenty-five people (Kelly, 1995, pp. 209–16). Agglomerations of bands or family groups were often around two hundred people. Service’s (1971) categories of band, tribe, chiefdom, and state correspond to groups of roughly this size, with perhaps one thousand as a modal unit for chiefdoms. Johnson and Earle (2000, p. 32) transform Service’s categories into family groups (including the family/camp and family/hamlet), local groups (both acephalous and big man), and regional groups (both chiefdoms and states). The size of these different societies is bounded at the upper range by problems of managing violence.²² Although the evidence is contested, small societies seem to experience high levels of violence (Keely, 1996; LeBlanc, 2003; Steckel and Wallis, 2006).

Although new archeological evidence is accumulating, evidence about what happened five or ten thousand years ago is too limited to make inferences about social organization in the Neolithic societies. However, three sets of available evidence are relevant to the question of social organization and scale. Ethnographic studies of small-scale societies in anthropological research abound and suggest similarities in organization of societies that increase in size from local to regional polities, or from tribes to chiefdoms.²³ Related work includes the literature on the structure of “pristine” societies, the first large-scale civilizations to arise in different parts of the world that appear to have done so without external influence (Trigger, 2003). Finally, we have evidence on the incidence of human-induced violence in New World societies based on archeological evidence of skeletal remains (Steckel and Rose, 2002).

Johnson and Earle (2000) draw together evidence on nineteen ethnographies of societies at different scales. Of particular interest is the transition between local and regional groups, or between big-man societies and chiefdoms. In big-man societies, an individual or family leads the group and enjoys more wealth but is subject to considerable constraint from the larger group. The big man leads by building a personal following. The big man usually possesses the key privilege (or performs the function) of controlling trade between his group and other groups.

²² Johnson and Earle (2000, p. 246) provide a table with population breakdowns across 19 of the ethnographic cases consistent with size breaks at 25/30, 200/250, and 1,000 (although the fit is not perfect, as one would expect). Service (1971) discusses size categories and types. Dunbar (1996) suggests that the optimal size of a human group, based on studies of brain size and group size in primates, is about 150. Bandy (2004) studies the fusion–fission process in Mesoamerican societies.

²³ Earle (1997, 2003), Johnson and Earle (2000).

The emergence of chiefdoms involves a more complicated form of political economy of social organization. Earle identifies three types of power – economic, military, and ideological – and argues that chiefdoms emerge when “the material flows of the political economy provide the wire that binds the sources of power together” (1997, pp. 207–8). In Earle’s words:

In chiefdoms, control over production and exchange of subsistence and wealth creates the basis for political power... Economic power is based on the ability to restrict access to key productive resources or consumptive goods... Control over exchange permits the extension of economic control over broader regions... The real significance of economic power may be that the material flows through the political economy can be used by the chief to nurture and sustain the alternative power sources... (1997, p. 7).

The coalition, which the chiefheads and that always controls trade, is distinct from the common people through some form of social identification (an important function of ideology). Violence and coercive power are inherent elements of larger social groups, but the groups are held together not only by the threat of coercion but by mutual interests (Earle, 1997, p. 106). The creation of elites requires the social construction of social personas. The emergence of chiefdoms, of societies of more than one thousand members, appears to be associated with institutions that reflect the logic of the natural state.

State is a term of art with a specific meaning in anthropology, but less so in political science and economics. States are distinguished from chiefdoms by size and structure and include formal administration of government. For anthropologists, states do not appear until populations rise into the hundreds of thousands.²⁴ In contrast, what we define as the natural state arises as societies reach populations of one thousand or more, and new forms of integrated political and economic organization develop to limit violence. As Earle recognizes, “The fundamental dynamics of chiefdoms are essentially the same as those of states, and... the origins of states is to be understood in the emergence and development of chiefdoms” (1997, p. 14). We add the logic of natural state to the approach of Johnson and Earle: the key link that constrains military power is embedding the individuals who direct military power in a network of privileges. By manipulating privilege, interests are created that limit violence.

The emergence of social hierarchy in early societies results directly from the creation of privileged elites. Ancient civilizations do not afford us a

²⁴ For recent investigations into the archaeology and anthropology of pre-modern states see Smith (2004), Farghar and Blanton (2007), and Blanton and Farghar (2008).

direct view of early developments at the dawn of the Neolithic revolution, but they do give us a clear picture of early natural states. All these societies were theocracies.²⁵ All had interlocking sets of religious, economic, political, military, and educational elites; indeed, in many of the highest leaders were simultaneously priests, warriors, and kings or princes. As Service (1975) argued, any state that rules by coercing subjects and rivals must continually risk war against both its own subjects and its rivals. Successful societies “wage peace,” in Service’s evocative phrase. Securing peace prepares the ground for beliefs to grow up in the population at large about the legitimacy of the system. Those beliefs are consistent with positive incentives for powerful individuals to maintain the peace, in contrast to societies where a balance of terror is all that ensures order. Ideology, both in material culture and in the religious and educational organizations of elites, interlocks with the logic of the social order.

The implication is that chiefdoms and states exhibit the characteristic organization of natural states, and they should therefore experience lower levels of violence than the foraging order. Using recently developed forensic techniques for inferring the existence of human-induced violence from skeletal evidence, Steckel and Wallis (2006) show that the rate of human-induced violence in a sample of New World individuals declined as the size of the population increased. Individuals living in small foraging groups had significantly higher rates of human-induced trauma. We summarize the results in the appendix to this chapter.

The anthropological evidence suggests that the increasing scale of human societies is associated with the emergence of social organizations implied by the logic of the natural state. The political economy of chiefdoms expresses the logic of the natural state. All of the pristine ancient civilizations were societies with strong theocratic hierarchies in which limited access to economic, political, military, and religious functions played a key role in identifying the social persona of elites. The evidence from skeletal remains suggests that, as the scale of societies increased, human-induced violence declined.

²⁵ Service includes Mesopotamia, Egypt, the Indus River, China, Mesoamerica, and Peru. Trigger excludes the Indus River civilization, and includes the Aztecs and Incas of the fifteenth and sixteenth centuries and the Yoruba and Benin peoples of West Africa in the eighteenth century. Trigger does not require civilizations to be “pristine” in the sense of being the first civilization to develop in its part of the world (2003, pp. 28–9). Freid (1967), Feinman and Marcus (1998), and Yoffee (2005) offer alternative frameworks for interpreting early civilizations and more recent anthropological work on the origins of “archaic states.”

2.6 Natural State Dynamics: Fragile to Basic Natural States

How do fragile natural states move toward basic natural states? How does a society whose major accomplishment is holding the dominant coalition together become transformed into one where the state establishes durable institutions and under the spreading umbrella of state support develops economic, educational, and religious organizations that sustain the integration of the society into a larger geographic and cultural entity?

Two historical examples illustrate how a basic natural state forms out of a fragile natural state: the Mesoamerican Aztec Empire (1428–1519) and the European Carolingian Empire (751–840). Each empire was founded in a society living in the shadow of historical greatness, where disintegration and decline had occurred over several centuries. Each empire experienced a quick rise to power, yet lasted a relatively short time. Their key similarities relate to the way in which the dominant coalitions structured their internal relationships and how they used military, economic, political, religious, and educational institutions and organizations. Each empire built on existing social materials, and each borrowed extensively from preexisting institutions, making small modifications in them that dramatically changed the way the social order performed. Finally, in each empire the state was the only framework within which organizations could be supported. These were truly basic natural states.

The Aztec Empire. Mesoamerica had seen three large civilizations rise and fall: the coastal Olmec from about 1500 to 300 B.C.E.; Teotihuacan in the Valley of Mexico from about 100 to 750 C.E.; and the loosely organized Toltec society from roughly 900 to 1200 C.E. Each empire was followed by a period of disintegration. In the thirteenth and fourteenth centuries, the Valley of Mexico was composed of competing city-states and fragile alliances.²⁶ One of many migrant ethnic bands, the Mexica, settled on the island of Tenochtitlan in Lake Texcoco. The lands around the lake were fertile, and the area was a center of military and political conflict over control of the region. In the late fourteenth century, the most powerful local alliance was headed by the Tepanecs, which included the Mexicas in their alliance.

²⁶ The large literature on Aztec society is continually growing and changing. We have drawn on Smith (2001, 2003), Smith and Montiel (2001), Lockhart (1992) Carrasco (1999), Conrad and Demarest (1984), Berdan (1982), Berdan et al. (1996), Blanton (1996), and Brumfiel (1987), Hassig (1988), among others.

The basic units of Mexica society were the *altepetl* and the *calpulli*. The *altepetl* were ethnic states that were primarily organizations of people controlling a given territory. The *calpulli* were the constituent parts of the *altepetl*, whose land was owned by nobles who granted use of the land to *calpulli* members. The *calpulli* served as the unit of organization for schools, temples, and military units.²⁷ The Mexica aristocracies, the *pipiltin*, were the first families of the *calpulli*, and higher political leaders, the *tlatoani* (ruler) were drawn from the *calpulli* leadership. Integration into the Tepanec alliance gave the Mexicas an opportunity to develop their own elites, to serve in military campaigns, and to acquire administrative and military skills. The Mexicas petitioned the city of Culhuacan and the Tepanec alliance to provide them with a prince of Toltec blood. The first *tlatoani*, Acamapichtli, came to Tenochtitlan around 1370. Whether he came because of a request or was imposed on the Mexicas by the Tepanecs is not clear.

The death of a powerful Tepanec leader in 1426 was followed by a period of confusion. In the power void, a coalition of three city-states – the Triple Alliance of Tenochtitlan, Texcoco, and Tacuba – defeated the Tepanecs in 1428. The Triple Alliance became the Aztec Empire (Smith, 2001). The Triple Alliance was always a coalition. Despite the relative ascendancy of the Mexica Aztecs within the coalition, it remained an alliance until the Spanish arrived in 1521. The southern valley was divided into three regional states, each governed by one part of the coalition. The alliance turned its military forces to the conquest of other city-states in the southern part of the valley by the 1430s.

The Aztec leaders immediately began distributing land from conquered areas to leading elites, particularly military leaders.²⁸ The land went to the *tlatoani*, the military warriors, and the principal members of the *pipiltin* aristocracy. A new leadership group was formed called the Council of Four, which consisted of all members of the Imperial family, who exercised executive power within the state and determined succession upon the death of the emperor. Religious and government offices were restructured to reflect the new order. A tight relationship among wealth, conquest, and land developed in which a small dominant coalition mobilized the economic resources of the society to provide both social order and the wherewithal to continue a program of conquest.

²⁷ Lockhart (1992), pp. 14–20; Conrad and Demarest (1984), p. 23; Zorita (1963), pp. 105–11 on land tenure and pp. 135–51 on schools; Carrasco (1999), pp. 16–21 on *calpulli*.

²⁸ Hodge (1996) describes the Aztec methods of governance.

Three features of the Aztec system are of interest. The first is the use of religion. Aztec religious beliefs were rooted in the larger Mesoamerican religious tradition. Aztecs held that the sun god daily battled his way across the sky against his brothers the stars and his sister the moon. In order to triumph in this battle, the sun needed to be strong and vigorous, for which he required the *chalchihuatl*, the precious liquid, the blood of man (Caso, 1958, pp. 12–13). Failure to nourish the sun would result in the destruction of the earth. The expanding use of human sacrifice fueled the expansionary military goals of the Aztecs, and it provided an ideological framework that justified the labor of the masses even if they did not participate fully in the gains (the earth would be destroyed): the religious structure permeated the way in which the dominant coalition defined itself. The last feature is critical. Because changes in status within the Aztec society were associated with ritual sacrifices, the religious structure identified who was powerful. The religious structure held the same position for the *pipiltin* aristocracy and for the merchant class (which we discuss in a moment).

The second feature is that access to the political-religious structure required access to education. The Aztec *calpulli* traditionally included a temple and a school, the *tepochcalli*, where male youths were taught the skills that would make them warriors. A parallel system of schools for the elites, the *calmecac*, instilled the same warrior skills and values, but also taught the students how to read “sacred books, the arts of painting, of counting the days and years, of placing the festivals of the gods, of chanting the hymns, and of understanding the demands made by the supernaturals” (Leon-Portilla, 1963, pp. 134–52). Education in a *calmecac* was an essential prerequisite for admission to the higher reaches of elite status. The combination of religion, education, and politics created a dominant coalition with a strong identity, clear shared beliefs with strong educational roots, and limited access.

The third feature is that the Aztecs limited access to economic activity by empowering a specific group of merchants, the *pochteca*. The *pochteca* possessed exclusive privileges to trade in long-distance and luxury goods, organized in guilds, with their own elaborate social ranks signified by special privileges in dress and behavior.²⁹ The Aztec Empire’s fiscal system ran on a

²⁹ Conrad and Demarest (1984), pp. 50–1; Hodge (1996), pp. 43–4; Brumfiel (1987); and Blanton (1996), pp. 47–84. Blanton’s conclusions, pp. 83–4, are particularly relevant to our argument. For the role of *pochteca* in the outlying provinces of the empire see Berdan (1996), pp. 115–35 and more generally Berdan’s analysis of the role of trade in the Aztec system in Berdan (1982, 1985, 1987) and Smith (2003).

combination of tribute and market taxation. Local markets for agricultural products were under the control of local elites, but control of long-distance movement of goods was either directly under the control of the state in the form of tribute payments or indirectly under the control of the state through *pochteca* networks and formal regulation of markets for long-distance and luxury goods (Hodge, 1996). The Aztec Empire possessed a market economy, but the state harnessed the market economy to the needs of the political system.

The Aztec Empire created sophisticated organizational structures in politics, trade, religion, education, and the military. Reflecting the essence of a basic natural state, none of these organizations existed outside of the state.

The Carolingian Empire. Carolingian history, institutions, organizations, and culture differed from that of the Mexica. Nonetheless, striking similarities exist between the two cases. Both the Carolingians and the Aztecs harnessed the productive power of organizations. As with the Aztecs, the Carolingians built organizations only within the framework of the state. Both were basic natural states.

The decline of the Roman Empire in the west was associated with a series of barbarian migrations beginning with the Goths in 376 and continuing through 476 when the last emperor fell. Relatively small barbarian populations, in comparison with the Roman population, moved into and eventually controlled strategic areas of the empire (Heather, 2006). By the eighth century, barbarian kingdoms occupied most of Western Europe, with a few vestiges of Roman authority in Italy in the Papal States and in the Exarch at Ravenna.³⁰ The Catholic church remained an important influence, both in Italy and throughout the parts of Western Europe that had converted to Christianity.

The chaotic and ever-changing political map of Europe from 476 to 750 tracked kingdoms of Visigoths, Vandals, Franks, Lombards, Saxons, and Frisians; the invasions of Moslems; and the political machinations of the eastern Roman Empire (Geary, 1988). The Kingdom of the Franks was just one of several kingdoms when Charles Martel, then Mayor of the Palace, subdued his competitors for power within the Franks, conquered Frisia, and repulsed the Saracen invasion of Provence at Tours in 732. Charles died in 741, leaving the state between his two sons Carloman and Pippin. When Carloman retired to a monastery in 747, Pippin III took unified command of the Kingdom of the Franks.

³⁰ See Hodges and Whitehouse (1983), Hodges (1989), McCormick (2001), Heather (1996, 2006), and Geary (1988) for a review of the evidence.

In the meantime, the pope had his hands full in Italy. The Lombards had occupied much of northern and eastern Italy, and the Lombard king was attempting to deal with rebellious Lombard dukes in Benevento and Spoleto. The Byzantine Empire was fighting the Moslem expansion throughout the eastern Mediterranean and trying to maintain a foothold in Italy in the form of the Exarch at Ravenna. The pope still nominally acknowledged the Byzantine emperor, and the emperor still nominally acknowledged the pope as the first personage in the church, but East and West continued to be split by doctrinal disputes. The pope faced (at least) a four-cornered struggle for power within Italy.

As with the Aztecs, the beginnings of the Carolingian rise to power lay in an alliance. Charles Martel had cooperated with the pope in the face of the Saracen invasions. Despite an existing Merovingian king, the pope agreed to recognize Pippin as King of the Franks in 751. Pippin cemented the alliance by coming to the aid of the pope in Italy, sending an army to deal with the Lombards who had taken an army to the walls of Rome. In 754, “the Pope solemnly renewed the consecration which Boniface had bestowed upon Pippin, and, on penalty of excommunication, forbade the Franks ever to choose a king who was not descended from Pippin” (Pirenne, 2001[1954], p. 225).

By recognizing Charlemagne (Pippin’s son) as emperor in 800, the pope decisively broke with the Eastern Empire and the Eastern emperor. By pledging themselves as protectors of the Catholic church everywhere in the world, the Carolingians gained entry to all of Christian Europe, an entry that Charlemagne made masterful use of. Through diplomacy and war, Charlemagne built a political coalition based on control of land and personal obligations. Military service and land tenure became closely linked, forming the origins of the feudal system.

Carolingian integration utilized natural state mechanisms. A coalition of military specialists was tied economically to the land through vassalage. Charlemagne granted land to powerful individuals in return for personal service and obligation (Bloch, 1961; Ganshof, 1968, pp. 50–2). Another coalition of traders was built up by promotion of long-distance trade under the protection and direction of the emperor.³¹ Trading emporia were established at Quentovic and Dorestad. Contentious relations between Charlemagne and the Danes, for example, arose in part by Charlemagne’s desire

³¹ This was nothing new. Long-distance trade was always subject to political control; indeed, given the dangers of long-distance trade it was not feasible to trade without explicit political guarantees of protection and passage. See the discussion of “Emporia” in Hodges (1989), pp. 47–65.

to control and maintain the trade with the Eastern Empire and the Abbasid Caliphate (Hodges and Whitehouse, 1983, pp. 111–22, 171).

The alliance with the church was an integral part of the Carolingian economic structure. Long-distance trade was centered in a few cities and in a number of monasteries. By mobilizing wealth through the coinage of silver and the movement of interregional goods within the empire along rivers to approved trading centers, the Carolingians generated an economic surplus at the service of the Carolingian coalition (Hodges and Whitehouse, 1983, p. 171). The evidence that trade did grow in Europe during the eighth and ninth centuries now appears overwhelming, and it was a trade controlled by the state (McCormick, 2001).

Finally, Charlemagne encouraged the development of an elite educational system, again in alliance with the church. The Carolingian literary renaissance was ecclesiastic (Ullmann, 1969). The expansion of literacy, the dissemination of books, and the development of a literary elite were concentrated almost exclusively within the church and its ecclesiastical structure. The emphasis on Latin as the written language of the empire at a time when it had ceased to be the spoken language of most of the population highlights the elite nature of the educational system: because only elites learned Latin, elites alone had access to the state or ecclesiastical administration, courts, and business.

The Carolingian Empire was a basic natural state. From our modern perspective, the alliance of church and state may appear strange and forced, and our attention is drawn first to the political, military, and economic arrangements of the Carolingians. The defining characteristic of a basic natural state, however, is an inability to support organizations outside of the state.

The Carolingian natural state coalition combined many social elements – political, military, religious, economic, and educational – into a coalition of interlocking interests that for a short time existed in one large geographic entity, the Carolingian Empire. However, even after the political entity broke into smaller pieces, this structure helped order most of Europe for the next six centuries. Despite being composed of many organizations and actors, a key distinction made in most histories of Europe is between the church and the state. As we have emphasized, in basic natural states, organizations all exist within the framework of the state, so disentangling the connections between church and state in medieval Europe illuminates how a basic natural state can be structured and opens a window into understanding how Europe eventually developed the institutions and organizations characteristic of a mature natural state.

Figgis described the key to understanding medieval society:

In the Middle Ages the Church was not a State, it was the State; the State or rather the civil authority (for a separate society was not recognised) was merely the police department of the Church. The latter took over from the Roman Empire its theory of the absolute and universal jurisdiction of the supreme authority, and developed it into the doctrine of the *plenitudo potestatis* of the Pope, who was the supreme dispenser of law, the fountain of honour, including regal honour, and the sole legitimate earthly source of power, the legal if not the actual founder of religious orders, university degrees, the supreme “judge and divider” among nations, the guardian of international right, the avenger of Christian blood. All these functions have passed elsewhere, and the theory of omnipotence, which the Popes held on the plea that any action might come under their cognizance so far as it concerned morality, has now been assumed by the State on the analogous theory that any action, religious or otherwise, so far as it becomes a matter of money, or contract, must be matter for the courts (Figgis, 1923, p. 4).

What did Figgis mean when he said that the church was the state? On one level, the actual provisions of public services that we associate with states were undertaken both by the church and by secular political organizations. The church was primarily responsible for education and social welfare (of a variety of types including poor relief, hospitals, disaster relief, and food storage). The church shared responsibility for the provision of justice with secular authorities. The ownership and management of land, both in land’s generalized legal characteristics defined through courts and in the day-to-day operation of operating political units, were also shared by the church and secular authorities. After the tenth century, the church ceded police and military power to the secular princes.

On another level, the church occupied critical nodes in the organization of political power, just as secular lords occupied critical nodes in the organization of religious authority. The Electors of the Holy Roman Emperor, for example, included the archbishops of Mainz, Trier, and Cologne. The archbishops and bishops of England were members of the House of Lords. The church constituted the first estate in France and, down to the revolution, maintained an interdependent fiscal relationship with state finance. Symmetrically, Charlemagne encouraged his feudal retainers to organize, support, and defend religious institutions. Proprietary churches under the control and support of a secular donor were sufficiently common that attempts to reform the system led to clashes between church and secular leaders, the most spectacular of which was the Investiture Crisis that we consider in detail in the next section. Archbishops and bishops were appointed by kings and lesser lords. Just as the archbishops and bishops exerted significant fiscal and political influence on the secular powers, so

too the secular powers exerted influence over the church. We cannot make much sense out of European history from Charlemagne to the Reformation without acknowledging the close interlocking and interdependent nature of ecclesiastical and secular elements of the dominant coalition.

The interdependence of secular and ecclesiastical authority created recurrent constitutional issues within the church, within the state, and in their interactions. The church needed to define and maintain a vertical structure of authority between the church in Rome and the church throughout Europe. At each horizontal level within the church, a structure of authority had to be sustained, including the pope and the College of Cardinals in Rome; at the level of bishops, their dioceses, their cathedrals, and the college of canons; and at the level of the individual parish, abbey, monastery, and convents. Because the churches and her officers were large landowners playing central roles in the allocation of privileges within local secular states, the constitutional structures at the individual levels corresponded closely to political and other secular institutions and organizations.

2.7 Moving to Mature Natural States: Disorder, Organization, and the Medieval Church

The Aztec and Carolingian histories are starting points for understanding how mature natural states develop. Both civilizations grew out of environments where multiple small social units competed with each other for power, resources, and security. It was a world where conditions changed frequently, rulers were deposed, cities despoiled, and populations displaced. This, of course, is the normal world for most of human history. Even within Aztec Mesoamerica, large parts of the population lived outside the empire and faced ongoing warfare and confusion. Similarly, although the Carolingians united part of Europe for a time, at the death of Louis the Pious in 840, the empire was permanently dismembered so that by 900 Europe faced conditions similar to those of 700. There was nothing inevitable about the increasing scale of society under the Aztecs or the Carolingians, or about the decrease in scale and increasing disorder of Europe after 840.

All natural states face the problem of maintaining their dominant coalitions by granting elite individuals and organizations economic and political incentives to cooperate with one another. In basic natural states, all organizations – political, military, economic, religious, and educational – are integrated closely into the structure of the dominant coalition. For convenience, historians and social scientists often identify the military organization as the “state,” but as Figgis argues for Europe, the church was as much

a part of the state as the military leadership. How does a society manage to develop institutional support for elite organizations outside the immediate framework of the state? Beyond that, when independent elite organizations come into existence, how is it possible to endow them with perpetual life?

The essence of a natural state is personal relationships. The legal system cannot enforce individual rights if every individual is different, if every relationship between two individuals depends uniquely on their identity within the dominant coalition. Following their Roman origins, in Europe both public and private law evolved.³² Private law structured a limited number of relationships between individuals who are recognized as persons in the law. Not all legal persons enjoyed equal protection under the law; and the law recognized more than one category of legal persons, such as kings, nobles, and freemen. Public law structured relationships among and within organizations, including the most important organization in society, the state.³³ Public law entities were typically idiosyncratic. Each public law organization possessed rights and bore unique responsibilities, consistent with the logic of the natural state.

The combination of the public law that creates and sustains organizations designated as legal persons and the private law rules about how persons relate to one another creates the possibility for a rule of law for organizations. The interaction of public law organizations with the rest of society is determined by both public law (to the extent that the organization possesses special privileges) and private law (to the extent that the organization functions as a legal person).

The close integration of church and state organizations under the Carolingians posed a fundamental structural problem. Charlemagne's coronation did not answer the question of whether the pope created the emperor or the emperor created the pope. Under the original institutions by which the Catholic church became the state church of the Roman Empire, there was no confusion. The Catholic church was grounded in the Roman law, particularly the law as codified by Justinian: *Ecclēsia vivit iure Romano*, or "the Church lives according to the Roman law" (Ullmann, 1975, p. 54). Justinian's Code opens with the law by Theodosius the Great that made the Christian religion the sole religion of the empire. In Roman law, the

³² This explanation is clearly influenced by the development of European law, and may not be applicable to other societies, but it is a starting point.

³³ These distinctions and the argument in this section are developed in Wallis, Weingast, and North (2006).

emperor was the source of law and above the law. The emperor was above the law both as an individual person and in his social persona.

In the years after the fall of the empire, the church played an increasingly central role in the provision of order. Here was the problem: the pope's legal foundation was the emperor. If the pope created an emperor and recognized him as the emperor in law, then the church's legal foundation stemmed from the emperor as the *kosmokrator* – the ruler of the world – and as the *pantokrator* – God's earthly representative. How could the pope create an emperor, when the emperor could then create a pope? Was the pope or the emperor the *pantokrator*?

Charlemagne was crowned emperor by the pope. With respect to the church, he came to the defense of the pope in his struggles with the Lombards. He promoted the construction of cathedrals and support of the monasteries. He made large grants of lands to the church and to specific churches and monasteries. He issued diplomas and charters that formally granted immunity from specific taxes and regulations to individual churches and monasteries (Ganshof, 1968, pp. 45–50). Charlemagne acted within the framework of the law accepted by him and the church. Charlemagne also encouraged other nobles to found churches and arrange for their continued support. The practice developed into the establishment of proprietary churches where the patron appointed the priest or the bishop in return for the patron's continued support.

The division of the empire in 840 and the subsequent disruption of Western Europe by another wave of Saracen and Viking incursions removed the question of papal or imperial supremacy from active consideration for a time. However, the issue did not disappear. The origin of legal systems is defining elite rights and structures, and the existing law for the alliance of church and state in medieval Europe had a fundamental contradiction at its root. After the return of peace in the eleventh century, the issue came to a head in what is known as the investiture crisis.

Gregory VII became pope in 1073. He was a Cluniac monk and a reformer dedicated to eliminating two sources of corruption within the church: simony (the sale of church offices for money) and concubinage (priests marrying or living with women) (Ullman, 1972b, pp. 129–31). As with many reformers, Pope Gregory was concerned that the church had become too intimate a part of the worldly dominant coalition. Reform of simony struck directly at the proprietary church system whereby local lords selected local bishops and priests. In 1075, Gregory declared that only the church could select bishops and priests. The proprietary system was an integral part of the basic natural state in Europe promoted by Charlemagne. Gregory's

reforms had serious implications for both the structure of the church and the larger society.

At its most visible level, the investiture crisis was about patronage and the structure of patron-client networks within the dominant coalition: Would the pope or the secular lords appoint bishops, archbishops, and other members of the ecclesiastic hierarchy? Traditional views of the crisis pit the church against the state, the pope against the Holy Roman Emperor Henry VI. However, the crisis was a broader conflict within the dominant coalition.³⁴ The conflict was officially resolved in 1222 when a more complicated procedure for selecting and investing bishops was developed in which the church nominated, but the emperor/king approved, candidates.

The fascinating reaction came not from the secular lords, but within the church. The pope declared himself not just the Bishop of Rome, but the Vicar of Christ.³⁵ When the pope asserted his right to appoint bishops and archbishops he claimed a much greater position than the first bishop among essentially equal bishops; the pope claimed direct authority over all the powerful lords of the church. Because most bishops held their lands, and thus their wealth, as vassals of secular lords, the bishops had conflicting allegiances and interests. The pope's attempt to extend a more direct control over the entire administration of the church was a direct threat to the working alliance that most bishops and archbishops had with their secular lords. Papal appointment would have significantly reduced the independence of both the bishops and the kings. Papal appointment would have restructured the entire dominant coalition within Europe.

In order to function, the church needed two interdependent constitutional structures: one for vertical relationships among the levels of the church, and another for horizontal relationships at the different levels. The horizontal constitution structured ecclesiastical dioceses and other corporate entities of the church (abbeys, monasteries, universities, and orders). The constitutional relationship between a bishop and his college, for

³⁴ It was more than a conflict between a secular power and an independent church, although the emperor and pope did battle. The pope excommunicated Henry IV, which hurt Henry with his German supporters. Henry went barefoot in the snow to petition Gregory at Canossa. The pope relented and restored Henry to the communion, but in the meantime the pope had thrown his support behind another candidate for the empire. When Henry regained his position in Germany, he again opposed the pope, who again excommunicated him.

³⁵ The title had been claimed by kings before the crisis; after all, the king was the protector of the faithful in his kingdom.

example, affected the internal workings of the cathedral and diocese, in general when and to what extent the bishop had to seek the counsel or explicit consent of the cathedral college to make specific decisions. It also affected the external relations between the diocese and the external world in terms of the freedom and responsibilities that the bishop exercised as a representative of the diocese. Finally, it concerned the process of selecting bishops when a vacancy occurred and the exercise of a bishop's functions while the office lay vacant (Tierney, 1955, pp. 106–31).

Had the investiture crisis been solely about the appointment of bishops by the pope or the emperor, only the last aspect of the horizontal constitution would have been affected. However, lay investiture also affected the economic and social status of the bishop, because the proprietary church system gave the bishop and the diocese control over land and resources that the secular lord was pledged to honor. The constitutional structure of a bishop's diocese and its relationship to the secular authority therefore came directly into the matter.

The other major constitutional issue concerned the vertical relationship among the pope, the church in Rome (including the College of Cardinals), and the archbishops and bishops: How were the parts of the church to interact? How much independent authority did Rome have over the rest of the church? This debate occurred within the church itself, but involved the entire structure of power within European society. At issue was whether the pope as an individual had discretion over the entire church, or whether there existed some group within the church who had the power to depose a pope, as in the case of a heretical pope. The argument that the supreme authority in the church rested not with the pope, but with a general council, was put forward by theorists later known as conciliarists.³⁶ In terms of identity and personality, was the social persona of the pope above or below the law? Was the social persona of the bishop above or below the corporate identity of the diocese and cathedral college?

These questions were never answered definitively in the Middle Ages, although several positions were intensely debated. What emerged from the investiture crisis was a significant innovation in the church as an organization. Important steps were taken that led toward the establishment of

³⁶ One can immediately see how debates about the constitutional structure of the church played a critical role in debates about the constitutional structure of European society as a whole, and as a result, in the political theory of the Middle Ages. See Tierney (1955), Ullman (1972b, 1975), and Kantorowicz (1997[1957]) for an introduction to this enormous literature.

self-constituting bodies within the church. The corporate identity of the diocese was strengthened, although it was never completely detached from the identity of its bishop.

The innovation that changed the corporate identity of the church was a new agreement about the ownership and alienation of church property. From the thirteenth century onward, it became general practice to acknowledge that ownership of church property did not reside with the individual person of the bishop, nor did it reside with the social persona of his office. Specifically, the bishop took an oath not to alienate the fiscal resources of the church without the consent of the cathedral college. Presaging the modern debate about ownership, control, and fiduciary responsibility, the developing canon law began to identify the interests of the church as a corporate body with respect to real property. By privileging the corporate interest in property, the body of the church, the *congregatio fidelium* – the body of the faithful (as represented by the cathedral canons) – had obtained leverage with the bishop as leader and a control over the bishop as the representative of the corporation in the wider world. Formalization of the process of consent and consultation between the bishop and the cathedral chapter created an organizational structure that gave an independent life to the corporate body.³⁷ This structure extended further into the structure of the church when popes also began including promises not to alienate church property as part of their coronation oaths.

The seeds of an elite organization independent of the state are visible in these new arrangements: a congregation or cathedral whose corporate identity was created by and associated with the property of the corporate group. In the thirteenth century, the church was not about to set individual churches free. The bishops were too critical and powerful an element in the political balance of the age, and the dominant coalition was not about to strip the bishops of their source of rents. By identifying the corporate identity of the individual cathedral with the body of property that generated the bishop's rents, however, the law began moving the organization of the

³⁷ "The inevitable starting point for such inquiries was the accepted fact that, when a prelate appeared in a court of law on behalf of his church, it was not his own possessions that he defended; he did not possess legal *dominium* over the ecclesiastical property entrusted to his protection; his status, therefore, had to be defined as that of one who represents the interests of another party. The canonists often discussed the question where actual *dominium* did reside and usually agreed that, while God himself was the ultimate owner of all the goods of the church, *dominium* over them in an earthly sense belonged to the ecclesiastical community. Hugguccio attributed it to the *congregatio fidelium*" (Tierney, 1955, p. 118).

church out of the structure of the state. The cathedral college became, in a limited but important sense, a self-constituting corporate entity. The diocese began to take on a form of legal personality independent of the grant of such personality by a pope, bishop, or king.

Just as the Carolingians and the popes cast their agreements in the classical forms of Roman law, so too did Gregory VII claim that his policies were lawful. The investiture crisis heightened the interest of Roman and canon law within the church. Moreover, Gregory's claims against the secular lords stimulated the development of civil law studies. The new ideas about the corporate identity of the church were embedded in the public law of the church and of the state. As we consider in Chapter 5, Kantorowicz shows that the secular lords also began swearing not to alienate common property in their coronation oaths.

The next major constitutional crisis occurred in 1378. The papacy had been located in Avignon for seventy years and was in the process of returning to Rome when the reigning pope died. The College of Cardinals initially elected Italian Bartolomeo Prignano as Pope Urban VI. A few months later, the College of Cardinals declared that they had been forced to elect Urban under duress and wished to change their mind. The College then elected a French pope, Clement VII, who promptly returned his papacy to Avignon. The church was stuck in another constitutional anomaly: it had two popes.³⁸ Because the investiture crisis had not settled the issue of who had the ultimate power of decision over the church in a time of crisis, the pope or a general council, the church faced a serious dilemma.

The issue was not resolved until a general council, called with the assistance and influence of the French king and the Holy Roman Emperor, met at Constance in 1414. The Council of Constance affirmed that a general council had authority over the pope in critical matters and resolved that councils would thereafter meet on a regular basis. The conciliarists triumphed, and a new constitutional structure for the church was decreed. The schism produced a concrete articulation of the church as a corporate body, independent and potentially self-governing. In a crisis, ultimate authority in the church rested with the body of the faithful, the *congregatio fidelium*, as represented by a general council. The councils were composed of elites – archbishops, bishops, abbots, and cardinals – and in no sense was there representation of the typical churchgoer. However, an identifiable corporate entity had been recognized.

³⁸ For a history of the Great Schism see Ullman (1972a).

The great schism was a constitutional crisis. Although the locus of the crisis lay within the church, it spread to larger European society through the identification of factions within the church with factions in European politics. The close parallels between canon and civil law would bring the innovations of the Council of Constance to the larger structure of the dominant coalition within Europe. Churches were developing a corporate identity outside the umbrella of the state, following a principle actuated at the Council of Constance.

Ironically, the constitutional provisions of the Council of Constance, although firmly embedded in the canon law, did not prevent the resurgence of a powerful papacy. The popes soon ceased to call regular councils. The pope remained not just the head of the church, but as God's representative on earth he remained above the law of the church. The effective concentration of power in the papacy and the church in Rome would be a powerful force leading to the reformation after 1520. If the decrees of the Council of Constance were in fact canon law, then a pope above the law was unconstitutional. Resolution of this issue did not come easily to the Catholic church.

2.8 Mature Natural States: France and England in the Sixteenth, Seventeenth, and Eighteenth Centuries

The previous section focused on the church because in the legal and political environment of medieval Europe, organizational changes in the structure of the church immediately applied to the wider society. The problem of guaranteeing elite organizational independence was not solved by the church, however. These are subtle questions. For our last examples, we turn to France and England after the sixteenth century.

Increasing support for the corporate identity of elite organizations proceeded steadily in France in the seventeenth and eighteenth centuries. Few French corporations functioned like a modern business, however. Corporate identities were created for courts, municipal governments, and financial offices within the government. The corporations were central institutions in the development of the French economy and polity. France was a nation of independent and diverse geographical units, each with unique histories, institutions, and relations with the central government. The corporations provided an important element by which the dominant coalition was able to cement relationships across the nation through the explicit creation of limited entry corporate privileges. The crown created rents that could then be shared with the corporate office holders, and, in turn, the sale and

taxation of corporate offices provided an important source of royal revenues. The rise of corporate groups was tied closely to the fiscal structure of the French crown.³⁹ In return for creating corporate privileges, the crown sold offices within the corporation to elite individuals. Office holders could sell their office or transfer it to descendants under a complicated set of rules.⁴⁰

Cities and towns were key corporations. Bossenga's (1991) study of the city of Lille offers a window into the inner workings of corporations in the dominant coalition in seventeenth- and eighteenth-century France. In 1667, Louis XIV conquered Walloon Flanders and accepted the keys to the city of Lille. After negotiations with the city leaders, Louis confirmed the city's ancient rights and prerogatives, and signed a capitulation treaty detailing the privileges the city was to enjoy. Upon the ascension of Louis XV in 1726, Lille's privileges were again confirmed, this time after the payment of 703,300 livres, a payment the town felt was a violation of the very privileges it was purchasing. In 1774, Louis XVI ascended to the throne, and although he did not press Lille for a monetary payment, he was not much interested in swearing mutual oaths or honoring the ancient privileges of the city (Bossenga, 1991, pp. 1–4).

As with most cities, Lille possessed ancient privileges that endowed the city with a corporate identity vested in its leading citizens. Leadership was not selected by election, but by membership in the corporation of municipal government. The corporate privileges and identity were durable, but subject to the lives and personality of the city's mortal overlords. In some dimensions, the city was a perpetual organization. Membership in the municipal corporation came with a significant social persona, but the identity of the city itself as a corporate entity was independent of the individuals who made up the corporation. The city could borrow money at lower rates than the king. When under fiscal duress, the king forced loans from the city largely by offering new offices in the municipal corporation for sale. The town council was forced to buy the new offices with borrowed funds. The

³⁹ Although France often comes off badly in comparisons with Britain in the study of economic and political development in the eighteenth and nineteenth centuries, the recent research of Lamoreaux and Rosenthal (2004, 2005), and their wider research with Guinaame and Harris (2007), consistently shows that France was organizationally more sophisticated with respect to business corporations than Britain or the United States through much of the nineteenth century. We return to their work in Chapter 6.

⁴⁰ Doyle (1996), Kwass (2000), and Root (1994) explain how the French utilized corporate privileges and Collins (1995) gives a more general background on France in the early modern period.

crowns also imposed financial responsibilities on the city, both in kind for expenditures on troops and barracks and in money as the town was required to remit a share of its taxes to the center (Bossenga, 1991, pp. 22–46).

Municipalities like Lille were halfway down the path to perpetual life. Some aspects of the corporate identity of Lille were perpetually lived, others were subject to the will of the king; and the entire corporation was potentially dependent on the personal goodwill of the sovereign. The recognition that Lille possessed ancient privileges was a tacit acknowledgment that Lille possessed a corporate identity of its own, an identity not directly dependent on the crown. In normal times, the city and the king would negotiate changes in their arrangements, but in times of duress the king possessed the ability to change the agreement unilaterally by manipulating several dimensions of the corporate structure, like the number of offices in the corporation. The king could and did recognize the city of Lille and its ancient privileges, but he was subject to continually changing circumstances both inside and outside of France. Thus, the king could not credibly commit to honor his agreements with Lille, and he broke them when circumstances compelled him to do so. The same conditions held for thousands of cities, towns, courts, and corporate entities throughout France. In one sense, France in 1700 was the most corporately organized society the world had ever seen to that point in history. Nevertheless, the needs of France as a mature natural state, a coalition of interests subject to forces continually in flux, prevented those corporations from taking on a perpetual life of their own that was completely independent of the crown.

The evolution of corporate entities in England followed a different path than in France. England also had a long history of chartering municipalities and, within them, craft guilds. These formal chartered entities possessed economic privileges and limited access in several different dimensions.⁴¹ In the mid-sixteenth century, England began chartering joint-stock companies, the largest of which engaged in overseas trade and colonization. The first was the Russia Company, followed by the Virginia Company, the East Indies Company, the Massachusetts Bay Company, and the Hudson Bay Company, among many others. These were all natural state creations: organizations controlled by elites who enjoyed, in this case, the explicit legal support of

⁴¹ For municipal charters and governments see Webb and Webb (1908). The interpretation of medieval craft guilds underwent steady revision in the twentieth century, beginning with the collections of charters in Ballard (1913) and Ballard and Tait (1923) through Scott (1917), Thrupp (1948), Munro (1977, 1990), Swanson (1988, 1989, 1999), and the recent work of Richardson (2001, 2004). The literature has moved from treating urban guilds as monopolies to economic entities with a much more varied set of privileges.

the state. Commercial and trading interests had always been a part of the dominant coalition in England, but in the seventeenth century, the growing importance of commerce, both domestic and international, engendered a shift of emphasis within the coalition, away from landed interests and toward commercial and manufacturing interests. These adjustments were not accomplished without bloodshed and disorder.

Two aspects of the seventeenth-century English struggles are worth noting. First, these struggles were not simply between parliament and king, as so many histories suggest, but a struggle between the elements in the dominant coalition — particularly commercial interests — which sought more power and new rules to benefit its interests. Second, many of the new rules sought by the commercial constituency were in the form of greater rights and impersonality. Paralleling the demands of earlier landed elites who sought to restrict the king's ability to adjust the dominant coalition through redistributing rights in land, the subject of the next chapter, the new commercial constituency sought greater security from arbitrary decisions by the king to adjust the coalition to changing circumstances.

Similar to France, it was difficult to institutionalize the impersonal aspects of organization. In a very important sense, aspects of the seventeenth-century claims of the king's arbitrary action parallel our observations about the Roman emperor and the popes being above the law. As heads of natural states, all of these rulers used their powers to maintain the natural state order. From a modern, constitutional perspective of an open access coalition, this behavior looks arbitrary; but from the perspective of the natural state, this behavior is a logical consequence of natural state coalition maintenance under changing circumstances.⁴²

2.9 Natural States

The progression of natural states involves increasing more complex societies, requiring increasingly complex institutions that support more complex organizations. In all natural states, economics is politics by other means: economic and political systems are closely enmeshed, along with religious, military, and educational systems. The close interrelationship between the church and the state in medieval Europe from Charlemagne through the sixteenth century illustrates the futility of drawing hard and fast

⁴² Our perspective helps explain the behavior of the king prior to the Glorious Revolution, which North and Weingast (1989) took as given.

lines around governments in natural states and calling them the state. Power is dispersed in natural states.

We have stressed the importance of institutions that enable the organization of elites as the primary element in moving from fragile to basic to mature natural states. Organizations in fragile natural states are usually closely tied to powerful individuals. The emphasis on personal identity begins with the identification of specific individuals with specific privileges in a dominant coalition. The dominant coalition is the organization of organizations that powerful elites are associated with. As societies move toward basic natural states, these identities become less associated with specific individuals and more with social personas that become associated with powerful organizations. As societies move from fragile to basic natural states, these organizations become clearer and better defined. Organizations begin to become institutionalized. This process initially occurs simultaneously in the public and private sector; indeed, it is a primary reason that governments in most limited access societies appear so corrupt to observers from open access societies: Most important basic natural state organizations are closely associated with the (private) individual identities of the elites who inhabit them. These organizations span the boundary of public and private, personal and social.

In mature natural states, credible institutions evolve that provide organizations a measure of rule of law. As more complex organizations develop, both inside and outside of the formal government, the distinction between public and private organizations begins to appear. The first steps toward Weberian states or governments with consolidated, monopoly control over the military occur in mature natural states. Sustaining some amount of rule of law for elite organizations appears to be incredibly complicated to pull off and is the beginning of the doorstep conditions.

No teleology pushes states through the progression from fragile to basic to mature natural states. The dynamics of natural states are the dynamics of the dominant coalition, frequently renegotiating and shifting in response to changing conditions. If adjustments lead to more power and rents based on personal identity, institutions become simpler and organizations less sophisticated, and the society moves toward the fragile end of the progression of natural states. If adjustments lead to more power based on durable agreements, institutions become more complex and organizations become more sophisticated, and societies move toward the mature end of the progression. No compelling logic moves states in either direction. As governments become more sophisticated and institutionalized across the natural state progression, they also become more resilient to shocks. Mature

natural states are more stable than basic natural states, which are more stable than fragile ones.

We define the progression of natural states – fragile, basic, mature – in terms of the organizations they can support. Paralleling this increasing organizational sophistication is an increasing institutional sophistication. Growth in state capacity is equally important as that of private organizations. Creating more sophisticated and complex organizations therefore requires that the state becomes more sophisticated and complex; it must be capable of greater tasks and a greater range of credible commitments that create durability and predictability, including the emergence of the rule of law.

Basic natural states are richer in organizations and more institutionally complex than fragile natural states. They have greater specialization and division of labor among organizations and institutions, often including organizations specializing in trade, education, religion, production (e.g., mineral extraction), and taxation. These societies often also have a range of other institutions, such as succession rules for determining the next ruler when the current one dies; institutions for dividing spoils of conquest; and perhaps an official forum in which nobles formally negotiate with the ruler, such as a cortes or parliament. Basic natural states also tend to differentiate public from private law.

Mature natural states are yet more institutionally complex than basic ones. To support a wider range of private organizations, these societies must develop institutions to better support private organizations, including a form of legal system for administration of intra- and inter-organizational contracts and more secure commitments to protect private organizations from expropriation by the state and the dominant coalition. In other words, a more complex rule of law must emerge in mature natural states. Because rule of law cannot be improved by fiat – which is too easily undone – many attempts to enhance and extend the rule of law in basic natural states fail. The institutions providing these services must also be embedded in the state and society in a way that protects them from the rearrangement of privileges among members of the dominant coalition.

One of the principal institutional issues that emerged in this chapter concerned the problem of constraining personality: putting the king under the law. At the level of societies, the head of the dominant coalition – whether the pope or the Catholic Church, the emperor of Rome, or the king of a European state – reflects the realities of these natural states: the ruler is often above the law. This allows him or her to adjust the rules, privileges, rights, and laws to suit the needs of the coalition as the fortunes of various elites rise and fall. Elites gaining power must be granted more privileges and

rents while those losing power also lose privileges and rents. The ruler is not free to make these decisions at his discretion, but must instead attempt to maintain a coalition to support the natural state. Failure to do so risks coups, civil war, and other forms of disorder.

The ruler is not the only person above the law; in many natural states, this problem extends down through the institutions and organization of society. How are the powerful personalities of elites to be constrained within institutional structures that subject and commit them to the organizations of which they are a part? How the bishop is constrained by the cathedral college, the duke by his manorial court, or the corporate leader by his corporate bylaws are critical reflections of the entire social order.

The issue of constraining the ruler plagued the West for two millennia. A major feature of this chapter's cases involved the emergence of a dual identity of the leader – the personal and the social identity – whereby the social identity began to embody a series of duties and constraints for the ruler (Kantorowicz, 1997[1957]). We return to this issue on several occasions in later chapters.

APPENDIX: SKELETAL EVIDENCE AND EMPIRICAL RESULTS

Richard Steckel and Jerome Rose invited a large group of physical anthropologists, economic historians, demographers, and medical historians to document and analyze the history of health in the Western hemisphere using data from archaeological skeletons (Steckel and Rose, 2002). Anthropologists contributed data on several skeletal indicators of health for individuals who had lived at sites scattered from South America to southern Canada. The combined data set includes 12,520 skeletons from 65 localities representing populations who lived from 4,500 B.C.E. to the early

Table 2.1. *Expected probabilities of violent trauma*

Group	Expected probability (%)	% male	Sample size
Pre-Columbian, hunter-gatherer	13.39	47.13	715
Pre-Columbian, city	2.70	48.63	183
Early post-Columbian, village	9.48	44.43	673
European-American, city	7.25	59.88	496
African-American, city	18.53	49.90	511

Source: Western Hemisphere database (Steckel and Rose, 2002).

Session 2: The West and the Near East – Past and Present

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

CHAPTER THREE
POLITICS, ETHICS, AND
WAR IN PREMODERN ISLAM

Given the history of Shari'a reasoning, it is not surprising that we should turn to Iraq for an early example of the ways Muslim scholars dealt with questions of war. The Qur'an, of course, contains passages justifying military force, as do standard accounts of the Prophet's life and work. In the context of struggle between the early Muslims and their Qurayshi opponents, these sources suggest a developing sense that fighting is an appropriate means by which Muslims should seek to secure the right to order life according to divine directives. The notion of a just war is therefore an aspect of the foundational narrative of Islam.¹

Similarly, we know that questions about the justice of fighting emerged in the context of the early conquests or campaigns to "open" neighboring territories to Islamic governance. The notion of the Prophet's letter to the rulers of Byzantium and other empires of late

Copyright © 2009. Harvard University Press. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

antiquity is instructive in this regard, as are the stories of arguments between ‘Ali and others regarding the proper mode of response to Mu‘awiya, the Kharijites, and other perpetrators of civil strife. Many of the texts cited as precedents in the developing discourse associated with Shari‘a reasoning are in fact reports of actions taken, or of arguments confirmed or denied, in relation to the justification and conduct of war in this early period of Muslim expansion and self-definition. The brief chapter on *jihad* in the *Muwatta* of Malik ibn Anas is instructive.² It contains a collection of reports regarding the example set by the Prophet and his companions. Topics range from the advisability of taking copies of the Qur’an into enemy territory, to the distribution of war prizes and the proper care of horses. Overall, the point is to suggest the importance of participation in fighting under those conditions and according to those norms consistent with the notion of submission to the will of God. Thus we find Muhammad saying: “The man who fights in the cause of the Lord may be compared to one who fasts and prays and is not weary of prayer or fasting, until he returns from the fighting.”³ Endurance or persistence in the task one is given, whether fighting or fasting and praying—that is the key, for one who gives his or her life to the Lord of the worlds. Or again, we find Muhammad in conversation, as follows:

The Prophet of God said: “One who spends his wealth in the cause of the Lord will be called on at the gate of heaven—‘O servant of God, this is bliss!’ The man of prayer will be admitted from the prayer gate, the man of fighting will be admitted from the fighting gate, and the man of fasting from the gate of fasting.” Abu Bakr al-Siddiq said: “Prophet of God, for a man who is called from one gate, there would be no trouble, but would there be one who is called from all the gates?” The Prophet of God said: “Yes, and I would hope you would be among them.”⁴

Such texts provided the material for the more developed treatises of the Iraqi school. Of these, Abu Yusuf's "Book on the Land Tax," which deals with questions arising from the administration of conquered territories, seems to be the earliest. For our purposes, al-Shaybani's "Book of the Foundations," and within that work, the chapters on *al-siyar*, are more interesting and worthy of sustained analysis.⁵

From the outset it is clear that al-Shaybani's text approaches the question of war in connection with broader issues of political ethics. The term *siyar*, for example, means "movements." The movements in question are those between and within two broadly defined political-territorial associations, namely the "territory of Islam" and the "territory of war." The first of these terms indicates the area within which Islam is the dominant political reality. The second is a generic term for political-territorial associations not governed by Islam. In the territory of Islam, the ruler is a Muslim, the system of government is organized to serve the expansion of Islam, and the people within the territory are identified according to their participation in various groups, each of these defined by its relation to Islam. The primary group is Muslims, but there are also "protected people" (*ahl al-dhimma*, mainly Christians and Jews, though other groups could come under this category), "rebels" (*bughat*, indicating groups of Muslims dissenting from the policy of the established leadership), "apostates" (*al-murtadd*, meaning those who were once Muslims but have in some way disassociated themselves from the true faith), and "brigands" (*muharibun*, indicating criminals who prey off ordinary people and live as "highwaymen"). In the territory of war, the overarching category is *harbi* (roughly, "war person"). Subcategories, for example women, children, the elderly, or, in some cases, Muslims residing in the territory of war, are referred to in discussions of the means of war, especially with respect to the protection of noncombatants. Al-Shaybani and his colleagues are thus concerned with the justification and conduct of war in the context of the kind of

political-territorial arrangement outlined in previous chapters. Their judgments are developed in response to questions like “What is the proper behavior for Muslim fighters who cross into the territory of war?” or “What is the appropriate way to respond to Muslim rebels?”

War is a means to a political end, which is the establishment and governance of a political-territorial association governed by Islam. We may take this further: the establishment of an Islamic state is itself a means by which the Muslim community can carry out its divinely mandated mission of calling humanity to the relationship with God signified by submission, *al-islam*. In this connection, war is a means to a political end (establishing an Islamic state), which is itself a means to an overarching religious goal (calling humanity to Islam). This approach suggests that there is nothing particularly good or bad about war in itself. It is a means to an end, and should be viewed as such. Resort to war is thus a matter of estimating its probable effectiveness in attaining certain goals. Notions of just conduct in war are similarly suggested by the desire to gain particular objectives.

Perhaps the simplest way to illustrate this way of thinking is to cite a report with which al-Shaybani’s text begins. Here the Prophet Muhammad is depicted as giving directions to Muslim fighting forces—first, to those in command, and through them, to all members of the army. The text establishes precedents to which everyone subsequently engaged in the practice of Shari‘a reasoning about war will recur.

Whenever God’s Messenger sent forth an army or a detachment, he charged its commander personally to fear God, the Most High, and he enjoined the Muslims who were with him to do good.

He said:

Fight in the name of God and in the path of God. Fight

the *mukaffirun* [ingrates, unbelievers]. Do not cheat or commit treachery, and do not mutilate anyone or kill children. Whenever you meet the *mushrikun* [idolaters], invite them to accept Islam. If they do, accept it and let them alone. You should then invite them to move from their territory to the territory of the émigrés. If they do so, accept it and leave them alone. Otherwise, they should be informed that they will be in the same condition as the Muslim nomads in that they are subject to God's orders as Muslims, but will receive no share of the spoil of war. If they refuse, then call upon them to pay tribute. If they do, accept it and leave them alone. If you besiege the inhabitants of a fortress or a town and they try to get you to let them surrender on the basis of God's judgment, do not do so, since you do not know what God's judgment is, but make them surrender to your judgment and then decide their case according to your own views. But if the besieged inhabitants of a fortress or a town ask you to give them a pledge in God's name or in the name of God's Messenger, you should not do so, but give the pledge in your names or the names of your fathers. For if you should ever break it, it would be an easier matter if it were in the names of you or your fathers.⁶

There is much to learn from this text, which is basic to all Shari'a reasoning on the topic of war. In it, for example, al-Shaybani and others found precedent for limiting the right of war to the head of state. Muhammad, as W. M. Watt's biographical study suggests, was both prophet and statesman.⁷ Scholars in the Shari'a tradition saw in this report an example of his statesmanship, which was to be followed by those who succeeded him in the office of *khalifa*.

Resort to war requires an order from a legitimate authority. It also requires a just cause and righteous intention. So scholars in the

Shari'ah tradition understood, as the words attributed to the Prophet echoed in their minds: "Fight in the name of God and in the path of God. Fight the *mukaffirun*." The purpose of war is the establishment and governance of an Islamic state. That state, in turn, finds its purpose in connection with the Muslim community's mission of calling human beings to Islam. A just war must be tied to these purposes. And those authorizing it show that their intention is consistent in this regard by carrying out a specific protocol, namely, the issuing of an invitation to the enemy.

One might well ask how the "enemy" is identified. The Prophet's statements seem to characterize them as *mukaffirun*, which literally means "ingrates" and is typically translated "unbelievers." Staying strictly with the terms of this report, we may say that such people are identified by their response to the invitation proffered by Muslim fighters. The mention of *mushrikun* a few sentences later provides further specification. Signifying "idolaters" or perhaps more literally "associationists," the term often bears the meaning of "polytheists." That would constitute a considerable limitation regarding resort to war, since the category specifically does not include Jews and Christians. For al-Shaybani and his colleagues, however, Jews and Christians are clearly among those to whom an invitation should be issued, unless and until they pay tribute and thus come under the sway of Islam as protected people. Thus practitioners of Shari'ah reasoning would find just cause for war in the fact that a given people refused the invitation of the Muslims to accept Islam or to pay tribute and come under the protection of the Islamic state.

The procedures outlined in this report establish a notion of war as a means of pursuing goals considered legitimate. The nature of the invitation makes it clear that war is not the first or primary means recommended for pursuing these goals. Fighting is prescribed only when other means have failed. Although this formula is not a precise equivalent of the Western just-war criterion of "last resort," it is nev-

ertheless an indication that resort to war must follow an attempt to pursue legitimate goals by nonlethal means. Thus the tradition of Shari‘a reasoning already provides an equivalent to the just-war criteria of legitimate authority, just cause, righteous intention, and (at least) “timely” resort.⁸ We might also see in the overarching purpose of establishing and maintaining an Islamic state something equivalent to the just-war criterion of “aim of peace.” Proportionality and reasonable hope of success, by which just-war tradition instructed those deliberating about resort to war to make a good-faith effort to estimate the overall costs and benefits of fighting in particular cases, do not show up, at least in this report.

The preoccupation of the report is of course with those who are not already under the protection of the Islamic state. As we read al-Shaybani’s text, we come to understand that fighting against residents of the territory of Islam is also possible, though justified with somewhat different reasons. Governance of the non-Muslim “protected” peoples, for example, is set in terms of obligations established by treaties. If the protected peoples violate their obligations, for example by initiating fighting against the armies of the Muslim state, then the ruler of the Muslims is justified in disciplining them.⁹ Again, if rebels rise up against the state, fighting can be justified.¹⁰ In either case, the overarching purpose of fighting is to restore peace, order, and justice to the territory of Islam. Protected peoples are to be returned to their rightful status. The notion that rebels are to be reconciled implies that war not only is a means of discipline, but is to be conducted in such a way that any legitimate grievances of the rebels have may be redressed. In the somewhat different case of apostates, the point is to return those who have deserted it to the way of Islam or, failing success in the matter, to prevent injustice by killing them.¹¹ Similarly, in the case of highwaymen the objective is the security of the territory of Islam and its residents, and the ruler is justified in undertaking those measures necessary to protect it.¹² In

an interesting demonstration of confidence, al-Shaybani and his colleagues do not address the question of fighting in response to an enemy invasion.

Right authority, just cause, and other criteria already mentioned are not the only measures of justice in war, of course. It is noteworthy that the prophetic dicta also include matters related to lawfulness in the conduct of war. In particular, the saying of the Prophet includes prohibitions on cheating, treachery, mutilation, and the killing of children. This saying, along with other reports of the Prophet's practice, gave rise to more extensive restrictions on targeting that should be seen as analogous to the just-war tradition's concern for "discrimination" or the "immunity of noncombatants" from direct and intentional attack. Other considerations have to do with fair dealing in the distribution of booty and with prudence in making agreements. As mediated by al-Shaybani and his colleagues, the report does not contain any references to weapons that may or may not be used—the typical concern associated in the just-war tradition with just-war proportionality.

Al-Shaybani and his colleagues built on this and other reports of the practice of Muhammad and his companions, providing answers to specific questions raised in their own day. In reading their collected judgments, it is important to remember their role as advisers at the Abbasid court, which explains their preoccupation with matters of administration. The following account of an exchange between two of the early masters of al-Shaybani's school is typical:

Abu Yusuf said: I asked Abu Hanifa concerning the food and fodder that may be found in the spoil and whether a warrior in need may take from that spoil any of the food for himself and fodder for his mount.

Abu Hanifa replied: No harm in all that.¹³

Such questions have little to do with the concerns of modern military thinking. They are connected with the questions of the early

Hanafi jurists, however, because of their focus on movements between the territories of Islam and territories of war, and probably also because of movements in Muslim military organization in the direction of a professional army.¹⁴ The question emerges in connection with the repeated consensus of the school that booty should be divided only after its return to the territory of Islam. Such judgments attempt to ensure a just distribution of the booty, and also to keep the troops focused on their military task.

In dealing with the conduct of Muslim fighters in enemy territory, however, al-Shaybani and others addressed several questions of direct relevance to our concerns. Thus we read the following:

If the army attacks the territory of war and it is a territory that has received an invitation to accept Islam, it is commendable if the army renews the invitation, but if it fails to do so it is not wrong. The army may launch the attack by night or by day and it is permissible to burn fortifications with fire or to inundate them with water.¹⁵

The first sentence is reminiscent of the Prophet's directives to fighters. The requirement of an invitation is critical for establishing just cause and righteous intention. It also reveals much regarding the overarching religious and political purposes that war is supposed to serve, and places war within a hierarchy of means to be deployed in pursuit of those purposes. Here we are to envision a case in which hostilities have begun. An invitation having been given and refused, the enemy is tasting the steel of the Muslims. Should the Muslims now renew the invitation, giving the enemy a second chance to submit, or should fighting be carried to a conclusion—as al-Shaybani seems to assume, a conclusion that involves decisive victory for the Muslims? The judgment is in the nature of a recommendation: renewal of the invitation would be a good thing but is not required. Commanders in the field have discretion in this matter.

The second sentence, however, moves in a direction not covered by the Prophet's statements. Here scholars directly address the question of military means, and they do so in ways that give a great deal of latitude to the fighters. The army may fight by night or by day—the former likely involving a greater risk of injury or death to the Muslim fighters, as well as to certain people on the enemy side who fall into the category of noncombatants. Similarly, the army may utilize tactics that increase the potential for injury, specifically burning fortifications with fire or inundating them with water.

Throughout, the text reveals a strong inclination toward a position one might characterize as “military realism.” Given that requirements associated with just war are satisfied, and Muslim troops are thus engaged in a legitimate war, al-Shaybani and his colleagues are willing to grant wide latitude to commanders in the determination of appropriate means. Such latitude is not total, of course. The Prophet's dicta are clear regarding cheating, treachery, mutilation, and the direct targeting of children. With respect to the last, in particular, al-Shaybani and his colleagues demonstrate respect for early Muslim practice in classifying a number of enemy persons as, in effect, immune from direct attack. In addition to children, slaves, women, old people, the lame, the blind, and the helpless insane are immune from direct attack.¹⁶ In addition, the text indicates a concern regarding the presence, or even the possible presence, of Muslims in enemy territory—a condition that must have been ubiquitous along the border between their respective territories, where Muslim and Byzantine forces often took turns conquering and reconquering particular towns and regions.¹⁷ Muslims in the midst of the enemy are also presumed “innocent,” and thus immune from direct attack, for the lives and property of Muslims had been considered inviolable from the earliest times. Direct and intentional targeting of any of these categories of people violates an express directive of the Prophet, or a legitimate extension of that directive. It also in-

volves dishonor to fighters, who win their reputation in direct engagement with other fighters. As the Qur'an has it (2:190):

Fight in the path of God
Those who are fighting you;
But do not exceed the bounds.
God does not approve the transgressors.

“Do not exceed the bounds”; even in the revealed text, the reference is to a warriors' code. Now al-Shaybani and his colleagues, guided by the example of the Prophet and his companions, must elaborate such a code for their day.

As they do so, their judgments indicate both their adherence to precedent and a disposition toward military realism. Thus we read the following series of questions and responses:

I asked: Would it be permissible to inundate a city in the territory of war with water, to burn it with fire, or to attack with hurling machines even though there may be slaves, women, old men, and children in it?

He replied: Yes, I would approve of doing all of that to them.

I asked: Would the same be true if those people have among them Muslim prisoners of war or Muslim merchants?

He replied: Yes, even if they had Muslims among them, there would be no harm to do all of that to them.

I asked: Why?

He replied: If the Muslims stopped attacking the inhabitants of the territory of war for any of the reasons that you have stated, they would be unable to go to war at all, for there is no city in the territory of war in which there is no one at all of these you have mentioned.¹⁸

One who fights should employ the means necessary to win—or so it would seem, given the final statement in this sequence of judgments.

That is not all there is to it, however. In just-war tradition, military policy can be highly realistic even while adhering faithfully to the guidelines indicated by the just-war criterion of discrimination or noncombatant immunity. As more than one interpreter suggests, the idea is that noncombatants are immune from attacks that are direct and intentional, but they cannot be immune from any harm whatsoever. An attack that is legitimate, in the sense that it is directly intended as a strike against the enemy's ability to fight, may result in the deaths of noncombatants. This outcome may be purely accidental, such as when a child happens to wander by a military target and is caught in the crossfire. It may also be foreseeable but unavoidable, as when an enemy's military resources are deployed in the midst of a civilian population. In this case the deaths of civilians are the indirect or secondary (in the sense of "unintended") consequence of the deployment of means necessary to accomplish a legitimate military purpose. Soldiers whose actions take place under such conditions are excused from the guilt associated with unjust killing. Alternatively, the actions undertaken may be described as justified, though with some regrettable (secondary) effects. This judgment of acceptability holds so long as the incidence of unintended, collateral damage is proportionate with the military objective.

Given this line of thinking, it is not surprising to read:

I asked: If the Muslims besieged a city and its people positioned behind the walls shielded themselves with Muslim children, would it be permissible for the Muslim fighters to attack them with arrows and hurling machines?

He replied: Yes, but the warriors should aim at the inhabitants of the territory of war and not the Muslim children.

I asked: Would it be permissible for the Muslims to attack

them with swords and lances if the children were not intentionally aimed at?

He replied: Yes.

I asked: If the Muslim warriors attack with hurling machines and arrows, flood cities with water or burn them with fire, thereby killing Muslim children or men, or enemy women, old men, blind, crippled, or lunatic persons, would the warriors be liable for blood money or acts of expiation?

He replied: They would be liable neither for blood money nor for acts of expiation.¹⁹

The reasoning is quite reminiscent of Western just-war tradition and its approach to collateral damage. One would be quite wrong, in the case of just war or of Shari'a reasoning, to read such a passage as negating respect for the immunity of noncombatants. The point is that the attacks are not directly and intentionally aimed at noncombatants. Without this overarching categorization, the military acts described would be unjust, and those engaging in them would need to make restitution. As it is, al-Shaybani and his colleagues are best read as trying to combine military realism with respect for rules that measure just conduct in the midst of war. In so doing, they mean to establish norms appropriate to the conduct of a professional army.

The judgments advanced by al-Shaybani became part of a historical deposit with which subsequent generations had to deal. Not all later practitioners of Shari'a reasoning agreed with al-Shaybani. Even the title of a text written by the great al-Tabari (d. 923), the "Book of the Disagreement of the Practitioners of *Fiqh* on the *Jihad* and the Tax [imposed on conquered peoples]" (*Kitab al-Ikhtilaf al-Fuqaha fi al-jihad wa 'l-jizya'*), indicates the conversational nature of the discipline. Shari'a reasoning about war constituted a lively discourse, in which there could be considerable disagreement about the administrative matters so important to al-Shaybani and his col-

leagues. There could even be disagreement over whether soldiers who participated in an action in which noncombatants were unintentionally killed should perform acts of expiation, or over the precise categories of people who belonged to the class of noncombatants.²⁰

In Shari'a reasoning, it is best to think of judgments like those of al-Shaybani as precedents. As recent debates in American constitutional law show, the status of precedent in legal reasoning can be contentious. That seems the case with Shari'a reasoning as well, with some opting for a stronger, some for a weaker notion of precedent. It is clear, however, that the notion of Shari'a reasoning requires a balance between respect for the judgments of one's predecessors and the need for independent judgment (*ijtihad*), reflecting the idea that changing circumstances require fresh wisdom.

By the time of al-Mawardi (d. 1058), the preoccupation of Shari'a reasoning appears to have shifted from the details of administrative law to the superstructure of right authority. Consider, for example, the following:

When the Caliph appoints a governor of a province or city, the latter's jurisdiction may be either general and unrestricted or special and restricted. General governorship is of two kinds: regular (by recruitment), in which appointment is made by the sovereign's free choice; and usurped, in which the appointment is made through coercion . . .

Governorship by usurpation is coercion in the sense that its holder acquires by force certain districts over which the Caliph gives him a decree of appointment, assigning him their management and the maintenance of public order therein. By seizing power, the governor becomes an independent and exclusive controller of political matters and administration, while the Caliph, by his permission, becomes

the implementer of the dictates of religion, thus transforming unlawfulness into legality, and the forbidden into the legitimate. Although by doing so he does not adhere closely to the conventions of appointment in respect of conditions and procedures, his action upholds the canon law and its provisions in ways that are too important to be disregarded. Thus, decisions that would normally be unacceptable in regular appointment based on choice are permissible under usurpation and necessity, owing to the difference between ability and incapacity.²¹

As al-Mawardi's text makes clear, he presumes that designation of a ruler is a requirement of Shari'a; that the office must be singular (that is, there cannot be two legitimate rulers, even if they reside in different cities); and that the person holding leadership must have certain qualifications, most notably descent from the Quraysh, the clan into which the Prophet Muhammad was born. It is noteworthy, given other trends in the Muslim community of his day, that al-Mawardi does not require that the ruler be "of the Prophet's household" or "of the family of the Prophet (or of 'Ali)"; this was a position held by the various Shi'i groups. Al-Mawardi further stipulates that the ruler may be either designated by his predecessor or chosen by electors. The latter may be any of the number of Muslims who are truthful, who possess knowledge of the required characteristics of the ruler, and whose prudence and wisdom make it likely that they will choose the right person. As a matter of custom, he writes, this role usually falls to the leading citizens of the capital of the territory of Islam.

Thus far, there is nothing controversial about al-Mawardi's argument. Indeed, it reflects the standard presentation of Abbasid imperial practice. Similarly, when al-Mawardi comments on "the ministry" (*al-wizara*), noting that appointment to this office is the right of

the sovereign ruler and that appointment may constitute the minister's authority as "delegated" or "executive," he seems to reflect the standing practice of the Abbasid court. No ruler is able to administer the day-to-day affairs of a far-flung empire; the Abbasid caliphs tended to assign many of these tasks to court advisers, the most authoritative of whom held positions as ministers. "Delegated" authority of a minister, al-Mawardi says, suggests greater independence; "executive" authority indicates that the ruler himself is setting policy and restricting ministerial practice to carrying out his orders.

It is in the comments on "governors" that al-Mawardi's reasoning becomes creative. Many, perhaps most, commentators read his notion of a governor "appointed" by coercion as a pure example of *realpolitik*.²² It was in fact the case from about 935 on that no Abbasid ruler really exercised power in the style of the High Caliphate. Instead, power was concentrated in the provinces of the empire, whose governors commanded armies, collected taxes, and generally carried out the duties of statecraft. Among these, the more powerful carved out larger territorial units than Abbasid practice allowed. Powerless to remove such "usurpers," what was the caliph to do? By al-Mawardi's lights, the caliph should do his duty, namely, maintain the unity of the territory of Islam by authorizing such governors to rule as agents of the Abbasid court. However irregular such a strategy may seem, he writes, a caliph who acts in this way "upholds the Shari'a and its provisions in ways that are too important to be disregarded." In effect, the ruler "delegates" authority to the usurping governor. The latter thereby achieves legitimation, and the former hopes to preserve the unity of Islam, as is his charge. Governors must therefore "uphold the office of the sovereign [the caliph] as a vicar of prophecy and manager of religious affairs, in order to fulfill the legal requirement of ensuring its existence and preserving the rights arising therefrom."²³ Insofar as the usurper lacks formal qualifications, "it is up to the Caliph to sanction his appointment as a

means of winning him over and putting an end to his disobedience and intransigence.”²⁴ In certain conditions, the caliph may even appoint a minister to work with the usurper and to exercise executive power.

With respect to war, al-Mawardi’s judgment legitimating governorship by usurpation is a crucial modification in the notion of right authority. “General” governorship, in particular, comes with full power to make war.²⁵ There is a sense in which one who seizes power becomes the caliph’s designated minister for political and military affairs. After al-Mawardi’s time (and even, to some extent, during his lifetime), leaders of the Seljuq Turks played such a role, holding positions as *wazir* (minister) or *al-sultan* (the power) in concert with the Abbasid caliphs’ role as guardian of Islam.²⁶

If general governorship comes with the full power to make war, it also imposes duties. Al-Mawardi’s comments on this matter are a subject for study in themselves. As commander, the general governor sees to the formation, equipping, and morale of a fighting force. He manages fighting against “the idolaters in enemy territory,” as well as against apostates, rebels, and brigands within the territory of Islam. In the first group, al-Mawardi designates two classes: “those who have received the call to Islam but rejected it and turned away from it” and “those whom the call to Islam has not reached.”²⁷ The former are subject to attack or the threat of attack at any time, depending on the commander’s sense of what is best. The latter, of whom there would be “very few today on account of the victory the Almighty has accorded His Prophet’s mission, unless there be nations unknown to us beyond the Turks and Greeks we meet in eastern deserts and remote western areas,” cannot be subjected to a surprise attack.²⁸ The Prophet stipulated that an invitation is required before fighting. For al-Mawardi, such an invitation includes the presentation of material intended to persuade the unbelievers of the truth of Islam: “making the Prophet’s miracles known to them, and informing them of such

arguments as would make them to respond favorably.”²⁹ In a style indicative of the conversational nature of Shari‘a reasoning, he notes that, should a commander (improperly) initiate fighting without making an attempt to persuade such an uninformed enemy about the truth of Islam, “he will owe them blood-money, which, according to the more correct view in the Shafi‘i school, is equal to that due to Muslims, although it has been suggested that it should be the same as the various amounts of compensation paid to the heathens according to their different beliefs. Abu Hanifa has, on the other hand, said: ‘There is no compensation for killing them, and their lives are to be taken freely.’”³⁰ One must cite precedents in Shari‘a reasoning; one need not follow each and all.

“Killing women and children is not permitted in war or otherwise so long as they do not fight, owing to the Prophet’s injunction against killing them . . . God’s Prophet . . . has also forbidden the killing of servants and slaves.”³¹ Here al-Mawardi acknowledges the authority of the Prophet’s directives delimiting the range of legitimate targets. He also moves toward the identification of a general principle that we might see as underlying the listing of groups that fighters are not supposed to target for direct attack. The point is not simply that the Prophet identified women, children, and others as protected groups. Rather, these persons are listed because, as a general matter, they “do not fight.” Indeed, if women or children do take up arms, they “should be fought and killed,” albeit in ways that maintain honor.³² In some ways al-Mawardi’s account of the limits placed on just killing is more stringent than that of al-Shaybani and his colleagues. Al-Mawardi argues that Muslim fighters faced with an enemy that shields itself with Muslims should stop killing; if the Muslim fighters find it impossible to continue without killing the Muslim shields, they should stop attacking. Should the enemy take advantage of this proscription, surrounding the Muslim army while still protecting itself with Muslim hostages, the Muslim fighters

should defend themselves, trying all the while to avoid killing their co-religionists; if one of the hostages is killed, the killer must pay blood money and/or make expiation, depending on whether he knew the religious identity of the one killed.³³

Al-Mawardi shows both the continuity and change characteristic of Shari'a reasoning about war. Subsequent writers provide further illustration. Shortly after al-Mawardi's death in 1058, for example, al-Sulami of Damascus (d. 1106) produced the *Kitab al-Jihad* (Book on Armed Struggle). Al-Sulami was not and is not widely known for his practice of Shari'a reasoning; he was evidently an instructor in Arabic language and letters at the Grand Mosque in his home city. Much of his book on *jihad* has to do with language intended to evoke action by Muslims, especially those in positions of authority, in response to incursions into the territory of Islam by *al-faranj*, the Franks—that is, the Crusader armies. Nevertheless, al-Sulami's mode of reasoning and above all his indication of the formal duty to fight indicates participation in the practice of Shari'a reasoning.³⁴

For our interests, it is particularly striking that al-Sulami's preoccupation is with defensive or "imposed" war. Al-Shaybani and other early Hanafi jurists seem entirely unconcerned about fighting in a context defined by the success of an external, non-Muslim enemy in occupying Muslim territory. For al-Shaybani and his colleagues, the focus of war with non-Muslims had to do with the ways Muslim fighters should conduct themselves during an incursion into enemy territory. Similarly, al-Mawardi's preoccupation is with the rights and wrongs of war aimed at expanding the territory in which Islam is established. His brief mention of the right of a governor "with limited powers" whose province is located on the border between the territory of Islam and the territory of war to fight non-Muslims if they strike first, and without obtaining the caliph's authorization, is exceptional.³⁵

With al-Sulami, we have a different idea, and a different context:

the Crusades. Specifically, he writes in response to the First Crusade, which by 1099 yielded Christian control of Jerusalem, as well as significant portions of greater Syria.³⁶ Al-Sulami argues that the duty to extend the territory of Islam, expressed at least through an annual expedition organized and led by the ruler, was established by the Prophet and the early caliphs. Subsequent rulers followed this custom, until a particular caliph (unnamed in the text) failed to carry through, either through simple neglect or for reasons of state. Thereafter others failed in the custom for similar reasons. The incursions of Franks into Islamic territory are a judgment of God, executed in response to this sin of omission. And now, al-Sulami says, Muslims find themselves in a distinctive situation, which affects how they construe the duty to struggle for justice. Citing al-Shafi'i, al-Sulami notes that the raising of an army for purposes of carrying out expeditions into enemy territory is the minimum obligation of the Muslim ruler, and that if sufficient forces are not raised, it becomes the duty of those "in the rear" to fulfill God's command. The duty of fighting (al-Sulami uses the term *ghazw*, raiding) is, in cases of necessity, incumbent upon all the members of the community. The current situation, in which enemy forces are making inroads into the territory of Islam, constitutes such a case. With further citations from al-Ghazali (d. 1111), the greatest scholar of his day, al-Sulami goes on to stipulate that if a town in, say, Syria is attacked by the Franks and cannot defend itself, all the other (Muslim) cities of the region are obligated to come to its aid. Supporting and extending his argument with additional Shari'a references, he concludes with the judgment that "the Book [the Qur'an], the example of the Prophet, and the consensus [of representative practitioners of Shari'a reasoning] with respect to the community's obligation to fight makes clear that in cases of necessity, this duty becomes one of the individual duties."³⁷ As the last reference makes clear, al-Sulami's judgment is that in the type of context he describes, the duty to fight is like the

duty of fasting. Only those unable by reason of sickness or other incapacity may be excused from its performance.

It seems unlikely that al-Sulami was the first to argue in this way. However, the precedents he cites do not say precisely that fighting becomes an “individual duty.” Rather, they suggest that those who are “distant” from the affected area are called to step forward and give aid. Just how much distance there is between this and al-Sulami’s judgment that the incursions of the Franks make fighting an individual duty is a matter in need of further exploration.

In any case, by the time one gets to the work of Ibn Taymiyya (d. 1328), reference to fighting as an individual duty in circumstances of emergency seems to come easily. If, he writes, the enemy attacks the Muslims, then the duty to fight becomes a personal or individual duty for both “those against whom the attack was made and those not directly affected by it.”³⁸ Citing the Qur’an and the example of the Prophet, this scholar, much of whose career was spent in the prisons of those ruling in Damascus and in Cairo, writes that every Muslim must provide assistance, even if he is not a professional soldier. In the battle of the Trench (627), when the Muslims in Medina came under siege, “God permitted none to abandon the jihad.” Once the siege was broken, and fighting involved pursuit of the enemy, a different set of judgments prevailed. Ibn Taymiyya writes that the defense of Medina “was a war to defend the religion, the family honour and the lives [of community members]; an obligatory fighting . . . [Pursuit of the enemy following the siege] was voluntary fighting to increase the prestige of the religion and frighten the enemy.”³⁹

Different circumstances suggest distinctive judgments about the nature of the duty to fight. Do they suggest distinctive rules? The answer, it seems, is no. Those who “do not constitute a defensive or offensive power, like the women, the children, the monks, old people, the blind and the permanently disabled should not be fought.”⁴⁰ This general rule is supported as the consensus of the Muslims, the exam-

ple of the Prophet, and Qur'an 2:190: "fight in the way of God against those who fight against you, but do not commit aggression. Truly, God does not love those who commit aggression." Ibn Taymiyya comments that "we should only fight those who fight us, if we really want the Religion of Allah to be victorious."⁴¹

Ibn Taymiyya does suggest that noncombatants may forfeit their protected status if they "carry on a kind of fighting with words" or serve as spies, provide transport for munitions, or in other ways participate more directly in the enemy war effort. Earlier writers had typically stipulated that women, children, and others moved to combatant status if and when they took up arms. The principle, in either case, seems to be that those who do not, or may be presumed not to, participate in the enemy's war effort are not to be the target of direct attack by Muslim forces.⁴²

This matter brings us to an emphasis that many commentators view as a special focus of Ibn Taymiyya's judgments about war. He writes that the "heaviest jihad should be directed against the unbelievers and those who refuse to abide by certain precepts, like the abstainers from paying zakat-alms and the Kharijites."⁴³ The case of the tax evaders has drawn attention from many students of Ibn Taymiyya's work.

Ibn Taymiyya's wording suggests the notion of fighting against apostates (*al-murtadd*) and rebels (*al-bughat*). Discussion of the rules governing fighting against these groups was an established practice of Shari'a reasoning. Al-Shaybani and his colleagues discussed the matter, as we have seen. Apostates were understood along the lines of the precedent attributed to Abu Bakr, discussed in Chapter 1. Having received a communication from a particular group regarding its intention to maintain faith in God and God's Prophet, but to forgo payment of *al-zakat* to the common treasury, Abu Bakr declared his intention to fight in order to compel obedience. *Al-zakat*, he said, was not optional for Muslims. Believers must contribute to the com-

mon treasury, administered by a legitimate authority. Those who refused to do so had “turned” (the literal meaning of *al-murtadd*) away from the faith, even if they continued to recite the basic creed and to perform obligatory prayers. Ibn Taymiyya is consistent with his predecessors on this matter, even in his discussion of “the preacher of heresy contrary to the teachings of the Book and the Sunna.”⁴⁴ In either case, the wrong has to do with public departure from Islam, as established and protected by a legitimate authority. Such wrongs are connected with Qur’an 2:214 and other texts that stipulate that *fitna*, or civil strife, is more blameworthy than the shedding of blood.

Similarly with rebels. The Kharijites, whom Ibn Taymiyya mentions explicitly, were an early group understood as “secessionists” (the literal meaning of the name *al-khawarij*). According to traditions consistently cited by practitioners of Shari’a reasoning, the group seceded from the rightful leadership of ‘Ali ibn Abi Talib (d. 661), the fourth successor to the Prophet. Their secession was motivated by an objection to ‘Ali’s approach to conflict with Mu’awiya, governor of Syria and an opponent of ‘Ali. In justifying their actions, the Kharijites cited Qur’anic texts. They understood themselves as advocates for justice and saw others as moral slackers, in need of encouragement or, in some cases, of punishment. ‘Ali responded with a mixture of persuasion and force, the goal being to limit the damage done by this group of “irregular fighters” while at the same time hoping for reconciliation. Over time the Kharijites became a standard reference for those who, while well-motivated, committed acts that did more harm than good. In justifying fighting against such persons, Ibn Taymiyya again was consistent with Shari’a precedent. It is the duty of a Muslim ruler to establish and protect the practice of Islam. Such a ruler does not use military force to make new converts. Rather, the ruler fights to extend or protect the hegemony of Islamic values in what one might call “geopolitical space.” Even within that space, fighting is not authorized to force conversion. As

Ibn Taymiyya writes, the protected peoples (*ahl al-dhimma*), for example Jews and Christians living under Muslim rule, may be fought only if they violate the terms of their treaty with the Muslim ruler. Their difference in religion is not itself a justification for fighting.⁴⁵

Force is authorized, however, to protect the establishment of Islam. So apostates and rebels are seen as a greater threat than those unbelievers who live outside the boundaries of the Islamic state. A legitimate ruler is authorized, even required, to use force to limit the ability of such groups to “spread corruption in the earth” (Qur’an 5:33).

Thus far Ibn Taymiyya seems consistent with his predecessors in the practice of Shari’a reasoning. He does write as one intent on purifying the practice of Islam; his treatise on the institution of *al-hisba*, or the public exhortation of Muslims to fulfill their duties, is indicative of this concern, as are numerous tracts and formal Shari’a opinions about the danger of practices that depart from true Islam.⁴⁶ For Ibn Taymiyya, any Muslim refusing to perform those duties or to observe those prohibitions established by Shari’a reasoning is subject to punishment. If such refusal is a matter of performance only, the miscreant may be considered an unbeliever, an apostate, or a “half Muslim”—this, Ibn Taymiyya writes, is a matter on which Shari’a authorities disagree. If refusal is accompanied by explicit statements to the effect that regulations established by Shari’a reasoning do not hold, the person is simply an unbeliever. In any case, however, punishment does not rise to the level of fighting unless departure from duty is accompanied by armed resistance. If such resistance is present, then punishment with fighting is necessary. Depending on the circumstance, participation in fighting to discipline those who have departed from their duties may be a matter for the standing army, or it may be an individual duty incumbent on each and every Muslim.⁴⁷

Ibn Taymiyya’s zeal for purity in Islamic practice may or may

not involve an expansion of the typical concerns of Shari'a reasoning. Where he does seem a bit different from his predecessors is in his construal of the relations between rulers and subjects. For Ibn Taymiyya, these are reciprocal. Rulers owe it to their subjects to provide leadership, to govern wisely, to establish institutions that will ensure that subjects understand their duties and are encouraged to fulfill them.

Subjects owe obedience; this is a standard judgment of Shari'a reasoning. At some times and in some places subjects are compelled to "omit to obey," as when a ruler commands something contrary to the Qur'an and the example of the Prophet. This again is standard. Ibn Taymiyya takes this thinking further, however; he says that it is a duty of subjects to remind rulers of the demands of obedience to God's law. Indeed, Ibn Taymiyya presumes that "commanding right and forbidding wrong" is a task in which all Muslims have a part. Certainly, not all Muslims are public officials with rights and duties associated with enforcement of the Shari'a. But all Muslims do have a share in the task of commanding good and forbidding evil, and this notion may be understood in ways that expand the rights of subjects in relation to rulers. For rulers, as well as subjects, may be negligent in their duties. Rulers, as well as subjects, may depart from Islam in ways suggestive of apostasy, rebellion, or unbelief. When this situation holds, who has the duty, and with it the right, of punishing the ruler? The establishment of legitimate authority, Ibn Taymiyya holds, is a requirement of the Shari'a. Who will establish such authority if and when those who hold the reins of power are themselves corrupt?

It is difficult to see Ibn Taymiyya as a full-blown advocate of just revolution. He wrote in a time when the unified caliphate presumed by his predecessors no longer existed. With the invasion of the Mongols, the Abbasid caliphate proved unsustainable, even in the special role envisioned by al-Mawardi. The Muslim state was

in a time of transition, with competition between various claimants to power. The Mamluk sultanate in Cairo was preeminent. The Mongols themselves were in transition; having conquered large portions of the territory of Islam, they appeared to be quintessential unbelievers. They were, however, in the process of converting to Islam, and by the time of Ibn Taymiyya they governed their portion of the territory of Islam by a kind of mixed legal regime. Judgments associated with the Shari'a were set alongside those characteristic of the Mongol code of honor. In this context, Ibn Taymiyya's concerns for pure Islamic practice seem to suggest the necessity of fighting against the Mongols, and in certain opinions he appears to say that such fighting is a duty for Muslims. Such judgments are put forth with care, however; much depends on the trajectory of the Mongol state with respect to adherence to Shari'a norms. Thus the transitional nature of authority in Ibn Taymiyya's time makes it difficult to read his judgments as constitutive of a theory of just revolution.⁴⁸

Sunni and Shi'i Perspectives

At this point it is worth pausing for a brief discussion of the differences between Sunni and Shi'i approaches to matters of politics and war. As we have seen, the Shi'i insistence that leadership rightly belongs to the designated leader or imam had implications for the practice of Shari'a reasoning—specifically, for the idea that the authority of *ahadith* needed careful examination, since one could not trust that people who rejected 'Ali and his successors would faithfully transmit the *sunna* of the Prophet. Trustworthy or "sound" reports were those transmitted by, or on the authority of, one or more of the imams. Practically speaking, this meant that the biographies of those mentioned as transmitters should be scrutinized, since even the inclusion of the name of a designated imam rested on the witness of a subsequent reporter. Indeed, in the case of Twelver Shi'ism,

a small group of community leaders, noted for learning and piety, came to serve as “deputies” of the Hidden Imam, and thus to function as authorities in his stead, albeit with certain restrictions tied to the fact that, despite their standing as the imam’s representatives, “they” were not “he.”

The role of the *‘ulama* as, in some sense, deputies of the Twelfth Imam is the topic of a wide-ranging study by Abdulaziz Sachedina, *The Just Ruler in Shi‘ite Islam*.⁴⁹ Sachedina shows that the precise dimensions of the authority attached to this designation are controversial. Nevertheless, it is clear that the Shi‘i *‘ulama* perceived themselves as advisers to rulers. In that capacity, they issued opinions intended to “bind and loose,” that is, to set limits upon and provide justifications for the exercise of political and military authority by those wielding power.

In connection with the Sunni precedents outlined in this chapter, the most important of these opinions have to do with limiting the ruler’s authority to initiate military action to instances in which war is “imposed” by the action, or threatened action, of an enemy. The great work of al-Muhaqqiq al-Hilli (d. 1277), for example, begins with a description of the rules of engagement which is compatible with the judgments offered by al-Shaybani, al-Mawardi, and others described above. Al-Hilli then makes clear, however, that he is describing the justification and conduct of fighting ordered by the designated leader, who, as the imam of his age, is protected from the commission of serious sin, and thus knows when war is a proper means of statecraft. In the absence of the designated leader, his deputies are not authorized to carry out wars in which the intention is to expand Islamic territory. The deputies’ authority is limited to situations in which the invasion or threat of invasion by a non-Muslim power provides a context for “defensive” war.

In such a context, the normal rules governing the conduct of Muslim forces hold. That is, those fighting in order to defend the terri-

tory of Islam are to avoid directly targeting noncombatants, and to use weapons and tactics that are proportionate to the task. If the progress of enemy forces raises the specter of an emergency, then the ruler, and with him the Shi'i *'ulama*, call for help from neighboring Muslim rulers.

The Shi'i trajectory, like the Sunni, thus establishes a set of precedents for the justification and conduct of war. The distinctively Shi'i emphasis on leadership has important implications for the notion of right authority, however. As in the case of Ibn Taymiyya's writings, the Shi'i emphasis on leadership is critical for an understanding of contemporary militancy, not least in the writings of the late Ayatollah Khomeini. Indeed, the logic of contemporary Muslim argument makes clear the role of precedent in Shari'a reasoning about politics and war. As a *living* tradition, Islamic judgments about the order of society and the uses of military force always respond to new situations. Yet both the shape of argument and the form of reasoning employed by militants and their critics alike rely upon interpreting and applying historical, textual precedents set by premodern scholars—or, even better, by the Qur'anic and Prophetic precedents to which these scholars referred. Shari'a reasoning is a system of argument about the fit that obtains between precedents and the facts of political life. And so Muslim arguments about war and politics in contemporary settings proceed just as they did in preceding centuries: by asking "What guidance has God provided for the conduct of life?" "What is the path that leads to refreshment" corresponding to the example of the Prophet and the true nature of human beings?⁵⁰



Why Is There so Much Conflict in the Middle East?
 Author(s): Mirjam E. Sørli, Nils Petter Gleditsch and Håvard Strand
 Source: *The Journal of Conflict Resolution*, Vol. 49, No. 1 (Feb., 2005), pp. 141-165
 Published by: Sage Publications, Inc.
 Stable URL: <http://www.jstor.org/stable/30045102>
 Accessed: 11/12/2013 09:01

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

 NOTICE:
 This material may be protected by copyright law (Title 17 U.S.
 Code)



Sage Publications, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Conflict Resolution*.

<http://www.jstor.org>

Why Is There So Much Conflict in the Middle East?

MIRJAM E. SØRLI

*Centre for the Study of Civil War
 International Peace Research Institute, Oslo (PRIO), Norway*

NILS PETTER GLEDITSCH

*Centre for the Study of Civil War, PRIO
 and*

*Department of Sociology and Political Science
 Norwegian University of Science and Technology*

HÅVARD STRAND

*Department of Political Science
 University of Oslo, Norway*

The Middle East is one of the most conflict-prone regions—but why? The Collier-Hoeffler model of civil war provides the starting point for our analysis. In an application to Africa, Collier and Hoeffler found poverty to be the most significant predictor of conflict. For conflict in the Middle East, a more complex picture emerges. Consistent with Collier and Hoeffler, the authors find that economic development and economic growth, in addition to longer periods of peace, generally decrease the likelihood of conflict. They also find that ethnic dominance is significant, while social fractionalization is not. Contrary to Collier and Hoeffler, they find that regime type matters. Variables for the Middle East region, Islamic countries, and oil dependence are not significant. Conflict in the Middle East is quite well explained by a general theory of civil war, and there is no need to invoke a pattern of “Middle Eastern exceptionalism.”

Keywords: *Middle East; conflict; Collier-Hoeffler model; civil war*

Conflict in the Middle East is a recurring feature in international politics, academic literature, and current news coverage. The fifty-five-year-old Israeli-Palestinian conflict is one of the most enduring conflicts anywhere, but over the past twenty-five years, the region has also hosted two of the wars with the most international participants (Iraq in 1991 and 2003), as well as the bloodiest interstate war of that period (Iran-Iraq, 1980-1988). The region is also surrounded by other long-term conflict zones: Afghanistan, the Caucasus, the Horn of Africa, and Sudan. Internal and regional instabilities have combined with the close ties between Middle Eastern and arms-producing governments to make the Middle East the most militarized region in the world (Bureau of Verification and Compliance 2000; Sköns et al. 2002). In the extensive literature on international water disputes, the Middle East figures very

JOURNAL OF CONFLICT RESOLUTION, Vol. 49 No. 1, February 2005 141-165
 DOI: 10.1177/0022002704270824
 © 2005 Sage Publications

prominently (see, e.g., Lonergan 1997). The Middle East lags behind in economic, social, and political development (United Nations Development Program [UNDP] 2002). Reference is frequently made to "Middle Eastern exceptionalism"—that is, that there is something unique about the Middle East that makes the region prone to conflict, autocracy, and economic misery (see, e.g., Rubin 2002).

Since the 1950s, civil war has been the dominant type of conflict. In 2003, twenty-seven out of twenty-nine armed conflicts with more than twenty-five battle-related casualties were domestic and internationalized civil wars (Eriksson and Wallensteen 2004). In this study, we use a modified version of the model of civil war developed by Collier and Hoeffler (1998, 2002, 2004) to analyze the onset of civil conflict globally and in the Middle East region. Like Collier and Hoeffler, we restrict our analysis to intrastate armed conflict.

We first show that the Middle East indeed is a conflict-ridden region but that Asia and Africa are now equally or even more conflict prone. We compare the Middle East to other regions and find it to be characterized by authoritarian regimes, oil-dependent economies, Islam, and the protracted Israeli-Palestinian conflict. We discuss some contending approaches to the study of civil war. For a long time, the language of grievance dominated the civil war research agenda. Recently, more focus has been paid to the economic opportunity for rebellion. The Collier-Hoeffler model (subsequently called the CH model) has gained considerable prominence in the empirical study of civil war and has served as a basis for a widely publicized World Bank report on civil war (Collier et al. 2003). A modified version of the model used in their work on conflict in Africa (Collier and Hoeffler 2002) provides the design for the multivariate analysis. We next describe the model and the variables before presenting our empirical results and discussing their interpretation. Consistent with Collier and Hoeffler, we find that economic development and economic growth, in addition to longer periods of peace, generally decrease the likelihood of conflict. We also find that ethnic dominance is significant, while social fractionalization is not. In contrast to Collier and Hoeffler, we find that regime type matters. Variables for the Middle East region, Islamic countries, and oil dependence are not significant. Conflict in the Middle East is quite well explained by a general theory of civil war, and there is no need to invoke a pattern of "Middle Eastern exceptionalism."

CONFLICTS BY REGION AND TIME

Despite the prevalence of the image of the Middle East as conflict ridden, this point needs empirical scrutiny. Is the Middle East really more prone to conflict than other

AUTHORS' NOTE: An earlier version of this article, which draws on Sørli (2002), was presented to the 43rd Annual Convention of the International Studies Association, New Orleans, LA, March 24-27, 2002, and Sørli records her gratitude to the discussant and the participants in that meeting. We are also grateful to Scott Gates, Håvard Hegre, Pålvi Lujala, and Gudrun Østby, as well as the referees of JCR for their comments on this version. The data used in the analysis, as well as several unpublished appendices, are available at www.prio.no/cwp/data.asp and at www.yale.edu/unsy/jcr/jcrdata.htm. The authors' work was supported by grants from the Research Council of Norway.

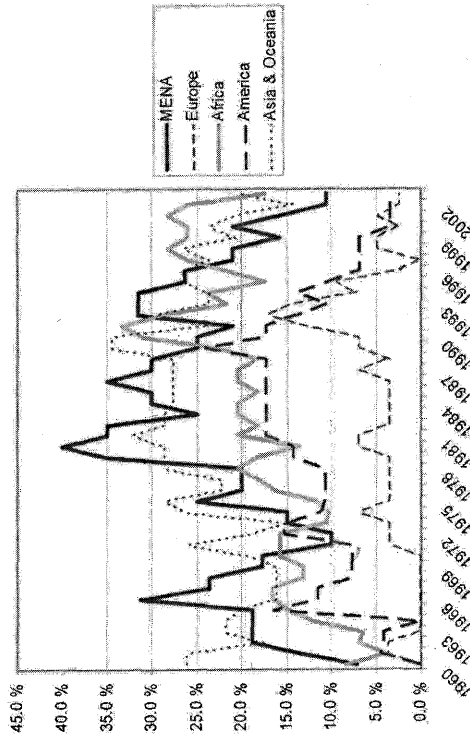


Figure 1: Relative Incidence of Armed Civil Conflict by Region, 1960-2003

SOURCE: PRIO/Uppsala conflict data (Gleditsch et al. 2002; Eriksson and Wallensteen 2004).

NOTE: In tallying the incidence of conflict, we include all years with ongoing conflicts but only for the conflict theater itself. We do not include outside countries that intervene in the conflict. We define the Middle East as follows: Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, Turkey, United Arab Emirates, and Yemen (including North and South Yemen). See Appendix 4 (online) for the definition of all regions.

regions? The distribution of conflict per region in Figure 1 is based on the International Peace Research Institute (PRIO)/Uppsala conflict data set (Gleditsch et al. 2002; Eriksson and Wallensteen 2004). The analysis includes all internal armed conflicts (including internationalized internal conflicts) with more than twenty-five battle-related deaths per year. The figure shows that, relative to the number of countries in each region, the Middle East has competed with Asia for the position of the most conflict-ridden region for the better part of the period. However, after the end of the cold war, this is no longer true. The incidence of conflict in the Middle East has declined from the end of the 1980s, as is the case for other regions. The Middle East now has a slightly lower incidence of conflict than Africa and Asia. Lacina and Gleditsch (2004) find that conflicts in the Middle East are usually not among those with the highest fatalities. The one exception is an interstate war, the Iran-Iraq War, which made the Middle East the bloodiest region in the 1980s.

The relationship between armed civil conflict and an explanatory variable is often endogenous. As a result of war, political institutions and economies are weakened, ethnic minorities might flee, and so forth—in brief, the presence of an armed conflict influences the main variables used to explain conflict, with the exception of geographical variables. Ongoing conflict also influences the quality and availability of data.

TABLE 1
The Regional Distribution of Armed Civil Conflict, 1960-2003

	The Americas		Europe and the Caucasus		Sub-Saharan Africa		Middle East and North Africa		All Asia Countries	
Number of armed civil conflicts	21	24	56	22	39	162				
Percentage of country years (onset)	1.6	1.7	3.2	2.7	3.2	2.6				
Percentage of country years (incidence)	11.8	4.4	19.5	23.6	24.3	15.9				

NOTE: Numbers based on the PRIO/Uppsala conflict data set (Gleditsch et al. 2002; Eriksson and Wallensteen 2004). Intervening powers not included.

Gathering reliable statistical data is rarely the top priority of governments involved in armed conflict. These are two strong arguments for analyzing the onset (outbreak) of a new civil conflict rather than its incidence (or prevalence).¹ Also, the *incidence* of civil war conflates onset and duration. Many scholars argue that the causes of civil war duration are different from the causes of civil war onset (see Hegre 2004). Table 1 displays the regional pattern of conflict up to 2003.

In terms of conflict onsets, the Middle East does not stand out as the most conflict-prone region (see Appendix 5 [online] on our replication data page for annual data on conflict onsets by region). The average onset of conflict in the period from 1960 to 2003 in the Middle East is higher than the Americas and Europe/Caucasus but lower than sub-Saharan Africa and Asia. The incidence of conflict is nevertheless almost as high as Asia and higher than any other region because conflicts in the Middle East tend to last for a long time.

In Appendix 2, we list all the conflicts in the Middle East ongoing in the period from 1960 to 2003. A variety of conflicts can be observed. Many are between a government and militant Islamists (Algeria 1991-2003, Egypt 1992-1998, Saudi Arabia 1979, Syria 1979-1982, and others). Several countries have also fought military factions (Iraq in 1963, Morocco in 1971, and Syria in 1966). In Iran, Iraq, and Turkey, changing governments have fought Kurdish separatists—since the early 1960s in the Iraqi case. The conflict between Israelis and Palestinians is persistent but mostly limited to Israel. For more than a decade (1975-1989), Morocco fought secessionists in former Spanish Sahara, and the various unifications of Yemen have also been followed by fighting between the government and secessionists (1962-1970, 1986, 1996). But what general patterns can be found in this variety?

THEORIES OF CIVIL WAR

Accounts of civil war have frequently referred to ethnic and religious hatred and to economic, political, and social discrimination. In the recent turn away from the tradi-

tional grievance-based rhetoric, some researchers suggest that greed (or opportunity) serves as a better explanation for civil conflict: "The true cause of much civil war is not the loud discourse of grievance, but the silent force of greed" (Collier 1999, 8). Following Gurr (1970) and Ellingsen (2000), we see the outbreak of armed civil conflict as a combination of *frustration* (repression, suffering), *opportunity* (enough freedom to organize; access to finances, weapons, and soldiers), and a common *identity* (cohesion facilitates mobilization). Although we feature the CH model prominently, we include in the model variables relevant to all three general factors in civil war.²

The CH model provides valuable insight into the role of economic factors in civil war. A substantial body of literature deals with the relationship between civil war, economic growth, and natural resources (Addison, Le Billon, and Murshed 2002; Auty 2001; de Soysa 2002; Le Billon 2001). This literature emphasizes the crucial importance of access to finances in the form of foreign support for the insurgency, the availability to raise revenue through the extraction of resource rents, or indirectly through taxing goods passing through rebel-controlled territory. With the end of the cold war, foreign support for rebel groups has decreased, and such movements have increasingly been forced to generate their own revenue. In this new conflict environment, the line between civil war and crime has become increasingly blurred.

In brief, Collier and Hoeffler (1998, 2002, 2004) present two partially competing perspectives on civil war: high levels of grievance and opportunities for forming a rebel organization. Their analytical model focuses mainly on conditions that favor the formation of rebel organizations, and their econometric model predicts the probability of a civil war being initiated in a country during a five-year period. While both the opportunity and grievance factors are investigated, Collier and Hoeffler (2004, 25) conclude that "a model that focuses on the opportunities for rebellion performs well, whereas objective indicators of grievance add little explanatory power."

While economic inequality, political rights, ethnic polarization, and religious fractionalization were insignificant, natural resource dependence emerged as a significant factor in the CH study. Economic development and access to finance, in the form of foreign contributions or control over natural resources, provide the backbone of the opportunity explanation for civil war. A short time since last conflict always increases the risk of a new conflict. In their work on Africa, Collier and Hoeffler (2002) conclude that the main source of conflict in sub-Saharan Africa is the region's poor economic performance. In contrast to several empirical studies, Collier and Hoeffler find no effect of regime type. We reexamine this finding. We also question the usefulness of their natural resource dependence variable. In our analysis, we therefore extend and modify the CH model.

MIDDLE EAST EXCEPTIONALISM?

Middle East exceptionalism is mostly used to denote the region's resistance to democracy but also includes the region's lack of success in economic and social devel-

2. A useful recent review of the civil war literature is found in Sambanis (2002).

opment, a prominent example being the position of women, in addition to the persistent state of conflict (Kedourie 1994; Kramer 1996; Aarts 1999; Rubin 2002). Can conflict in the Middle East be explained by factors that are unique to the Middle East? Or is the Middle East just an unfortunate region where conflict risk factors have accumulated?

The borders of the modern Middle East region were drawn up by the victorious powers of the First World War, as laid out in the 1916 Sykes-Picot agreement. The former core areas of the vast Ottoman Empire became the state of Turkey. The Armenian provinces were given to Russia, Lebanon and most of Syria to France, and modern-day Iraq and the rest of Syria to Britain. France controlled Morocco, Algeria, and Tunisia, while Britain was in charge of Egypt, Palestine, Jordan, and southern Yemen. By the mid-1960s, the people living in the areas drawn up as states or mandates by the Sykes-Picot agreement had achieved their independence.

The establishment of the state of Israel in 1948 has been the source of considerable regional conflict and instability, and the protracted conflict has hampered political and economic development in the entire region. Due to its geographic location and resources (notably oil), the Middle East is generally considered to be of great strategic importance to the main players in world politics. The political economy of the Middle Eastern states has given them a particularly strong level of integration in the world economy. There is also a uniquely close relationship between Israel and the United States. Great-power bickering and external intervention continue to influence the region today, with the U.S.-U.K. invasion and occupation of Iraq in 2003 as just one of several examples. We will proxy Middle East exceptionalism by using a Middle East dummy in our empirical analysis. There is no compelling theoretical reason for including such a variable. It simply measures whatever factors that are peculiar to the Middle East that we have been unable to measure more directly. The empirical analysis will show that, in fact, there is no reason to posit any such regional exceptionalism.

POLITICAL INSTITUTIONS

The Middle East region remains the world's authoritarian stronghold and has yet to experience a wave of democratization (Huntington 1991). Of the region's nineteen states, only Israel and Turkey qualify as more or less democratic. Israel obtains maximum scores on the Polity index in the entire period (1960-2000). On the Polyarchy index, which includes participation, it scores from 28 to 37 (the maximum is 47). Turkey's score roller-coasts from very low to very high on both indices. The democratic deficit in the Middle East resides in a complex set of circumstances. There has been extensive debate about the role of Islam. Fish (2004) finds, in a multivariate cross-national analysis, that Islam is robustly associated with autocracy and attributes this mainly to the subordination of women in Islamic states (although this is not required by the Koran). Donno and Russett (2004), on the other hand, find that this is more characteristic of Arab states than of Islamic countries generally. The colonial experience, continued foreign influence, and persistent conflict have not encouraged democratic tendencies either. Yet, other countries have broken out of such postcolonial authoritarian-

ism. More important, in most of the Middle Eastern countries, the basis of the economy has not produced social and economic development with modernism and democracy as favorable by-products. In particular, heavy dependence on oil production is not conducive to democracy (Ross 2001).

ECONOMIC DEVELOPMENT

The main sources of economic income and activity in the region stem from natural resources, workers' remittances, foreign aid, agriculture, and small businesses. Industry plays a minor role in most Middle Eastern economies, and economic development has generally been state driven (Owen and Pamuk 1999). The role of natural resources is particularly important to an analysis of the causes of civil war. While some researchers emphasize the importance of resource scarcity to conflict (Homer-Dixon 1999), others argue that resource abundance drives conflict because control of natural resources provides a source for financing rebellions and personal wealth for rebel leaders (Collier and Hoeffler 2002; Le Billon 2001). Along the same lines, rentier-state theory argues that resource-dependent states are likely to develop authoritarian political systems and economies with low diversity (Beblawi and Luciani 1987). Sachs and Warner (1995) found that resource-abundant economies grow at a slower rate than resource-poor ones. A combination of population growth and an increasing portion of people with higher education has led most Middle Eastern governments to swamp their bureaucracies with new graduates. Further hampering economic development is the high military spending, largely funded by oil income.

RELIGION

The birthplace of three world religions, the Middle East today is overwhelmingly Muslim. Much conflict and friction has occurred between the two major strands of Islam: Sunni (about 85 percent) and Shia Islam. Iran is the only Shia Islamic republic. Shia Muslims constitute the majority of the populations in Iraq and Bahrain, and significant Shiite minorities are found in Lebanon and Syria. Sunni and Shia each have several sects and subbranches. Inter-Islamic rivalry and conflict has been present since the religion's early history. Several countries have experienced internal Shia-Sunni conflicts (Lebanon, Iraq, Syria), and the Shiite Iranian regime is feared by its Sunni Islamic neighbors. Jews constitute 80 percent of the population in Israel, while Christians constitute significant minorities in Lebanon, Syria, Israel, Egypt, and Iraq.

Islam and Islamic law (Sharia) are derived from the teachings of the Koran and the practice of the Prophet Mohammed (Sunna). Muslims vary in their rigidity of adherence to Islamic teaching, in the same way as Muslim states vary in the degree of application of Sharia law. Consequently, very diverse expressions of Islam are found around the world.

The role of Islam in political and organizational life—variously called Islamism, Islamization, fundamentalist Islam, and political Islam—has increased in the period

that we study. In 1928, Hassan al-Banna founded the first Islamist group, the Muslim Brotherhood, in Egypt. However, it was not until the 1950s that the Brotherhood and other Islamist organizations gained popularity. Secular ideologies have generally failed in the region. Various versions of radical socialism failed in the 1950s and 1960s, and pan-Arabism did not fare much better in subsequent periods. Since the late 1970s, Islamic groups have gained greater support and have become more potent political and social players. Woltering (2002) suggests that there is little about Islam per se that accounts for the growth of political Islam and that one should look to political, social, and economic explanations. In some areas, Islamist groups have supplemented or even replaced the government as social service providers. An important membership base has been the masses of the educated unemployed. Frustration on a personal as well as political level—a continued sense of humiliation through covert and direct foreign support for repressive regimes as well as direct foreign intervention—has left many disillusioned with their own governments, as well as those of the major Western powers. This makes support for an Islamic solution an attractive option, if not the only one.

RESEARCH DESIGN

Collier and Hoeffler (1998, 2002, 2004) used a five-year period design to predict the probability of conflict breaking out in each period. This approach can be seen as a discrete-time duration analysis of peace periods (Beck, Katz, and Tucker 1998, 1261f.). All duration analyses raise the problem of time-varying covariates: some variables, such as regime type, can change over the duration being studied, so that neither the value at the start or the end of the period is representative for the whole period. The most common solution to this problem is to include more and shorter time periods (Box-Steffensmeier and Jones 2004, 95ff.). Most quantitative analyses on the causes of civil war are conducted on the basis of country-years, a short enough basis to ensure that time-varying covariates are representative. Therefore, we apply the CH model to an annualized data set. In addition to changing the structure of the data set, we use a different threshold for conflict. Instead of the threshold of 1,000 battle-related deaths required by the Correlates of War project, we include conflicts with a threshold of 25 battle-related deaths per year from the PRIO/Uppsala conflict data set. Due to data constraints for some of the explanatory variables, the statistical analysis is limited to the period from 1960 to 2000.

STATISTICAL MODEL

We employ a logistic regression model in our analysis. Logistic regression analysis is appropriate when the dependent variable is dichotomous and skewed. However,

3. On the basis of simulations, King and Zeng (2001, 157) conclude that these biases should be addressed when the distribution on the dependent variable is at least 5 to 95 percent.

when the dependent variable is very skewed,³ standard logistic regression models can report biased estimates (King and Zeng 2001). We correct for this bias by using King and Zeng's (2001) RELOGIT procedure. STATA Version 8.2 (StataCorp 2003) was used to run the analyses.⁴ We use a time-series, cross-section data set in which the assumption of independence between observations is rarely met. We employ the Beck, Katz, and Tucker (1998) method for the control of time dependence, using cubic splines.

DEPENDENT VARIABLE: ARMED CIVIL CONFLICT

The dependent variable is internal and internationalized internal conflict. We will use the term *armed civil conflict* to denote both types. An armed conflict is defined by the PRIO/Uppsala conflict data as "a contested incompatibility which concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" (Gleditsch et al. 2002, 618).⁵ Internal conflict "occurs between the government of a state and internal opposition groups without intervention from other states." Internationalized internal conflict "occurs between the government of a state and internal opposition groups with intervention from other states" (Gleditsch et al. 2002, 619). Onset is defined as the eruption of a new conflict (a unique ID in the data set) and a change in the conflict's sub-ID.⁶ The dependent variable is censored in the remaining years of the conflict since a country in a continuing state of civil war cannot experience a transition from peace to war. This censoring includes cases in which the onset occurred before 1960. The variable is coded 1 for all onsets and 0 otherwise.

In the period analyzed in this article (1960-2000), 168 armed civil conflict onsets were recorded. Armed civil conflict remains a rather rare event that occurred in only 3.4 percent of all country-years.⁷

4. The RELOGIT procedure is not a maximum likelihood estimator, so we do not report log-likelihood values.

5. A state is defined as an internationally recognized sovereign government controlling a specified territory or an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.

6. A sub-ID is by default 0 and changes when the conflict has fewer than twenty-five battle-related deaths over a period of ten years; the conflict type changes from internal to internationalized internal and when the opposition changes completely (Strand, Wilhelmsen, and Gleditsch 2003, 3).

7. The PRIO/Uppsala data set contains several conflicts that broke out at a time when there was already a conflict going on in the same country (e.g., in large countries such as India and Indonesia). Simultaneous conflicts pose a methodological challenge since studies of armed conflict onset usually study transitions from peace to war, omitting consecutive years of war. Urdal (2002) coded two different variables, "Onset1" (armed conflicts that erupted in a state of peace) and "Onset2" (all armed conflict onsets regardless of whether the conflict broke out when another conflict was taking place in the same country). The difference between the two coding schemes did not influence Urdal's results much. To follow the Collier-Hoeffler (CH) design, simultaneous conflicts are censored. This analysis includes all outbreaks of (new) armed civil conflicts as defined by the PRIO/Uppsala data set's conflict ID and sub-ID.

INDEPENDENT VARIABLES

The CH model includes gross domestic product (GDP) per capita, economic growth, natural resource dependence, ethnic dominance, social fractionalization, time since last conflict, population size, and geographical dispersion. We measure income and economic growth from the World Bank's real purchasing power parity (PPP)-adjusted GDP per capita, log-transformed. Natural resource dependence is proxied in three ways: data from *World Development Indicators* (World Bank 2002) are used to calculate the ratio of primary commodity exports to GDP. The same source was used to create the oil dependency and mineral dependency variables: the value of oil and mineral export to GDP, with the cutoff point set to 40 percent. We subtract the mean of oil and mineral dependence to minimize collinearity (Friedrich 1982, 822). The dummy for ethnic dominance is 1 if a single ethnolinguistic group makes up 45 to 90 percent of the total population and 0 otherwise, based on data from *Atlas Narodov Mira* (Department of Geodesy and Cartography 1964). This variable was copied and extrapolated from the original CH data set. We created a slightly modified variable, *ethnic dominance*, *Islam*. The eleven countries with significant Shia minorities were changed from 1 to 0 on ethnic dominance. Other religions could, of course, also have been subdivided, but this would have been a major project of its own. We gave priority to a more realistic coding of Islam as the dominant religion in the Middle East. Social fractionalization is a combined measure of ethnic and religious fractionalization. Following Collier and Hoeffler (2002, 26-7), ethnic fractionalization is measured by the ethnolinguistic fractionalization index, which measures the probability that two randomly selected individuals from a given country do not speak the same language. Data are only available for 1960. Using data from Barrett (1982) on religious affiliation, Collier and Hoeffler constructed an analogous religious fractionalization index. Following Barro (1997), Collier and Hoeffler also aggregated the various religious affiliations into nine categories: the fractionalization indices range from 0 (*completely homogeneous*) to 100 (*completely fractionalized*). The social fractionalization index is calculated as the product of ethnolinguistic fractionalization and religious fractionalization. Data for total population in each country were taken from the World Bank's *World Development Indicators* from various years. The variable was logged. For their measure of geographical dispersion, Collier and Hoeffler generated a Gini coefficient of population dispersion for each country, set to 0 if the total population is evenly distributed around the country and 1 if it is concentrated in one area. Data were only available for 1990 and 1995. For earlier years, we copied the 1990 value, and for 1991 to 2000, we used the 1995 value. We added two variables measuring regime type: the Polyarchy index of democracy, developed by Vanhanen (2000), ranges from 0 (*least democratic*) to 47.11 (*most democratic*). Polyarchy is a multiplicative index combining measures of political competition and participation, defined as in Dahl (1971). In additional analyses, not reported in table form, we used the Polity measure of democracy (Jaggers and Gurr 1995).

Following Beck, Katz, and Tucker (1998), we control for temporal dependence through a variable measuring time since last conflict and cubic splines. Time since the last conflict counts the number of whole years since the end of the last conflict in that

country. The spline variables were constructed with the BTSCS program (Tucker 1998).

We follow the Correlates of War (2002) country numbers in our definition of regions. We define a country as Islamic if it is a current member of the Organization of the Islamic Conference (OIC). The OIC was established in 1969, but all current members are assigned the value 1 and nonmembers the value 0 for the entire period. All current members had a substantial Muslim population before the OIC was established. The following countries are members of the OIC: Afghanistan, Albania, Algeria, Azerbaijan, Bahrain, Bangladesh, Benin, *Brunei*, Burkina Faso, Cameroon, Chad, Comoros, Ivory Coast, Djibouti, Egypt, Gabon, Gambia, Guinea, Guinea-Bissau, Guyana, Indonesia, Iran, Iraq, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lebanon, Libya, Malaysia, *Maldives*, Mali, Mauritania, Morocco, Mozambique, Niger, Nigeria, Oman, Pakistan, *Palestine*, Qatar, Saudi Arabia, Senegal, Sierra Leone, Somalia, Sudan, *Suriname*, Syria, Tajikistan, Togo, Tunisia, Turkey, Turkmenistan, Uganda, United Arab Emirates, Uzbekistan, and Yemen. Countries in italics were not included in the data set.

For a more detailed description of the independent variables, see Appendix 3 on our data Web page.

RESULTS

First, we present the general result of the CH model of civil conflict applied to an annualized data set and with some additional variables included. We then address conflict in the Middle East and North Africa region. Several variables are introduced to better capture the religious difference in the region, the importance of regime type, and more detailed data on natural resources. Collier and Hoeffler (2002) found that wealth, economic growth, social fractionalization, and longer periods of peace reduced the propensity for conflict (see Appendix 1 for their original table). Primary commodity dependence and a large population increased the risk for conflict. Ethnic dominance, geographic dispersion, and the sub-Saharan Africa variables were positive but not significant.⁸

Table 2 presents our annual analysis with the lower threshold for conflict and our specifications of the variables. Moving to an annualized design using a different conflict data set, we find a slightly different picture of the causes of civil war. Globally, high levels of economic development and growth and longer periods with peace decrease the likelihood for conflict, consistent with what Collier and Hoeffler (2002) found. We find ethnic dominance to be significant (at the 10 percent level), while social fractionalization is not. A high score on social fractionalization implies a large number of ethnic, linguistic, and religious groups within the country. According to Collier and Hoeffler, high fractionalization should impede mobilization to conflict. This is not supported by our findings. That the presence of a dominant ethnic group

8. However, Collier and Hoeffler (2004, 39) generally find ethnic dominance to be significant in their combined opportunity and grievance model.

TABLE 2
Accounting for Armed Conflict 1960-2000:
The Collier-Hoeffler Model with PRIO/Uppsala Data

	β	Standard Error	p	dPr (Onset) ^a
In GDP per capita	-0.37	0.082	< .0005	-0.035
GDP growth	-7.78	2.38	.001	-0.026
Primary commodity exports	0.072	2.53	.977	-0.0073
Primary commodity exports squared	1.35	4.33	.755	
Social fractionalization	0.000011	0.000058	.852	0.00082
Ethnic dominance	0.32	0.23	.158	0.0062
In Population	0.16	0.084	.063	0.015
Geographical dispersion	0.99	0.57	.082	0.011
Years in peace	-0.13	0.10	.204	
Spline 1	-0.0014	0.0014	.300	
Spline 2	0.00061	0.00089	.498	
Spline 3	-0.000056	0.00023	.814	
Constant	-2.93	1.33	.028	

NOTE: $n = 3,601$. Primary commodity exports and primary commodity exports squared are represented together. See King and Zeng (2001) for details on the computation algorithm used. GDP = gross domestic product.

a. Shows the change in probability for observing onset in a given year when each variable changes from the 5th to the 95th percentile, holding all other variables at their mean values.

yields a higher risk of conflict is in line with findings in Gurr (1993). Hegre et al. (2001) find a weak correlation between fragmentation and civil war, while Ellingsen (2000, 228) finds that "multiethnicity does increase the propensity for domestic violence." However, in both cases, political and socioeconomic factors were found to be more important in predicting conflict: political and economic factors influence the centrality of ethnic identities, play a major role in the possibility for conflict mobilization, and affect whether conflict escalates to a violent level.

Primary commodity dependence is not significant in our analysis. The natural resource variable lumps together all kinds of natural resources, ranging from beverages and food to oil, metals, and diamonds (Collier and Hoeffler 2002, 16). Several scholars have argued that the variable is too wide to measure conflict-related natural resources very meaningfully (de Soysa 2002). We therefore break down natural resource dependence into dependence on oil and minerals.

WHAT CAUSES CONFLICT IN THE MIDDLE EAST?

Collier and Hoeffler (1998, 2002, 2004) demonstrate that poverty is the most important element explaining conflict in sub-Saharan Africa. What explains conflict in the Middle East? We examine the descriptive statistics to see if the Middle East region scores systematically differently on one or more variables compared with sub-Saharan Africa and the global average (see Table 3).

TABLE 3
Descriptive Statistics, 1960-2000

	All Countries			Africa			Middle East			
	n	Mean	SD	Min.	Max.	n	Mean	SD	Min.	Max.
Gross domestic product (GDP) per capita	5,250	7.42	1.58	4.34	10.94	1,595	6.07	0.88	4.44	9.05
GDP cap growth	5,168	0.019	0.070	-0.722	1.55	1,556	0.090	0.068	-0.40	0.67
Oil dependence	3,082	0.00	0.109	-0.0349	1.10	710	0.00071	0.12	-0.035	1.05
Oil dependence squared	3,082	0.012	0.068	0.00	1.22	710	0.014	0.070	0.00	1.10
Oil dependence (dummy)	3,082	0.025	0.156	0	1	710	0.035	0.1844	0	1
Mineral dependency squared	3,276	0.00	0.038	-0.0142	0.40	799	0.013	0.060	-0.014	0.398
Mineral dependency	3,276	0.0014	0.0082	0.00	0.16	799	0.015	0.015	0.00	0.158
Social fractionalization	4,972	1,811.09	1,934.96	4	6,975	1,530	3,537.82	1,988.69	20	6,975
Ethnic dominance, Islam	5,077	0.46	0.50	0	1	1,530	0.36	0.48	0	1
Islam (dummy)	5,802	0.33	0.47	0	1	1,651	0.50	0.50	0	1
Logged population	5,652	8.94	1.55	4.81	14.06	1,625	8.42	1.27	5.38	11.62
Geographical dispersion	4,817	0.60	0.20	0	0.97	1,530	0.58	0.18	0	0.86
Polyarchy	5,159	0.00	12.81	-10.30	36.81	1,408	-7.23	5.57	-10.30	17.36
Polyarchy squared	5,159	164.11	206.00	0.0019	1,354.75	1,408	83.30	42.70	0.035	301.26
Middle East (dummy)	5,802	0.13	0.34	0	1	1,490	0.18	0.14	0	0.090
Primary commodity exports/GDP	4,982	0.15	0.15	0.0020	1.20	1,490	0.18	0.14	0	0.090

The Middle East is home to some of the richest as well as the poorest countries in the world but is, on average, much better off than sub-Saharan Africa and slightly better off than the global average. However, economic development in the Middle East has been oil dependent and state driven and has not followed the modernization path of economic diversification, occupational specialization, and mass education (Inglehart 1997). The Middle East is the world's most oil-rich region (but does not rank as high in minerals).

Middle Eastern countries stand out as extremely homogeneous, while African countries are highly fractionalized. Contrary to Collier and Hoeffler's (2002) work on Africa, we find that ethnic dominance increases the risk of conflict. The Middle East scores higher than Africa and the global average on ethnic dominance. However, as originally coded, this variable does not account for the divisions within Islam. When we use our recoded variable (ethnic dominance, Islam), where countries with significant Shia minorities are given a 0, the Middle East scores much lower on ethnic dominance than Africa and the global average.

Authoritarian and democratic regimes are generally less prone to conflict than regimes with middle-range scores (Hegre et al. 2001; Fearon and Laitin 2003). Both Africa and the Middle East score significantly less democratic than the global average. Populous countries are more prone to conflict, mainly because larger populations are more likely to host more sources of conflict. Middle East countries are generally small, and population size does not account for conflicts in this region. We now add modified ethnic dominance, regime type, and natural resource dependence to the analysis.

A REVISED MODEL

In Table 4, we examine the three variables proxying Middle East exceptionalism. To the CH model, we have added dummy variables for the countries in the Middle East region, Islamic countries, and countries where oil constitutes more than 40 percent of total exports. The Middle East dummy is not significant (model 2), so the region as a whole is neither more nor less conflict prone than the rest of the world. This corresponds well with Table 2. We then examine other features of the suggested Middle East exceptionalism.⁹

IS IT ISLAM?

Are some religions more aggressive and conflict prone than others? Huntington's (1996, 258) clash of civilization thesis predicts increasing conflicts between Islam and the West, claiming that "Islam's borders are bloody, and so are its innards." According to Huntington, Muslims are involved in a disproportionate percentage of violent con-

9. As a check on the robustness of the results, we have run the same model with incidence of armed civil conflicts as the dependent variable instead of onset. The Middle East dummy variable is now positive and significant, confirming that conflicts in the Middle East tend to last longer. The other results are quite similar.

TABLE 4
Accounting for Armed Conflict 1960-2000:
Revising the Collier-Hoeffler Model

	Model 1	Model 2	Model 3	Model 4
GDP per capita	-0.37*** (-4.52)	-0.37*** (-4.37)	-0.34*** (-3.99)	-0.45*** (-4.58)
Economic growth	-7.78*** (-3.27)	-7.72*** (-3.32)	-7.71*** (-3.34)	-10.96*** (-4.81)
Natural resource dependence	0.072 (0.03)	-0.035 (-0.01)	0.10 (0.04)	1.63 (0.55)
Natural resource dependence squared	1.35 (0.31)	1.07 (0.26)	0.94 (0.21)	2.51 (0.58)
Social fractionalization	-0.000011 (-0.19)	0.000019 (0.32)	-0.000075 (-0.13)	-0.000058 (-0.74)
Ethnic dominance	0.32 (1.41)	0.31 (1.38)	0.35 (1.55)	0.51* (1.83)
In Population	0.16* (1.86)	0.14* (1.73)	0.17** (2.03)	0.35*** (2.62)
Geographic dispersion	0.99* (1.74)	0.97* (1.73)	0.83 (1.42)	0.49 (0.69)
Middle East dummy		0.55 (1.64)		
Islam dummy			0.34 (1.37)	
Oil dummy				-1.14 (-1.04)
Peace-years	-0.13 (-1.27)	-0.13 (-1.32)	-0.13 (-1.31)	-0.0089 (-0.07)
Spline 1	-0.0014 (-1.04)	-0.0015 (-1.10)	-0.0015 (-1.11)	-0.00056 (-0.33)
Spline 2	0.00061 (0.68)	0.00065 (0.73)	0.00066 (0.74)	0.00026 (0.22)
Spline 3	0.000056 (0.24)	0.000049 (0.21)	0.000046 (0.19)	0.000044 (0.13)
Constant	-2.93** (-2.20)	-2.90** (-2.17)	-3.30** (-2.42)	-4.63** (-2.38)
n	3,601	3,601	3,601	2,272
Number of conflict onsets	89	89	89	58
Sample average onsets (percentage)	2.5	2.5	2.5	2.6

NOTE: Robust z-values in parentheses. GDP = gross domestic product.

*Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

licts between religious and ethnic groups. The Middle East is a region where Islam dominates. Perhaps our analysis should focus on Islamic countries rather than on the Middle East. However, the statistical analysis reveals no evidence for an effect of Islam on intrastate conflict (model 3). The Islam dummy variable is positive but far from significant.

What about Islam's "bloody innards"? Our modified variable—ethnic dominance, Islam (which takes account of the distinction between Shia and Sunni)—displays approximately the same values as the original one. This reinforces the conclusion that any dominant ethnic group increases the risk for conflict, but Islamic dominance no more so than other cases of dominance.

IS IT OIL?

On the background of the recent conflicts in Africa (Angola, Nigeria, Sierra Leone, and others), natural resource abundance has been suggested as an important factor in conflict. Easily lootable natural resources such as alluvial diamonds, gemstones, and drugs provide a financial base for rebel groups (Le Billon 2001). Oil is indisputably the most important natural resource in the Middle East. Does oil fuel or finance conflict? An oil dummy variable (1 = countries where more than 40 percent of exports derive from oil) does not increase the risk for conflict (model 4). Neither does an interaction term between oil and the Middle East.

Smith (2004) found, using a continuous measure of oil export dependence, that civil war and antistate protests were slightly less likely in oil-dependent states in the period from 1960 to 1999 and that regime durability was higher. Model 5 in Table 5 presents our findings with a continuous variable for oil and mineral resources. We do not find significant support for either a linear or a curvilinear relationship between internal conflict and oil dependence. (However, there was a high fraction of missing observations on this variable.) When we exclude all Middle East countries from the analysis (model 7), polyarchy is significant at the 10 percent level, and the significance of polyarchy squared is strengthened. These are the only important effects of excluding the Middle East countries.

IS IT AUTHORITARIANISM?

Collier and Hoeffler (1998, 2002, 2004) did not find any influence of regime type on conflict and excluded it from their model. However, several other empirical studies find a curvilinear relationship between democracy and conflict (Hegre et al. 2001; Fearon and Laitin 2003). Democracies are expected to experience less civil war because fewer sources of grievance are present, while authoritarian regimes have fewer conflicts because opposition is harshly repressed. Our analysis in Table 5 (model 6) clearly confirms the curvilinear relationship between conflict and democracy. The first term is positive, while the squared term is negative.¹⁰ The two terms combined are significant. Collier and Hoeffler's failure to find a curvilinear relationship between regime type and conflict is probably due to their use of five-year periods. Given the curvilinear nature of the relationship, the high level of authoritarianism cannot by itself account for the high level of conflict in the Middle East.

10. Running the analysis with the *Polity* variable yielded similar, although weaker, results.

TABLE 5
Accounting for Armed Conflict, 1960-2000:
Extending the Collier-Hoeffler Model

	Model 5	Model 6	Model 7
GDP per capita	-0.41*** (-4.40)	-0.34** (-2.53)	-0.39** (-2.43)
Economic growth	-11.05*** (-4.72)	-10.73*** (-4.31)	-13.00*** (-5.04)
Social fractionalization	-0.000030 (-0.39)	-0.000091 (0.12)	0.000080 (0.93)
Ethnic dominance	0.53* (1.85)	0.52* (1.79)	0.46 (1.41)
In Population	0.24* (1.95)	0.24* (1.95)	0.26* (1.94)
Geographic dispersion	0.41 (0.55)	0.62 (0.84)	1.60* (1.74)
Oil dependence	2.93 (1.12)	1.94 (0.72)	-3.02 (-0.68)
Oil dependence squared	-3.69 (-0.61)	-2.28 (-0.42)	12.34 (1.49)
Mineral dependence	2.75 (0.41)	2.11 (0.33)	-0.010 (-0.00)
Mineral dependence squared	-5.16 (-0.15)	-0.94 (-0.03)	6.42 (0.23)
Polyarchy		0.019 (1.04)	0.041* (1.95)
Polyarchy squared		-0.0026**	
Peace-years		-0.031 (-0.24)	(-1.91) (-0.025)
Spline 1		-0.00085 (-0.50)	(-0.17) (-0.00042)
Spline 2		0.00045 (0.39)	(-0.21) 0.000052
Spline 3		-0.00 (-0.01)	(0.04) 0.000015
Constant		-3.29** (-2.01)	(0.34) -4.35**
<i>n</i>	2,265	2,153	1,903
Number of conflict onsets	58	54	44
Percent onset (percentage)	2.6	2.5	2.3

NOTE: Robust *z*-values in parentheses. GDP = gross domestic product.

*Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

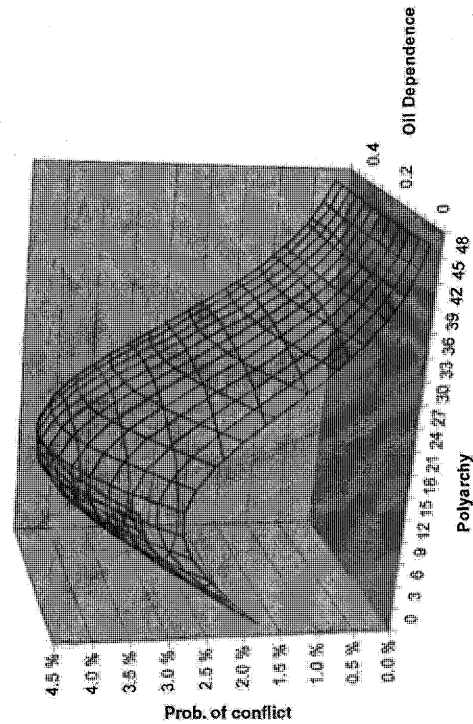


Figure 2: The Effect of Regime Type and Oil Dependence on the Probability of Armed Civil Conflict (Based on Model 6 in Table 5)

IS IT OIL AND AUTHORITARIANISM COMBINED?

Figure 2 displays a clear curvilinear relationship between regime type and conflict. The relationship between oil dependency and conflict remains weak, yielding an insignificantly higher risk for conflict for high levels of oil dependence and medium score on democracy. We found no significant results for interaction variables between economic development and, respectively, oil, Islam, and regime type; nor did we find one for oil dependence and semi-democracy.

Economic growth and economic development are the two most important variables in our analysis. Poor countries that are trapped in poverty seem to be the most war prone, with an average probability of conflict onset at 8.8 percent, more than 5 percentage points higher than the global average. We are unable to report any strong findings for raw material dependence. Figure 2 presents the effect of oil dependence from model 6 in Table 5. It is both small and statistically insignificant. The same figure illustrates the effect of the political regime variable. While this effect is statistically significant only at the 10 percent level, the effect portrayed in the figure clearly outperforms that of oil dependence. An average semi-democracy is close to 3 percentage points more likely to experience a conflict onset than is a full democracy.

TABLE 6
Comparing Observed and Predicted Probability of Civil Conflict

All Armed Civil Conflict	Standard		Minimum	Maximum	n
	Mean	Deviation			
Sample					
Conflict onsets (observed)	0.025	0.155	0	1	3,601
Predicted risk of onset (average of individual country predictions)	0.026	0.029	0	0.938	3,601
Sub-Saharan Africa					
Conflict onsets (observed)	0.032	0.177	0	1	1,180
Predicted risk of onset (average of individual country predictions)	0.036	0.027	0.0004	0.266	1,180
Middle East and North Africa					
Conflict onsets (observed)	0.035	0.183	0	1	347
Predicted risk of onset (average of individual country predictions)	0.026	0.019	0.0007	0.113	347

NOTE: This table is based on model 1 in Table 4.

A MISSING PIECE?

Are we missing one or several factors in the analysis of armed civil conflict in the Middle East? Do the variables in the CH model fail to register the sources of conflict in the Middle East? We examine the regional differences between sub-Saharan Africa and the Middle East and North Africa by comparing the data set's actual observed conflict onsets with the model's predicted probability of conflict for the sample and the two regions. The results are presented in Table 6. For Africa, the predicted incidence and the actual onset are almost the same. For the Middle East, the difference between observed conflict onset and predicted value is rather large. While observed conflict is 3.5 percent, the predicted incidence is almost 30 percent lower at 2.6 percent. Does the CH model favor variables that explain conflict in Africa and overlook variables that are of importance to conflict in the Middle East?

Including the modified ethnic dominance variable produced almost no change in the values. Thus, incorporating the divide between Shia and Sunni Muslims adds little explanatory value to the model.

The Israeli-Palestinian conflict has dominated domestic, regional, and world politics for more than five decades. Due to our research design, this conflict is not included in our dependent variable: the conflict broke out in 1948 and falls outside the period included in this study. Nevertheless, the Israeli-Palestinian conflict has spurred both interstate and intrastate conflict and hampered economic integration and social development in the region. Is it the conflict between Israel and Palestinians that accounts for conflict in the region? Dummy variables for all the countries neighboring Israel, as well as one for Israel and its neighbors, were constructed and added to the analysis. None of them was significant.

PEACE IN THE MIDDLE EAST?

This study has found no support for Middle Eastern exceptionalism regarding the causes of conflict. There is nothing mysterious or particular about conflict in the Middle East or in Muslim countries. Conflict is quite well explained by a general model of civil war, although the model underpredicts the probability of conflict in the Middle East. Collier and Hoeffler (2002) conclude that Africa would have seen less conflict if the region had achieved economic development at the world level. The Middle East has enjoyed much greater prosperity than Africa but would undoubtedly benefit from improved economic development. Nevertheless, deciphering the Middle East's correlates of conflict is a more complex task.

In the Middle East and North Africa, abundant access to oil has created a peculiar regional system as well as ties to the international markets and political actors. The failure to modernize and democratize has not led to major internal rebellions in the Middle East. So far, oil money has bought some of the countries enough carrots and sticks to keep their populations quiet. But what will happen when the oil wells run dry? Deteriorating economic conditions and the lack of democracy in the Middle East may well create a fertile base for grievance-based rebellions.

Loot-seeking behavior aimed at controlling oil revenue does not characterize civil war in the Middle East, and Collier and Hoeffler's looting argument appears to be less applicable than in Africa. The concept of the rentier state provides a more fruitful explanation for the current and upcoming economic and political challenges faced by oil producers in the Middle East. These economic and political challenges fit better into the traditional grievance-based perspective on civil war. Lack of economic and political opportunities provides a fruitful base for frustration and opposition. Unless the regimes become increasingly repressive, the future is likely to hold growing political unrest and even civil war in the Gulf area. Sick and Potter (1997, 12) suggest that the Gulf countries are experiencing "a crisis in slow motion." The economic and political effects of oil—corruption, slow growth, and authoritarianism—create a basis for grievance-based conflict. The growth of political Islam is, to a large extent, a result of these forces.

The regime change strategy in Iraq is not likely to dramatically change the Middle East's status quo in which the region's authoritarian regimes have remained in power to a large extent due to external (particularly U.S.) support. These regimes have little legitimacy and put the West in an awkward light in many Arab eyes. Although the authoritarian regimes so far have been successful in quelling serious political and domestic armed conflict, they have not successfully constructed solid economic and social platforms for the future. The incumbent regimes face serious challenges to provide food and jobs to their populations (UNDP 2002). Ensuring diversified economic development remains the key to preventing future conflicts. Improved management of resources, natural as well as human, and the development of more transparent and legitimate, if not democratic, regimes provide the most efficient means toward preventing new conflicts—in the Middle East as in other developing regions.

APPENDIX 1
Accounting for Civil War, 1965-1999:
Results from Collier and Hoeffler (2002)

	Standard Error	β	Standard Error	β	Standard Error	β
In GDP per capita		-0.244***		-0.950		-1.053
GDP growth $t - 1$		0.042**		-0.098		0.042**
Primary commodity exports		16.773		5.206***		16.691
Primary commodity exports squared		-23.800		10.040**		-23.532
Social fractionalization		-0.0002		0.0001***		-0.0002
Ethnic dominance		0.480		0.328		0.449
Peace duration		-0.004		0.001***		-0.004
In Population		0.510		0.128***		0.473
Geographic dispersion		-0.992		0.909		-0.994
Sub-Saharan Africa (SSA) dummy						-0.370
French SSA dummy						0.526
Pseudo- R^2	750	0.22	750	0.22	750	0.23
Log-likelihood	-146.84		-146.50		-146.10	
Number of wars included	46		46		46	

SOURCE: Reproduced from Collier and Hoeffler (2002, 16, Table 1). Used with permission.

NOTE: Collier and Hoeffler did not report the value of the constant in their article.

*Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

REFERENCES

Aarts, Paul. 1999. The Middle East: A region without regionalism or the end of exceptionalism? *Third World Quarterly* 20 (5): 911-25.

Addison, Tony, Philippe Le Billon, and S. Mansoob Murshed. 2002. Conflict in Africa: The cost of peaceful behaviour. *Journal of African Economies* 11 (3): 365-86.

Auy, Richard M., ed. 2001. *Resource abundance and economic development*. UN/WIDER Studies in Development Economics. New York: Oxford University Press.

Barrett, David B., ed. 1982. *World Christian encyclopedia*. Oxford, UK: Oxford University Press.

Barro, Robert J. 1997. *Determinants of economic growth*. Cambridge, MA: MIT Press.

Behawi, Hazem, and Giacomo Luiciani, eds. 1987. *The rentier state*. New York: Croom Helm.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42 (4): 1260-88.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge, UK: Cambridge University Press.

Bureau of Verification and Compliance. 2000. World military expenditures and arms transfers 1998. Fact sheet, August 21, U.S. Department of State, Washington, DC. Accessed from <http://www.fas.org/man/docs/wmeat98/wmeat98fs.html>.

Collier, Paul. 1999. Doing well out of war. Paper presented at the Conference on Economic Agendas in Civil Wars, April 26-27, London. Accessed from http://econ.worldbank.org/files/13200_CollierDoingWell.pdf.

Collier, Paul, Lant E. Haggard, Anke Hoefler, Martha Reynal-Querol, and Nicholas Sambanis. 2003. Breaking the conflict trap: Civil war and development policy. Washington DC: World Bank.

Collier, Paul, and Anke Hoefler. 1998. On economic causes of civil war. *Oxford Economic Papers* 50 (4): 563-73.

_____. 1999. *Justice-seeking and loot-seeking in civil war*. Washington, DC: World Bank. Accessed from http://econ.worldbank.org/files/13203_CHJusticeSeeking.pdf.

_____. 2002. On the incidence of civil war in Africa. *Journal of Conflict Resolution* 46 (1): 13-28.

_____. 2004. Greed and grievance in civil war. *Oxford Economic Papers* 56 (4): 563-95.

Correlates of War. 2002. *Civil war dataset*. Accessed from <http://www.umich.edu/~cowproj/dataset.html#CivilWar/>.

Dahl, Robert A. 1971. *Polyarchy: Participation and opposition*. New Haven, CT: Yale University Press.

Department of Geodesy and Cartography of the State Geological Committee of the USSR. 1964. *Atlas Narodov Mira* [Atlas of the people of the world]. Moscow: Department of Geodesy and Cartography of the State Geological Committee of the USSR.

de Soysa, Indra. 2002. Paradise is a bazaar? Greed, creed and governance in civil war, 1989-99. *Journal of Peace Research* 39 (4): 395-416.

Donno, Daniela, and Bruce Russett. 2004. Islam, authoritarianism, and female empowerment: What are the linkages? Typescript, Department of Political Science, Yale University.

Ellingsen, Tanja. 2000. Colorful community or ethnic witches' brew? Multiculturalism and domestic conflict during and after the cold war. *Journal of Conflict Resolution* 44 (2): 228-49.

Eriksson, Mikael, and Peter Wallensteen. 2004. Armed conflict, 1989-2003. *Journal of Peace Research* 41 (5): 625-36.

Fearon, James D., and David D. Laitin. 2003. Ethnicity, insurgency, and civil war. *American Political Science Review* 97 (1): 75-90.

Fish, M. Steven. 2004. Islam and authoritarianism. *World Politics* 55 (1): 4-37.

Friedrich, Robert J. 1982. In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science* 26 (4): 797-833.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. Armed conflict 1946-2001: A new dataset. *Journal of Peace Research* 39 (5): 615-37.

Gurr, Ted Robert. 1970. *Why men rebel*. Princeton, NJ: Princeton University Press.

_____. 1993. *Minorities at risk: A global view of ethno-political conflict*. Washington, DC: U.S. Institute of Peace Press.

NOTE: This list is derived from the PRIO/ppsata database on armed conflict (Gleditsch et al. 2002, www.prio.no/cwp/armedconflict). These conflicts add altogether up to 190 conflict-years and 16 years with conflict onsets in the Middle East in the period from 1960 to 2003. For a list of the countries in the Middle East, see the note to Figure 1. We have included here only the conflicts that took place in the Middle East. The only countries in the Middle East to intervene in conflicts outside the region were Libya (in Chad, 1965-1990; in Uganda, 1978-1979; and in the Central African Republic, 2001) and Turkey (in Cyprus, 1974, and in Yugoslavia, 1999). These conflicts are excluded here, along with the conflict between the United States and al Qaeda, where Jordan is listed as an intervening power. Several Middle East countries (such as Israel and Syria) have intervened in other countries in the region. These conflicts are included, but only the country where the conflict was taking place is coded in the data on conflict onset or incidence. The years marked as bold are those that are included as onset observations in our analysis. The remaining years are censored. Since the Uppsala definition of conflict focuses on the incompatibility and the parties involved in the conflict, it can be difficult to distinguish between the onset of a new conflict and a continuation of an existing one. The conflict ID identifies a new separate conflict, but we also include a new sub-ID as onset. A new sub-ID is coded if the conflict has failed to reach the twenty-five battle-related casualties threshold for ten consecutive years or if there is a complete change on the opposition side. We do not include shifts from internal armed conflicts to internationalized internal armed conflicts as new sub-IDs. Only conflicts initiated in 2000 or earlier are included in the statistical analysis.

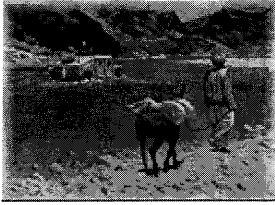
a. For details of the Islamic opposition groups in Algeria and for the full name of other groups indicated by their initials only, see Gleditsch et al. (2002, Appendix 1) and Eriksson and Wallensteen (2004, Appendix 1).

b. The conflict between Israel and the Palestinians is censored because the onset predates the period studied here.

ID	Sub-ID	Location	Side A	Side B	Begin
2930	0	Algeria	al-Qamaa al-Islamiyya, al-Jihad al-Islamiy, Tala 1 al-Fatih	KDPI	1991-2003
2980	0	Egypt		APCO	1992-1998
1060	1	Iran		KDPI	1966-1968, 1979-1988, 1990, 1993
2450	0	Iran		APCO	1979-1980
2440	0	Iran		Mujahideen e Khalq	1979-1982, 1986-1988, 1991-1993, 2000-2001
1620	1	Iraq		SARI	1963
1620	0	Iraq		Military faction	1982-1984, 1987, 1991-1996
1740	0	Iraq		KDP, PUK	1961-1970, 1973-1993
1370	0	Israel		Palestinian insurgents	1949-2003
1630	1	Lebanon	Various organizations, Syria, Israel	POLISARIO	1975-1990
2350	0	Morocco			1975-1989
2150	0	Morocco		Military faction	1971
2210	0	Oman	Oman, Iran, Jordan, United Kingdom	PFLAQ, South Yemen	1972-1975
2460	0	Saudi Arabia		Juhayman movement	1979
2020	0	Syria		Military faction	1966
2020	1	Syria		Muslim Brotherhood	1979-1982
2490	0	Tunisia		Resistance Armee Tunisienne	1980
2600	0	Turkey		FKK/KADEK	1984-2001
2900	0	Turkey		Devrim Sol	1991-1992
3100	0	Yemen		Democratic Republic of Yemen	1994
1330	1	Yemen (North)		Royalists	1962-1970
1330	1	Yemen (North)		National Democratic Front	1980-1982
2650	0	Yemen (South)		Faction of Yemenite Socialist Party	1986

APPENDIX 2
Intrastate and Internationalized Armed Civil Conflicts in the Middle East, 1960-2003

- Hegre, Håvard, ed. 2004. Duration and termination of civil war [Special issue]. *Journal of Peace Research* 41 (5).
- Hegre, Håvard, Tanja Ellingsen, Scott Gates, and Nils Petter Gleditsch. 2001. Toward a democratic civil peace? Democracy, political change, and civil war, 1816-1992. *American Political Science Review* 95 (1): 33-48.
- Homer-Dixon, Thomas F. 1999. *Environment, scarcity and violence*. Princeton, NJ: Princeton University Press.
- Huntington, Samuel P. 1991. *The third wave: Democratization in the late twentieth century*. London: Oklahoma University Press.
- . 1996. *The clash of civilizations and the remaking of world order*. New York: Simon & Schuster.
- Inglehart, Ronald. 1997. *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
- Jagers, Keith, and Ted Robert Gurr. 1995. Tracking democracy's third wave with the Polity III data. *Journal of Peace Research* 32 (4): 469-82.
- Kedourie, Elie. 1994. *Democracy and Arab political culture*. London: Frank Cass.
- King, Gary, and Langche Zeng. 2001. Logistic regression in rare events data. *Political Analysis* 9 (2): 137-63.
- Kramer, Martin. 1996. Islam and democracy. In *Arab awakening & Islamic revival: The politics of ideas in the Middle East*, edited by Martin Kramer, 265-78. New Brunswick, NJ: Transaction Publishers.
- Lacina, Bethany, and Nils Petter Gleditsch. 2004. *Monitoring trends in global combat: A new dataset of battle deaths*. Oslo, Norway: Centre for the Study of Civil War, PRIO. Accessed from <http://www.prio.no/csw/cross/battledeaths/>
- Le Billon, Philippe. 2001. The political ecology of war: Natural resources and armed conflicts. *Political Geography* 20 (5): 561-84.
- Lonegan, Steve. 1997. Water resources and conflict: Examples from the Middle East. In *Conflict and the environment*, edited by Nils Petter Gleditsch, 375-84. Dordrecht, the Netherlands: Kluwer Academic.
- Owen, Roger, and Sevet Pamuk. 1995. *A history of Middle East economies in the twentieth century*. Cambridge, MA: Harvard University Press.
- Ross, Michael L. 2001. Does oil hinder democracy? *World Politics* 51 (1): 325-61.
- Rubin, Barry. 2002. *The tragedy of the Middle East*. Cambridge, UK: Cambridge University Press.
- Sachs, Jeffrey D., and Andrew M. Warner. 1995. Natural resource abundance and economic growth. NBER Working Paper 5318, National Bureau of Economic Research, Cambridge, MA.
- Sambanis, Nicholas. 2002. A review of recent advances and future directions in the quantitative literature on civil war. *Defence and Peace Economics* 13 (3): 215-43.
- Sick, Gary G., and Lawrence G. Potter, eds. 1997. *The Persian Gulf at the millennium: Essays in politics, economy, security and religion*. New York: St. Martin's.
- Sköns, Elisabeth, Evamaría Loose-Weintraub, Wuyi Onitooogun, and Petter Stålenheim. 2002. Military expenditure. In *SIPRI yearbook 2002: Armaments, disarmament and international security*, 231-322. Oxford, UK: Oxford University Press.
- Smith, Benjamin. 2004. Oil wealth and regime survival in the developing world, 1960-1999. *American Journal of Political Science* 48 (2): 232-46.
- Sørli, Mirjam E. 2002. Resources, regimes and rebellion: A critical analysis of the greed and grievance model of civil conflict. Cand. polit. thesis, Department of Political Science, University of Oslo and Centre for the Study of Civil War at the International Peace Research Institute, Oslo.
- StataCorp. 2003. Stata statistical software: Release 8.2 (1). College Station, TX: Stata Corporation.
- Strand, Håvard, Lars Wilhelmsen, and Nils Petter Gleditsch. 2003. *Armed conflict dataset codebook, version 1.2*. Oslo, Norway: International Peace Research Institute (PRIO). Accessed from http://www.prio.no/cwp/ArmedConflict/old/v1_20/Codebook_v1_2a.pdf.
- Tucker, Richard. 1998. BTSCS software. Accessed from <http://www.vanderbilt.edu/rucker/programs/btscs/>
- United Nations Development Program (UNDP). 2002. *Arab human development report*. New York: United Nations Development Program.
- Urdal, Henrik. 2002. The devil in the demographics: How neo-Malthusian population pressure and youth bulges influence the risk of domestic armed conflict. Cand. polit. thesis, Department of Political Science, University of Oslo, Norway.
- Vanhaanen, Tatu. 2000. A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research* 37 (2): 251-65.
- Woltering, Robbert A. F. L. 2002. The roots of Islamist popularity. *Third World Quarterly* 23 (6): 1133-43.
- World Bank. 2002. *World development indicators*. Washington, DC: Development Data Center, International Bank for Reconstruction and Development.



SMALL WARS

JOURNAL

Primitivization of War and Prospects for Peace

By *Mehar Omar Khan*

Journal Article | Sep 14 2012 - 3:30am

...Because the Barbarians arrive today...

What laws now should the Senators be making?

When the Barbarians come they'll make the laws.[1]

The barbarians have come and the rules of war and peace stand transformed. Retired U.S. Army Colonel Ralph Peters once stated a historical fact rather tersely: "Technologies come and go, but the primitive endures." [2] Well, the primitive has endured and reared its head again. The rise and rise of intrastate, irregular, and asymmetric warfare has *primitivized* the war and this is, to use Thomas Kuhn's term, no "anomaly." [3] It is instead a paradigm shift that has the potential to have seminal, largely unintended, and hugely positive consequences for world peace. Wise people like General Rupert Smith have actually called it exactly that: a paradigm shift in conduct of war. [4] Unless stymied by the powerful military industrial complexes and parasitic security establishments, the *simplification* of war has the potential to change into a sprouting of peace.

It may be worthwhile to sit up and take note of the causes of this potentially watershed moment in the history of war and peace before highlighting just some of its consequences.

Globalized economy, a seemingly irreversible march of democracy, and the shadowy nature of our shared enemy have together brought about a seminal change in warfare. While states have increasingly shied away from marching on other states, technology has taken the back seat as the footsloggers duel with the insurgents and the terrorists for not just chunks of real estate but also the hearts and minds of the populace in objective zones. Unless the mighty hands of gods roll back the chessboard of global economy, impose a global tyranny, and make all bad people plotting to kill innocent men and women suddenly disappear, this change is here to stay. From Turkey to Afghanistan and from Pakistan's northwest to Thailand's southeast, nation states have their multi-million dollar bombers and hi-tech tanks parked idle as their foot soldiers battle the enemy that resides within the borders and fights both the state as well as the society.

The debate between those who believe in a "bright" future for state-on-state warfare (for example, Colin Gray)[5] and those (for examples, Martin Creveld,[6] Robert Kaplan,[7]) who argue that factors like nuclear weapons and economic interdependence on global scale have all but ended possibilities for any large-scale state-on-state warfare, is at best ongoing. However, most of the recent wars have been intra-state or ones that involved coalitions of states fighting non-state actors. Since 1991, world has seen few state-on-state wars. According to SIPRI Year Book 2012, during the first decade of 21st century the world has witnessed 69 armed clashes and 221 non-state conflicts. [8] Even the most spectacular showcase of conventional military hardware of the last century, Gulf War-I, also does not qualify as an inter-state conflict because it was actually fought by a large international coalition against a regime that had, for a long time, waged war on its own people and had lost all credibility. The same can be said about NATO's

campaign against Milosevic's Serbia, or against Qaddafi's Lybia.

Empirical evidence also suggests that the states that employ high-tech means and militaries to fight the non-state actors seldom win decisively. Over the last couple of centuries, insurgents' "rage against machines" has invariably prevailed.[9] Thus, trillions of dollars sunk into building state-of-the-art, often novel, weapon systems have increasingly become painfully irrelevant to modern, "fourth generation"[10] of warfare.

While the world will continue to be home to self-serving nation states with unequal military and economic powers, conflict in the military realm will, for the foreseeable future, ultimately and invariably be *decided* between the foot soldiers. Expensive space, sea, and air borne assets are condemned to the margins for next several years, possibly decades. As the recent counterinsurgency efforts in Iraq and Afghanistan have proved, the technologically superior state militaries will be obliged to climb the mountains, snake the streets, and jump the walls in order to fight the guerillas *like* the guerillas. While all the impressive air and ground based, hi-tech war-fighting means may help shape the battlefield, the footsloggers will ultimately decide the outcome of the conflict. The political and population-centric nature of the current and future wars will also oblige nation states to maintain militaries that value men more than machines to hold the ground and carefully reap the dividends of politico-economic investments.

This paradigm shift in *warfare* is bound to have consequences for the larger phenomenon of *war* as a tool of policy. Well into the foreseeable future, war will continue to become longer but cheaper and manageable for states as it becomes harder for the military planners. Longer, mainly infantry-fought, *primitivized* wars that can't be quickly and decisively won through sheer force of armament and which invariably need massive political and economic compliments will strengthen the role and dominance of policymakers and dilute the control of Generals. That means the world will see less and less of the *run-away* Generals marked for their disdain for the political oversight of the war effort. Similarly, the *primitivization* of war is also likely to enable political leaders to radically bring down investments in procurement and maintenance expenditures and focus more on people- and employment-friendly futuristic research and development.

Given the fact that war is a permanent reality of human existence, pacifists across the globe have enough reasons to celebrate the down-grading of war from an all-consuming national effort with millions of deaths to a sub-state undertaking that has increasingly cost less in human lives and that progressively demands a nice mix of military, economic, informational, and political means instead of blind force of deadly arms. Increased complexity and longevity of war is bound also to make it ultimately less and less attractive as a tool of statecraft. However, even in its present form, it will be increasingly fought by ground forces, with infantry-men in the van and through intelligence and surgical, targeted operations that will be limited in temporal and spatial dimensions. US-NATO's war effort in Afghanistan and Pakistan's counter-terror campaign in its tribal belt are the current exemplars of this type of war.

This transformative trend should lead to several encouraging possibilities for peace:

- Electorates across the world are entitled now to demand a cut in the defense budgets and greater investments in human security - food, shelter, fighting pandemics etcetera. Nation states may still be able to maintain their competitive edge in military technologies through progressive and futuristic R&D and without engaging in irresponsible shopping sprees for hi-tech aerial and naval platforms. As the now-former rulers of Egypt and Lybia have found out - and some more across the Arabian Peninsula may soon discover - their modern day rivals will not be impressed or defeated by multi-billion dollar fighter jets and tanks. Even the relatively mightier nations have found their machines irrelevant to wars in Turkish Kurd regions, Afghanistan, and Palestinian territories. It may therefore be time for the taxpayers to ask their governments: why invest in such systems?

- The long, hard, politically oriented 4GW may also resurrect Clausewitz and so arrange his trinity[11] that governments, guided by what the Master called “reason”, are actually placed on the top corner of the trinity. Policy may thus be able to regain *total* control of war. In a democratic world, ascendancy of policy and its control over war can have only positive consequences for peace. This, for one thing, will ensure that the strong arm of the nation does not run amuck and nullify the gains of the policy.
- It may also be time to de-glorify war as there is no real glory left in the complex, nerve-testing, hard-slog that modern war has become. Fortunately for world peace, humanity is not likely to see any glorious cavalry charges, awe-inspiring tank formations, or earth shattering strategic bombings any time soon. It may be time to worship heroes that do not deal in blood and steal, but instead are simple, boring, humanity-loving people.
- The *primitivization* of war is also an opportunity to firewall warfare from the limitless imaginative power of the human mind and divert our collective energies to perpetuation of peace. If we fail now, the dogs of war will tear the world with ferocity that humanity has never seen before. This is not hyperbole. Since times immemorial, the demons of war have used periods of relative peace to recover their destructive power and strike back with unprecedented force.
- This may also be time for security establishments to invest in the weapons that hurt the insurgents most: the foot soldiers. A progress in that direction will end up humanizing the warfare by, *inter alia*, minimizing the terrible cost in terms of collateral damage.

Summing up, seldom before has a step backward actually thrown up a range of opportunities to move forward. *Primitivization* of war may just prove a giant leap forward for mankind.

[1] From a poem entitled *Waiting for the Barbarians* by Greek poet C.P. Cavafy. (Translated into English by John Mavrogordato and reproduced in Readings in World Literature (HRW Library))

[2] Ralph Peters, “Our New Old Enemies”, *Parameters*, Summer 1999.

[3] Kuhn, Thomas S., *The Structure of Scientific Revolutions*, (University of Chicago Press, 1970), pp. 52-60.

[4] Smith, Gen Sir Rupert, *The Utility of Force: The Art of War in the Modern World*, (Allan Lane, 2005), pp. 1-20.

[5] Gray, Colin S., *Another Bloody Century: Future Warfare*, (London: Phoenix, 2006)

[6] Creveld, Martin Van, *The Rise and Decline of the State*, (Cambridge University Press, 1999)

[7] Kaplan, Robert D., *The Coming Anarchy: Shattering the Dreams of the Post Cold War* (Random House, 2000).


[8] Highlights of SIPRI data on the changing trends in conflict are available at <http://www.sipri.org/yearbook/2012>

[9] Lyall, Jason and Isaiah Wilson, III, "Rage Against the Machines: Explaining Outcomes in Counterinsurgency Wars," June 17, 2008. Available online at http://www.humansecuritygateway.com/documents/LYALL_RageAgainstTheMachines.pdf

[10] U.S. Marine Col T.X. Hammes has written a great book on the idea (*The Sling and The Stone: On War in the 21st Century*, Zenith Press: 2004) propounded originally by William S. Lind in his 1989 article for Marine Corps Gazette, entitled "The Changing Face of War: Into the Fourth Generation".

[11] The Prussian Master identified State, Society, and the Military as part of a trinity that conducts policy and its extension called war.

About the Author



Mehar Omar Khan

Mehar Omar Khan has graduate degrees in International Security Studies (Kansas State University) and International Relations (Balochistan University Quetta). He blogs on issues related to security with a focus on insurgency. Views expressed are his own. He can be reached at noorbakhte@yahoo.co.uk.

Available online at : <http://smallwarsjournal.com/jrnl/art/primitivization-of-war-and-prospects-for-peace>

Links:

{1} <http://smallwarsjournal.com/author/mehar-omar-khan>

{2}

<http://books.google.com.pk/books?id=evtlSoQIAdcC&dq=Rupert+Smith,+2005&source=bl&ots=ewqKbFIz/>

{3} <http://www.sipri.org/yearbook/2012>

{4} http://www.humansecuritygateway.com/documents/LYALL_RageAgainstTheMachines.pdf

{5} <http://smallwarsjournal.com/jrnl/art/war-and-words-the-menace-of-meaning>

Copyright © 2014, Small Wars Foundation.



Select uses allowed by Creative Commons BY-NC-SA 3.0 license per our [Terms of Use](#).

Please help us support the [Small Wars Community](#).



 NOTICE:
 This material may be protected by copyright law (Title 17 U.S.
 Code)

Tehran's Take

Understanding Iran's U.S. Policy

Mohsen M. Milani

Tehran's Take: Understanding Iran's U.S. Policy
 Author(s): Mohsen M. Milani
 Source: *Foreign Affairs*, Vol. 88, No. 4 (July/August 2009), pp. 46-62
 Published by: Council on Foreign Relations
 Stable URL: <http://www.jstor.org/stable/20699621>
 Accessed: 11/12/2013 09:06

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

ALTHOUGH A great deal has been written about the United States' policy toward Iran, hardly anything comprehensive has been produced about Iran's policy toward the United States. Given Washington's concerns that the United States faces "no greater challenge from a single country than from Iran," as the 2006 National Security Strategy put it, this lack of serious attention is astonishing. What does exist is sensationalistic coverage about Iran's nuclear ambitions and about mad mullahs driven by apocalyptic delusions and a martyr complex. That picture suggests that Iran's policy consists of a series of random hit-and-run assaults on U.S. interests and that its leaders, being irrational and undeterrable, must be eliminated by force.

In fact, Tehran's foreign policy has its own strategic logic. Formulated not by mad mullahs but by calculating ayatollahs, it is based on Iran's ambitions and Tehran's perception of what threatens them. Tehran's top priority is the survival of the Islamic Republic as it exists now. Tehran views the United States as an existential threat and to counter it has devised a strategy that rests on both deterrence and competition in the Middle East.

To deter any possible military actions by the United States and its allies, Iran is improving its retaliatory capabilities by developing the means to pursue asymmetric, low-intensity warfare, both inside and

MOHSEN M. MILANI is Professor of Politics and Chair of the Department of Government and International Affairs at the University of South Florida in Tampa.

[46]



Council on Foreign Relations is collaborating with JSTOR to digitize, preserve and extend access to *Foreign Affairs*.

<http://www.jstor.org>

outside the country; modernizing its weapons; building indigenous missile and antimissile systems; and developing a nuclear program while cultivating doubts about its exact capability. And to neutralize the United States' attempts to contain it, the Iranian government is both undermining U.S. interests and increasing its own power in the vast region that stretches from the Levant and the Persian Gulf to the Caucasus and Central Asia. Although it is being careful to avoid a military confrontation with the United States, Tehran is maneuvering to prevent Washington from leading a united front against it and strategically using Iran's oil and gas resources to reward its friends.

Iranian foreign policy today is as U.S.-centric as it was before the 1979 revolution. Mohammad Reza Shah Pahlavi relied on Washington to secure and expand his power; today, the Islamic Republic exploits anti-Americanism to do the same. Policy has been consistent over the years partly because it is determined by the supreme leader, who is also the commander of the security and armed forces and serves for life. Iran's defiance has in some ways undermined the country's national interests, but it has paid huge dividends to the ruling ayatollahs and helped them survive three tumultuous decades in power.

Today, Ayatollah Sayyid Ali Khamenei is the supreme leader, and he makes all the key policy decisions, usually after Iran's major centers of power, including the presidency, have reached a consensus. This means that the outcome of the presidential election in June will have some, although probably limited, ramifications for Iran's foreign policy. President Mahmoud Ahmadinejad and his two major reformist rivals, Mir Hossein Mousavi and Mehdi Karubi, have all supported engaging in negotiations with Washington—a political taboo just a few years ago. Ahmadinejad would be less likely to compromise than his more moderate competitors, but, thanks to the support he has among major anti-American constituencies inside and outside the Iranian government, he would be in a better position to institutionalize any shift in policy. Although Iran's president can change tactical aspects of the country's foreign policy, he cannot single-handedly alter its essence. Only Khamenei, the ultimate decider, can do that. And he will do that only if a fundamental change in policy would not undermine his own authority and if it enjoys broad support from among the major centers of power.

THE HUNGRY WOLF AND THE FAT SHEEP

THE ROOTS of anti-Americanism in Iran—really, an opposition to U.S. policies—can be traced to the 1953 coup against Mohammad Mosaddeq, which was backed by the CIA and MI6. Anti-American sentiment was strengthened when in 1964 Ayatollah Ruhollah Khomeini, who would become Iran's first supreme leader after the revolution, opposed a treaty granting legal immunity to U.S. military advisers in Iran and declared that Iran had become a “U.S. colony.” By 1979, the revolutionaries were portraying the Iranian monarch as “America's shah” and had made “independence” a defining slogan of their movement. After the taking of hostages at the U.S. embassy in Tehran that year, anti-Americanism became an enduring feature of the state's Islamic ideology. Since then, Iran's leaders have deftly linked the survival of the Islamic Revolution to Iran's independence, depicting the United States as antithetical to both. No one has drawn this link more vividly than Khomeini: he called the United States “the great Satan” and compared U.S. relations with Iran to those between a hungry wolf and a fat sheep. As hostility between the two states intensified, a Manichaean security paradigm developed in both of them. Each one came to perceive the other as a mortal enemy in a zero-sum game. Anti-Americanism and anti-Iranian feelings became two sides of the same coin.

For decades, the Iranian regime has used anti-Americanism to crush its opponents at home and expand its power abroad. After 1979, documents selectively released by the radical students who occupied the U.S. embassy were invoked to establish links between opponents of the Islamic Republic and the U.S. government. Hundreds of people were consequently defamed, jailed, or exiled. In the 1980s, when the young regime was simultaneously struggling to consolidate its rule and fighting a war with Iraq, allegations that the U.S. government was attempting to stage coups in Tehran and prevent Iran from winning the war strengthened the sentiment.

Over time, anti-American constituencies in Iran have proliferated and gained ground in various institutions. Some of them oppose the United States for purely ideological reasons. Others have substantial economic interests in preventing the normalization of relations between

Tehran and Washington: they profit from domestic black markets and international trade routes established to bypass U.S. sanctions. Moreover, the foreign organizations that Iran supports throughout the Middle East, and that Washington considers to be terrorist groups, have created effective lobbies in Iran that thrive on this animosity.

Yet now, as under Khomeini, the intensity of the anti-Americanism prevailing in Iran is ultimately determined by the supreme leader. As president, Khomeini declared in 1981, "We are not like [Salvador] Allende [a Chilean president ousted by a coup allegedly backed by Washington], liberals willing to be snuffed out by the CIA." Today, Khomeini still considers the United States to be an existential threat. Washington surrounds Iran with bases in Bahrain, Kuwait, and Qatar and massive troop buildups in Afghanistan and Iraq. It makes friends with the leaders of Iran's neighbors. And its nuclear-equipped naval carriers patrol the Persian Gulf. Khomeini sees the United States as isolating Iran, strangling it with economic sanctions, sabotaging its nuclear program, and beating the drums of preemptive war. He thinks Washington is pursuing regime change in Tehran by funding his opponents, inciting strife among Iran's ethnic minorities, and supporting separatist organizations such as the Baluchistan-based Sunni insurgent group Jundallah, which has killed scores of Revolutionary Guard members.

PREFER TO DETER

TEHRAN HAS responded to Washington's policy of containment with a strategy of deterrence. Tehran first developed this strategy against Iraq after Iraq invaded Iran in September 1980 and then, as the 1980s unwound and the menace of Iraq faded, redirected it toward the United States. Today, this approach is the result both of Iran's perception of its vulnerabilities and of the constraints that the international community has imposed on the country.

Iran's deterrence strategy has four components. The first is developing the means to fight an asymmetric, low-intensity war, inside and outside the country. In recent years, particularly after U.S. troops arrived in Afghanistan and Iraq, the Revolutionary Guards have played an increasingly important role in maintaining internal order. They

have also improved Iran's retaliatory capability in case of an invasion or surgical strikes against its nuclear facilities or the headquarters of its security forces. Khomeini's recent decision to decentralize the command-and-control structure of the Revolutionary Guards serves this purpose. So do the alleged ability of Iran's troops to transform themselves into nonconventional forces within days and Iran's thousands of small Iranian-made assault boats, which could create havoc for the U.S. Navy, as well as its thousands of motorcycles equipped with light artillery, which could impede the advances of an invading army. Iran's support for terrorist actions against U.S. interests in the Middle East is part and parcel of its strategy of asymmetric warfare. For example, in 1983, during the civil war in Lebanon, a group associated with the Iranian-backed Hezbollah killed 220 U.S. marines.

The modernization of Iran's weapons systems is the second component of its deterrence strategy. Decades of arms embargoes from the West have left Iran with limited access to advanced weapons, and Iran has consequently purchased relatively small arms supplies. Between 2002 and 2006, it spent \$31 billion on military purchases, compared with \$109 billion for Saudi Arabia and a total of \$48 billion for Bahrain, Kuwait, Oman, and the United Arab Emirates—four states with a combined population smaller than that of the city of Tehran. The embargoes have also caused an indigenous military-industrial complex to develop, controlled and financed by the state. It employs thousands of people and is connected to the country's major universities and think tanks. Most important, it is in charge of research and development for Iran's missile and nuclear technologies.

Developing indigenous missile and antimissile systems is the third leg of Iran's deterrence strategy. Tehran began building missiles during the Iran-Iraq War and accelerated its program after the "war of cities" in 1988, when both states showered the other's cities with missiles. Iran has used technical support from China and Russia to develop its missile technology. Now, it manufactures its own missiles and claims that two types, the Shahab and the Ghadr, can reach Israel. These missiles are known for their inaccuracy and limited offensive application. But they give Iran the power to retaliate against attacks, particularly in the Persian Gulf, where it could readily upset international navigation.

The fourth component of Iran's deterrence strategy is its nuclear program. The Iranian government claims that its program is designed for peaceful purposes—using nuclear weapons would violate Islamic law, it says—but Washington (and much of the West) accuses it of having a secret program to build the bomb. So far, the International Atomic Energy Agency has found no smoking gun or any evidence that Iran has diverted its nuclear program toward military purposes. But nor has it been able to confirm Tehran's peaceful intentions, because the Iranian government has refused to answer some major questions. Now that Iran has joined a small club of countries that can enrich uranium to a low level of purity, it seems unlikely to cave in to international pressure and accept zero enrichment in the future.

Why, given all the sanctions imposed by the United Nations, does Iran not cry uncle and stop its nuclear program? For one thing, by insisting that its nuclear project is essential for the country's domestic energy needs and scientific development, Tehran has effectively turned U.S. opposition to its program into a nationalist cause, pointing to it as proof that Washington intends to hold Iran back. (In an attempt to awaken national pride, the government has had the atom symbol printed on 50,000 rial bills.) For another, the nuclear impasse creates an excellent bargaining chip for Tehran in future negotiations. This may be the reason that Iran's leaders are cultivating uncertainty about the country's actual capability. It does appear, however, that they have decided to develop the infrastructure to build the bomb but not yet the bomb itself. (Former President Ali Akbar Hashemi Rafsanjani claimed in 2005, "We possess nuclear technology that is not operationalized yet. Any time we decide to weaponize it, we can do so rather quickly.") Iran and the United States seem to be engaged in a game of poker, with Tehran not showing its cards about its nuclear capabilities and Washington refusing to exclude the possibility of attacking Iran. Washington has the better hand, but the better hand does not always win.

COUNTERCONTAINMENT

FOR THREE decades, the United States has sought to contain Iran and has imposed on it a variety of sanctions in an effort to do so. To try to neutralize these measures' effects, Iranian leaders have played

major powers off against one another, forged alliances of convenience, and asserted Iran's interests at the regional and global levels.

First, Iran has tried to create a wedge between the United States and the United States' European allies. Iran's leaders believe that increased trade with the European Union will allow them to exploit differences among the organization's 27 members and discourage it from supporting regime change in Tehran, the total containment of Iran, or a military attack. In other words, they see the EU as a potential counterweight to the United States. After Iran restarted its uranium-enrichment activities in 2003, the EU ended its "constructive engagement" policy and imposed limited sanctions on Iran to tame its nuclear ambitions. Although these restrictions have prevented both Iran from gaining access to advanced technologies and European firms from making substantial investments in Iran, they have had a negligible impact on the overall volume of trade. The EU remains Iran's leading trading partner, accounting for about 24 percent of Iran's total international trade: the EU's total imports from Iran (mostly energy) increased from 6.3 billion euros in 2003 to 12.6 billion euros in 2007, and its exports to Iran (mostly machinery) remained the same, at about 11.2 billion euros, during that period. On the other hand, the fact that France, Germany, and the United Kingdom supported the referral of Iran's case from the International Atomic Energy Agency to the UN Security Council in 2005 proved the limits of Iran's wedge policy. The move, which highlighted the determination of the major Western powers to tame Iran's nuclear ambitions, was a major defeat for Tehran and a major victory for Washington.

The second component of Iran's strategy to undermine the United States' containment measures is to move closer to states that could counterbalance the United States. Iran has signed major economic and military agreements with China and Russia. It sees these two countries as natural allies, since they oppose the United States' unilateralism and its efforts to isolate Iran and have only reluctantly backed the sanctions against Iran. But the fact that they have supported the UN sanctions at all has proved to Tehran that when pressed, Beijing and Moscow are more likely to gravitate toward Washington than toward Tehran. Russia has not yet finished building the Bushehr nuclear reactor, which it had committed to completing by 2006, and

Moscow may be willing to pressure Tehran to change its nuclear policies if the Obama administration decides not to build antimissile systems in Russia's neighborhood.

But Iran continues its efforts. With little alternative but to rely on China and Russia as counterweights to the United States, it recently asked to upgrade its status in the Shanghai Cooperation Organization, a six-party security organization that includes China and Russia, from observer to full member so that it would receive assistance from other members if it were ever attacked. It has also become much more active in trying to popularize anti-Americanism within the Non-aligned Movement and the Organization of the Islamic Conference, and it has solidified its ties with Washington's most vocal opponents in the United States' backyard: Bolivia, Cuba, Ecuador, Nicaragua, and Venezuela.

The third facet of Iran's countercontainment strategy is to use its energy resources to reward its allies. No country in the region is as well endowed with energy resources as Iran is, not even Saudi Arabia: Iran's oil reserves total about 138.4 billion barrels, and its natural gas reserves total about 26.5 trillion cubic meters. (Although Saudi Arabia's oil reserves, which total approximately 267 billion barrels, are larger than Iran's, its natural gas reserves, which total about 7.2 trillion cubic meters, are much smaller.) Oil diplomacy has long been a strategy of Tehran's, of course. During the Rafsanjani presidency, it briefly served as a means to start normalizing relations with Washington. In that spirit, Iran signed in early 1995 a \$1 billion oil deal with the U.S. energy company Conoco, the largest contract of its kind since 1979. But the deal was soon terminated when, under pressure from the U.S. Congress and U.S. interest groups opposed to any opening toward Iran, President Bill Clinton issued an executive order banning U.S. companies from investing in Iran's energy sector. A year later, the Iran-Libya Sanctions Act was passed, which penalizes foreign companies that invest more than \$20 million in Iran's energy industry. In reaction, in 1997 Tehran signed a \$2 billion deal with the French oil and gas company Total.

Meanwhile, many Western companies that have continued to want to do business in Iran have struggled to bid for its huge and untapped natural gas reserves. And to immunize itself against the effects of the

sanctions and any potential boycott by the West, Iran has shifted its oil trade from the West to new markets. Before the 1979 revolution, the top five importers of Iranian oil were, in decreasing order, France, West Germany, the United Kingdom, Italy, and Japan. By 2008, they were Japan, China, India, South Korea, and Italy. Iran has also recently helped open up the Persian Gulf to China and Russia, signing multibillion-dollar contracts with the Chinese company Sinopec and granting Russia major concessions and access to the Azadegan oil field. Khamenei has even proposed forming with Russia a natural gas cartel modeled after OPEC.

Despite U.S. opposition, Iran has also made good progress on the construction of a so-called peace pipeline that would carry gas from the Persian Gulf to India through Pakistan, a project that would strengthen Iran's position as a major source of energy for those two countries. Nor should one underestimate the negative long-term implications for U.S. interests of China's and Russia's increasing involvement in the Persian Gulf, which Iran has facilitated, or of Tehran's recent move to use the euro in its international transactions, which has weakened the dollar. On the other hand, the sanctions have hurt Iran badly. Its plan for gas and oil pipelines that would connect the Persian Gulf to the Caspian Sea has stalled because of the restrictions. And its oil industry has been deprived of access to important modern technologies: as a result, Iran's oil production today remains significantly below what it was in 1979. In other words, the sanctions have been a lose-lose economic proposition for both the United States and Iran.

RIISING IN THE REGION

AFTER THREE decades of Washington's containment policies, Iran has nonetheless emerged as a regional power. The collapse of the Soviet Union allowed Tehran to expand its influence in the former Soviet republics, many of which it shares historical ties with. A decade later, the United States accelerated the process by overthrowing the Taliban and Saddam Hussein, Iran's neighboring nemeses. And by failing to reactivate the Arab-Israeli peace process and mismanaging the occupation of Iraq, Washington created enticing opportunities for Iran

to expand its power. For the first time in a long while, Iran's influence now radiates west, north, and east. Iran now rightly considers itself an indispensable regional player.

These ambitions pit Tehran against Washington. As Mohsen Rezaei, a former commander of the Revolutionary Guards, stated in 2007, "It is our principal and indisputable right to become a regional power," and the United States "would like to prevent us from playing such a role." According to a March 2007 article in *The New York Times*, a recent UN sanctions package against Iran's nuclear program was passed in order to rein in what U.S. officials saw as "Tehran's ambitions to become the dominant military power in the Persian Gulf and across the broader Middle East."

A pivotal element of Iran's strategy of neutralizing the United States' containment policy is to create spheres of influence in Syria, Lebanon, and among the Palestinians, as well as in Afghanistan and Iraq, by supporting pro-Iranian organizations and networks there. (As Rezaei put it, "Iran has no meaning without Iraq, Lebanon, Palestine, and Syria.") An especially controversial part of this strategy is Iran's support for Syria, Hezbollah, Palestinian Islamic Jihad, and Hamas—the rejectionist front in the Arab-Israeli conflict. Iran's three-decade-long alliance with Syria is one of the most enduring alliances between Middle Eastern Muslim countries since the end of World War II. Iran's support of the Shiites of Lebanon and the Palestinians goes back many years. Rafsanjani was incarcerated in the 1960s for translating a pro-Palestinian book into Persian; Khomeini condemned the shah in 1964 for his *de facto* recognition of Israel; Khomeini also authorized Hassan Nasrallah, the leader of Hezbollah, to collect religious taxes on his behalf in support of the Lebanese Shiites; and many of Iran's current leaders received training in Lebanon in the 1960s and 1970s. At first, Iran's support for Hezbollah and the Palestinians had an ideological basis; now, it has a strategic rationale. It gives Tehran strategic depth in the heart of the Sunni Arab world and in Israel's backyard, which translates into a retaliatory capacity against Israel, as well as bargaining power in any future negotiations with the United States. Moreover, after centuries of using its influence mostly to defend Shiites, Iran is now increasingly trying to transcend the sectarian divide by supporting

the Sunni groups Hamas and Palestinian Islamic Jihad. This, in turn, has undermined the regional position of such powerful Sunni countries as Egypt and Saudi Arabia.

That said, Iran's financial and logistical support for Hamas and Islamic Jihad should not be exaggerated. Tehran remains a peripheral player in the Arab-Israeli conflict. It has no compelling national interest in the dispute and is simply taking advantage of the failure of the peace process between Israel and the Palestinians. What makes Iran an influential player is not its financial support alone—Saudi Arabia and other Arab countries contribute substantially more to the Palestinians—but also the model of resistance it champions. Iran has helped Hezbollah develop an approach that combines Islamic solidarity, populism, some trappings of democracy, strict organizational discipline, extensive economic and social support for the needy masses, and pervasive anticolonial and anti-Western sentiments—all in an effort to mobilize the streets of the Islamic world against Israel and the United States and expand its own power. The effectiveness of that model, and of its asymmetric strategies, was on display during Hezbollah's 34-day war with Israel in 2006. The group's use of antitank missiles and portable rockets, which Israel claimed Iran had provided—a charge Iran has denied—inflicted enough damage on Israeli cities to create havoc and mass fear. Hezbollah appeared to have won because Israel could not score a decisive victory against it; the conflict marked the first time that an Arab force was not humiliatingly defeated by Israel. It boosted Hezbollah's popularity in many Sunni countries, gave Iran more credibility in the region, and undermined Washington's traditional allies, such as Egypt and Saudi Arabia, which had not supported Hezbollah. The war, along with the chaos in Afghanistan and Iraq and what the Iran expert Vali Nasr has called "the Shiite revival," has convinced Tehran that a new order is emerging in the Middle East: the United States no longer dominates, and Iran now plays a major role.

Tehran is trying to neutralize Washington's containment strategy by expanding its own influence in the Middle East.

claimed Iran had provided—a charge Iran has denied—inflicted enough damage on Israeli cities to create havoc and mass fear. Hezbollah appeared to have won because Israel could not score a decisive victory against it; the conflict marked the first time that an Arab force was not humiliatingly defeated by Israel. It boosted Hezbollah's popularity in many Sunni countries, gave Iran more credibility in the region, and undermined Washington's traditional allies, such as Egypt and Saudi Arabia, which had not supported Hezbollah. The war, along with the chaos in Afghanistan and Iraq and what the Iran expert Vali Nasr has called "the Shiite revival," has convinced Tehran that a new order is emerging in the Middle East: the United States no longer dominates, and Iran now plays a major role.

WHERE THE HARD THINGS ARE

THE COMPLICATED nature of the U.S.-Iranian relationship is most evident in Afghanistan and Iraq, where the convergences and divergences of the two sides' interests are the clearest. After the Soviet occupation of Afghanistan in 1979, Tehran became intensely engaged with its neighbor, and Iran subsequently became home to some two million Afghan refugees. Gradually, throughout the 1980s, it built new alliances and new networks with Shiite and Persian- and Dari-speaking minorities. (As the Afghanistan expert Barnett Rubin has put it, during that period, "ironically, the United States was indirectly aligned with 'fundamentalists' while Iran courted the 'moderates.'") Then, in the 1990s, while Pakistan and Saudi Arabia were providing critical support to the Taliban government, which itself backed al Qaeda, Tehran created a sphere of resistance in Afghanistan by supporting the Northern Alliance—a force that cooperated with the invading U.S. troops in 2001 in order to liberate Afghanistan from the Taliban. In helping dismantle the Taliban, in other words, Tehran effectively sided with the U.S. government—even providing Washington with intelligence.

Tehran maintained its policy toward Afghanistan even after U.S. President George W. Bush said Iran belonged to "an axis of evil." Today, still, it entertains close relations with the pro-U.S. government of Afghan President Hamid Karzai. And the convergence between Tehran's interests and Washington's interests in Afghanistan remains substantial. Both want to keep the country stable and prevent the Taliban's resurgence. Both want to control and possibly eliminate drug trafficking, the economic backbone of the region's terrorists and warlords. Both want to defeat al Qaeda (which considers Shiism to be a heresy). And both want to eventually rebuild Afghanistan.

At the same time, Iran's heavy involvement in the reconstruction of Afghanistan has allowed it to create a sphere of economic influence in the region around Herat, one of the most prosperous regions in the country. This, in turn, has helped stabilize the area by preventing al Qaeda and the Taliban from infiltrating it. Iran has also empowered the historically marginalized Afghan Shiites, such as the Hazara and the Qizilbash, who constitute about 20 percent of the Afghan population. At a donors' conference in Tokyo in January 2002, Iran pledged \$560 mil-

lion for Afghanistan's reconstruction, or approximately 12 percent of the total \$4.5 billion in international reconstruction assistance that was promised then. During a donors' conference in London in 2006, it pledged an additional \$100 million. And unlike many other donors, it has delivered most of its promised assistance. The bulk of the funds are targeted at developing projects for infrastructure, education, agriculture, power generation, and telecommunications. Iran hopes to become a hub for the transit of goods and services between the Persian Gulf and Afghanistan, Central Asia, and possibly also China.

This quest for influence in Afghanistan pits Iran against the United States in some ways. For example, Tehran opposes the establishment of permanent U.S. bases in Afghanistan. And to ensure that Washington will not be able to use Afghanistan as a launching point for an attack on Iran, Tehran is pressuring Kabul to distance itself from Washington. Uncertain about Afghanistan's future and Washington's intentions in the country, Iran is keeping its options open and trying to increase all its possible retaliatory capabilities against the United States. It maintains close ties with the Northern Alliance, as well as with warlords such as Ismail Khan, various Shiite organizations, and the insurgent leader Gulbuddin Hekmatyar and other anti-American fighters. It is turning the region around Herat into a sphere of influence: the bazaars there are loaded with Iranian goods, the area receives the bulk of Iran's investments in the country, and the Revolutionary Guards are reportedly visible and active.

The United States and Iran have tried to strike a fine balance in Iraq as well, but with much less success. If anything, Iraq has become center stage for their rivalry; there they have some common goals but also many more diverging ones. Iran's top strategic priority in Iraq is to establish a friendly, preferably Shiite government that is sufficiently powerful to impose order in the country but not powerful enough to pose a serious security threat to Iran, as Saddam did. Iran was the first country in the region to recognize the post-Saddam government in Baghdad. Since then, it has provided Baghdad with more support than even the staunchest of the United States' allies. It has a close relationship with the two parties that dominate the government of Prime Minister Nouri al-Maliki, as well as with the two major Kurdish parties. Like Washington, Tehran supports Iraq's stability, its new

constitution, and its electoral democracy, albeit in the parochial interest of ensuring the dominance of the country's Shiite majority. Like Washington, Tehran opposes Iraq's Balkanization, in its case partly for fear that such fragmentation could incite secessionist movements within Iran's own ethnically rich population. And like Washington, Iran considers al Qaeda in Iraq to be an enemy and seeks to eliminate it.

But as in Afghanistan, Iran is eager to engage in Iraq's reconstruction mainly in order to create an economic sphere of influence in the country, especially in the predominantly Shiite south, where many people of Persian descent live. It has pledged to spend more than \$1 billion for Iraq's reconstruction. Tehran seems to believe that with its existing influence in southern Iraq, including close ties to the major Shiite seminaries in Najaf, it can transform the region into a kind of southern Lebanon, creating a ministate within a state.

And then there are some major disagreements between Tehran and Washington. Tehran is determined to keep Washington mired in Iraq and prevent it from scoring a clear victory there. During the sectarian violence in 2004-7, Tehran supplied weapons to Shiite insurgents in Iraq, who then used them against U.S. troops. It supported the Mahdi Army and its founder, Muqtada al-Sadr, the radical Shiite cleric who opposes the U.S. presence in Iraq. Tehran is also vehemently opposed to the establishment of permanent U.S. bases in Iraq, for fear, as with those in Afghanistan, that the United States could use them to attack Iran. The status-of-forces agreement signed by the United States and Iraq in 2008 does seem to have diminished some of Iran's concerns, however. The agreement stipulates that U.S. forces will withdraw from Iraq no later than December 31, 2011, and that "Iraqi land, sea, and air shall not be used as a launching or transit point for attacks against other countries."

Iran's policies toward Iraq in the past few years suggest that when Iran feels threatened and its legitimate security needs and national interests are ignored or undermined, it tends to act more mischievously than when it feels secure. Its Iraq policy, therefore, is directly correlated with its perception of the threat posed by the United States. The security talks between Tehran and Washington launched at the urging of the Iraqi government in 2005 are thus very important. After these meetings began, and after the U.S. government launched its "surge"

strategy in Iraq, the level of violence in Iraq subsided. Iran played a role in stabilizing the situation by pressuring its allies, including the Mahdi Army, to refrain from violence against Sunnis or U.S. troops. The simple fact that Baghdad is a close ally of both Tehran and Washington offers a chance for those two governments to build on their interests in Iraq.

FULL ENGAGEMENT

ANTI-AMERICANISM IS not an insurmountable obstacle to normalizing relations with Iran. For one thing, Iran's elites are heterogeneous; they consist of two rival factions, both of which have come to favor, like a significant portion of the population, normalizing relations with the United States. For another, *mashhat*, or "expediency," is a defining feature of Iranian politics. Even the most ideological of Iran's leaders favor a cost-benefit approach to decision-making. According to the 2007 U.S. National Intelligence Estimate, Iran halted its nuclear weapons program in 2003 based on a cost-benefit calculation. Although the accuracy of that conclusion is debated, there is no question that Tehran has often resorted to that approach. When Iran needed advanced weapons during the Iran-Iraq War, Khomeini approved secret dealings with Israel and the United States, culminating in the Iran-contra fiasco. Despite its general opposition to the presence of U.S. troops in the region, Tehran remained actively neutral during the 1991 Persian Gulf War, seeing it as an opportunity to weaken its archenemy Saddam and improve relations with the West. The Revolutionary Guards, the most ideological group in the Iranian armed forces, rubbed shoulders with U.S. forces when they assisted the Northern Alliance in overthrowing the Taliban in 2001. Far from being a suicidally ideological regime, Tehran seeks to ensure the survival of the Islamic Republic while advancing the country's interests through negotiations.

Iranian policy toward the United States has a logic. It is a logic driven not by a single faction or a single issue but by a stable and institutionalized system of governance with both authoritarian and democratic features, with domestic constituencies and long-standing international alliances. It is a logic that made Iran into a regional power with

substantial influence in Afghanistan, Iraq, Lebanon, and the Palestinian territories, and among millions of Muslims around the world. And it is a logic that, despite mounting international pressure, has made it possible for Iran to make advances in asymmetric warfare, nuclear technology, uranium enrichment, and missile and satellite technologies. Now, Iran legitimately demands that Washington recognize these advances and Tehran's new role as a major regional power.

Unless Washington understands that Tehran's U.S. policy has a rationale, it will not be able to develop a reasonable long-term strategy toward Iran. Invading the country is not a viable option. Nor are so-called surgical strikes against Iran's nuclear facilities, which would most likely lead to a protracted retaliation by Iran; a Tehran more defiant and more determined to become a nuclear weapons power; more terrorism; greater instability in Afghanistan, Iraq, Lebanon, and the Persian Gulf; and higher oil prices.

The challenge for the U.S. government is to give Iran incentives to reevaluate its strategy toward the United States. A carrot-and-stick approach designed to stop progress on Iran's nuclear program is unlikely to work. Focusing on a few contentious issues, such as Iran's uranium-enrichment activities, would do little to change the fundamental logic of Iran's U.S. policy. Moreover, the stick part of that approach would only strengthen the anti-American constituencies in Iran while hurting its people. Nor will democracy promotion work. More a feel-good fantasy than a viable strategy, this approach misleadingly assumes that democracy can be exported, like cars, or imposed by force and that a democratic Iran would no longer have any serious conflicts with the United States or pursue nuclear ambitions. Iran was considerably more democratic under Mosaddeq than under the shah, and yet its relations with the United States were much worse then.

A better approach is a strategy of full engagement, one predicated on gradually increasing economic, educational, and cultural exchanges between the two countries; exploiting the commonalities shared by their governments; and establishing concrete institutional mechanisms to manage their remaining differences. Washington must recognize that there is no diplomatic magic wand that can fix its "Iran problem" overnight; normalizing U.S.-Iranian relations will be a long and difficult process. Unless Tehran and Washington make a strategic decision to

normalize relations, the many forces that continue to pull them apart are likely to derail the process.

As a first step, the United States should allay Iran's fears about regime change. It can do this by explicitly recognizing that Khamenei is the center of gravity in Iran's decision-making process and establishing a line of communication with his office. Holding direct, comprehensive, and unconditional negotiations with the Iranian government is Washington's least bad option. The two countries' negotiating teams must meet face-to-face to learn firsthand about each other's priorities and interests on all the important issues and break the psychological barriers that have kept the parties apart for three decades. Meanwhile, Washington should provide assurances to Israel and its Arab allies that they should not fear its rapprochement with Tehran and that Iran's nuclear policy will remain the main item on the United States' Iran agenda.

As the Obama administration reviews its options when it comes to Iran, it would do well to examine how, three decades ago, President Richard Nixon brought China back into the community of nations. It took almost eight years after the secret trip by Henry Kissinger, then U.S. national security adviser, to Beijing in 1971 for the United States and China to establish diplomatic relations. Anti-Americanism under Mao Zedong, China's support for North Vietnam, and China's arsenal of nuclear weapons were infinitely more threatening to the United States than Iran's policies are now. Yet Nixon and Kissinger had the foresight to map out a new strategic landscape for Beijing. They did not punish it for its policies of the past; they gave it a reason to want something better in the future. And then the two countries built a better relationship on their common recognition of the threat posed by Soviet expansionism. Washington can, and should, do something similar with Tehran today and finally end three decades of hostility by highlighting the two governments' shared interests in defeating al Qaeda and stabilizing Afghanistan and Iraq. Tehran, for its part, must recognize that without some kind of understanding with Washington over the issues that matter to the U.S. government, it will not be able to fully benefit from its recent ascent as a regional power—and could even lose much of what it has gained. ☺



Taylor & Francis
Taylor & Francis Group



AAG
Association of American Geographers

Just War and Extraterritoriality: The Popular Geopolitics of the United States' War on Iraq as Reflected in Newspapers of the Arab World

Author(s): Ghazi-Walid Falah, Colin Flint and Virginie Mamadouh

Source: *Annals of the Association of American Geographers*, Vol. 96, No. 1 (Mar., 2006), pp. 142-164

Published by: Taylor & Francis, Ltd. on behalf of the Association of American Geographers

Stable URL: <http://www.jstor.org/stable/3694149>

Accessed: 11/12/2013 09:04

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

NOTICE:
This material may be protected by copyright law (Title 17 U.S. Code)



Taylor & Francis, Ltd. and Association of American Geographers are collaborating with JSTOR to digitize, preserve and extend access to *Annals of the Association of American Geographers*.

<http://www.jstor.org>

Just War and Extraterritoriality: The Popular Geopolitics of the United States' War on Iraq as Reflected in Newspapers of the Arab World

Ghazi-Walid Falah,* Colin Flint,† and Virginie Mamadouh‡

*Department of Geography and Planning, University of Akron

†Department of Geography, University of Illinois

‡AMIDSt, Department of Geography, Planning and International Development Studies, University of Amsterdam

As with all wars, the U.S. military invasion of Iraq in 2003 needed to be portrayed as a just war in an attempt to garner support and legitimacy, domestically and internationally. The United States was acting as hegemonic power in the international state-system and, in light of this role, had imperatives and tools in creating the argument for a just war that differed from those used by nonhegemonic states. The United States acted extraterritorially by diffusing a message of moral right. Arab resistance to the war was evident in the construction of the United States and its leadership as immoral, precluding its ability to wage a just war. This article focuses on the Arab response by analyzing the portrayal in Arab newspapers of the imminent war on Iraq. Sixty-five newspapers of the Arabic language (plus the Iraqi news agency), published in seventeen Arab countries, of which four were Iraqi newspapers, were consulted for the purpose of this study. Interpretation of the geopolitical rhetoric within newspaper reports and political cartoons published in Arab newspapers highlights the way that arguments of morality and immorality were connected to understandings of territorial sovereignty and hegemonic extraterritorial influence into territorial sovereign spaces. *Key Words:* Arabic newspapers, Arab world, geopolitics, hegemony, just war, United States.

States need to justify the wars they wage: a war necessitates both internal and external legitimacy. It must be constructed as just and perceived as just by citizens and by other actors in the state system (Walzer 2000). How states justify their wars differs greatly however. The United States of America, acting as hegemonic power, constructed its 2003 war on Iraq as just with rhetorical tools unavailable to nonhegemonic states (Dalby 2003b; Flint and Falah 2004).¹ Moreover, the hegemonic power needs to justify its war to the citizens of other states, particularly to the peoples of the Arab world. We contend that the moral and geographic components of the hegemonic mission provided ammunition for its critiques. In other words, the tropes used by the United States to justify hegemonic military actions in Iraq were also used by Arab (and other) commentators to argue that they were unjust. The rhetorical battle over the justness of hegemonic military activity was founded on a geographic tension: the extraterritorial reach of the hegemonic power and its need to intervene in a sovereign state on the one hand, and the resort to the legitimacy of sovereign territory on the other.

All states must legitimize the wars that they fight, whether they are relatively powerful or weak, democratic or authoritarian. Also, given the role of the United

Nations and global media, all states must justify their actions to an international as well as a domestic audience. For example, Long (2004) argues that Saddam Hussein attempted to legitimate his 1991 invasion of Kuwait using a combination of Arab nationalism and distorted interpretations of Islam. The response was quite effective in the Arab world, though less so in Western countries, requiring a response by Arab countries supporting the U.S. military action (Long 2004). Further examples of political authority being undermined by military actions that do not resonate with the public are provided by the Vietnam War protests in the United States and the Bolshevik agitation in Czarist Russia.

The differential power of states is manifested in their ability to *make* others do what they want or to *persuade* them to act as desired, using what Boulding (1990) has described as coercive and integrative forms of power. Though states weak in coercive power are unlikely to attack stronger states, when weak states fight they must legitimize their actions using the same principles as strong states. These principles are used in diplomacy, but are built on the content of just-war theory. In this article, we argue that hegemonic powers, dominant in both coercive and integrative power, have particular impera-

tives for making a case that the wars they fight are just wars, and they have specific tools available to make their case.

Hegemonic powers are able to extend their influence into the sovereign spaces of other states (Arrighi 1994; Bacevich 2002; Flint 2003; Lustick 2003). Influence may be in the form of diplomatic pressure, emulation of economic and social practices, or military invasion (Agnew 1993; Agnew and Sharpe 2002). Such extraterritorial reach is facilitated by the projection of a prime modernity, or the hegemonic power's definition of modern society (P. Taylor 1999; Agnew and Sharpe 2002). In combination, extraterritoriality and the prime modernity are the foundations of hegemonic power that rest on a geography of universalism transcending the political spaces of nation-states. In turn, universalism requires the hegemonic power to legitimate the wars it initiates in ways that are different from those of non-hegemonic states.

Critics of wars waged by the hegemonic power have specific means of labeling the aggression as unjust that are different from the rhetoric used to claim that wars by nonhegemonic states are unjust. Specifically, two rhetorical battles over the legitimacy of the United States' military actions are the universalism of the hegemonic mission, which results in a symbolic conflict over the morality of its military actions, and its extraterritorial reach, which provokes a rhetoric of resistance centered on the just defense of sovereign space. Both of these rhetorical battles are examined in this article by specific reference to the 2003 invasion of Iraq by U.S.-led forces. Although supporters and critics all over the world engaged in such rhetorical battles, this article concentrates on the Arab world as the most critical region where the United States needed to convince the population of the justness of its war.

As Ó Tuathail and Dalby (1998, 5) have emphasized,

Geopolitics saturates the everyday life of states and nations. Its sites of production are multiple and pervasive, both 'high' (like a national security memorandum) and 'low' (like the headline of a tabloid newspaper), visual (like the images that move states to act) and discursive (like the speeches that justify military actions), traditional (like religious motifs in foreign policy discourse) and postmodern (like information management and cyberwar).

Our focus here is on some of those discursive production sites in the infospace realm of what Ó Tuathail dubs "popular geopolitics," centering in part on aspects of geopolitical information management within print discourse (Ó Tuathail 1996). Monuments and architecture and media sources such as television, music, film,

and the printed press can all be read as geopolitical texts (Dodds 2000; Sidaway 2003). Many interpretations of the media's portrayal of the U.S.'s hegemonic mission have appeared (Dalby 2003a; Mamadouh 2003; J. Taylor and Jasparo 2003), but none has focused on the response by opinion-makers in Arab countries. Our analysis scrutinizes satirical cartoons because they are expected to be effective tools to challenge both the foreign policy of a regime and the hegemonic position of the United States and to reach a large section of the population.

The article is organized into four sections. First, we explore the relationship between just-war theory and the particular case of U.S. hegemonic power. Second, we review arguments in the Iraqi media to expose the war as unjust. Third, we examine media coverage in the rest of the Arab world from the point of view of geopolitical relations between the United States and the Arab world. Fourth, we examine political cartoons to explore Arab constructions of the U.S. war against Iraq as an unjust war, and to illustrate how the notion of the universal morality of the hegemonic power was challenged.

Just War and Hegemony

The costs of war are high. Lives of combatants and noncombatants are lost, families are devastated; wounds are physical as well as psychological. Public coffers become exhausted, the opportunity cost is felt across the economy and public expenditure. Domestically, failure in war undercuts the authority of political elites. Internationally, legitimacy is at stake: Aggression results in diplomatic isolation and military failure can provoke subsequent political challenge. In light of these potential costs, political leaders must legitimate acts of war to domestic and international audiences.² All wars are portrayed by the propagators as being "just" (Ginsberg 1969). Any war that is judged as unjust is a war disliked (Walzer 2000, 12), provoking opposition and subsequent political conflicts within a country and in the international state system.

Modern just-war theory focuses on the state as the legitimate actor, in contrast to the premodern focus on individual actions for the sake of individual honor (Temes 2003). Just-war theory sits somewhere between realism, war cannot and must not be avoided, and pacifism, war must be avoided whatever the cost (Walzer 2000). Temes (2003, 13) sums up the just war theory as an understanding that war is "wrong, but necessary." Using the horrors of war to justify inaction in the face of obvious injustice (famously, Neville Chamberlain's appeasement of Hitler) is considered immoral.

The seed for modern just war theory was sown in the fifth century AD by Augustine, establishing the theoretical role of catholic jurisprudence in his statement "We go to war that we may have peace" (Temes 2003, 9): war was a project for the common good. During the Enlightenment, the common good was defined at the scale of the emergent modern state. Following the philosophies of Hegel and Kant, war was fought by individuals, but for a collective state project (Orend 2000c; Temes 2003). Grotius forced consideration of the manner in which a war was fought into the judgment of whether a war was or was not just (Temes 2003, 47–49), and Kant reflected on the establishment of a just peace (Orend 2000c). In combination, the modern approach to just war has focused on states as the legitimate actors, with the reasons for just war focusing on the defense of sovereign territory.

At an early moment in American hegemony, the geographic scale of the common good underlying a just war was expanded. In 1917, President Woodrow Wilson justified the American entry into World War I on the basis of defending human rights; war for an ideological mission, global or universal in scope (Temes 2003, 22). Universal claims had been the basis for the Jewish, Christian, and Islamic interpretations of just war (Temes 2003), but now the universalism was based on a definition of human rights that was not constrained by state sovereignty. Wilson's speech maintained the modern understanding of just war: states fight wars to maintain human rights; the standards of human rights are universal and are an issue to be policed at the global scale. The United States as an emerging hegemonic power defined universal standards of human rights and also identified itself as the global policeman protecting those standards.

Historically, the U.S. attitude to just war has, of course, been contested. On the one hand, government policy has centered on two strategies, both of which are vibrant in contemporary debates regarding the means of prosecuting the War on Terrorism. Temes sums up the two views as "wage war carefully—a multilateralist, humanitarian approach with all due attention paid to minimizing harm in every day of every battle; or wage war with overwhelming force, to make the process brutal, decisive, and brief, so that peace may be achieved sooner, even if on bloodier wings" (Temes 2003, 84). These historical perspectives are still relevant. The official description of the war on Iraq wrestles with both of these approaches; terms such as the "coalition of the willing," "shock and awe," and reducing "collateral damage" show that the Bush administration is well aware of the defining tropes of the language it must use

to justify the manner in which the War on Terrorism is fought.

In reference to the consistent themes across U.S. administrations, groups within civil society have critiqued the reasons for U.S. military action and have helped to redefine the reasons justifying war. A significant interjection was made through the "Pastoral Letter on War and Peace" issued by the Conference of Catholic Bishops in the United States in 1983 (Temes 2003, 76). The letter criticized the intentional destruction of cities by nuclear attack as an unjust means of waging war. Yet Pope John Paul II's catechism of 1992 reinforced the globalism of the just-war doctrine that had been introduced by Wilson. The catechism defined peace as "safeguarding the good of persons, free communications among men, respect for the dignity of persons and peoples, and the assiduous practice of fraternity" (Temes 2003, 77). In other words, war may be practiced to defend a global standard of human rights (though with apparent gender biases) rather than national self-defense. Interpretation of the catechism would allow for the type of claims to be found in a Cold War document such as NSC-68 as well as the 2002 National Security Strategy (or the "Bush Doctrine"), in which the United States was able to define human rights and judge when military intervention in their defense was just or not.

In contrast to the globalism and extraterritoriality of American hegemonic practice and ideology, theories of just war have rested on the geographic axioms of mainstream social science—that society can be equated with the sovereign spaces of nation-states and, hence, nation-states are the prime scale of analysis to the relative exclusion of nonstate actors at different geographic scales (Murphy 1996). Following this implicit geographic framework, Walzer's (2000) exposition of just-war theory focuses on invasion, and threat of invasion, of sovereign territory as the basis for fighting a just war.³ According to Walzer, another just reason for initiating war is to prevent severe abuses of human rights. But this is a harder case to argue, precisely because so many incidents of widespread atrocities are left unchallenged, with inactivity justified by resort to the world political map of nation-states and understandings of sovereignty. In sum, just-war theory rests on the geographic axiom of territorial sovereign nation-states.

Contemporary warfare rarely appears as a territorial conflict between states (O'Loughlin 2004). Civil wars, resource wars, various expressions of genocide, terrorism, and insurgencies are the prevalent modes of armed conflict (Flint 2004a). On that ground, the geographic axioms underlying the construction of wars as just must be reevaluated (Rengger 2002; Murphy 2003). In a step

toward that goal, this article focuses on the role of hegemonic powers in the international state system because, by definition, they extend their power beyond their borders and into other sovereign states. Hegemonic power rests on such an extraterritorial reach, and so faces particular hurdles in constructing its wars as just.

Hegemonic power is defined as the dominance of one state in the international state system. Originally, world-systems theorists (e.g., Wallerstein 1984) conceived of it as economic dominance, but subsequent work has emphasized the political and cultural components of hegemony (P. Taylor and Flint 2000, 67–77; Wallerstein 2002; Agnew 2003a; Anderson 2003; Smith 2003; Wallerstein 2003). Previous hegemonic powers, the Dutch in the seventeenth century and the British for much of the nineteenth century (P. Taylor and Flint 2000, 67), exercised extraterritorial power through their projection of philosophical and economic mandates, and through military strength. The United States is seen by some to be different, however, in that it has taken extraterritoriality to new levels that have fundamentally changed the world (Agnew 1993; P. Taylor 1993). In the historically consistent hegemonic goal of facilitating global economic flows, the United States has promoted the project of globalization (Hardt and Negri 2000). Some say this project has changed the balance between state sovereignty and global flows in an unprecedented and irreversible manner (P. Taylor 1993). Others have interpreted globalization as a moment in the hegemonic cycle of the United States, implying that globalization is a “period” rather than a “new age” (Arrighi 1994; Flint 2004b). Whatever the long-term significance of the contemporary moment, the extraterritorial reach of the United States is interpreted as being of an intensity that is unparalleled, and includes military power, economic influence, and cultural diffusion.

The world-systems approach to hegemony has adopted some of the Gramscian approach to hegemony, a cultural power of persuasion and normative values rather than the exertion of force (Boulding 1990; Arrighi 1994; Mitchell 2002). The exercise of hegemonic power by emulation is at the heart of P. Taylor’s concept of prime modernity. For P. Taylor (1999), each hegemonic power has defined what it means to be “modern” by depicting the social relations, regimes of production, social landscapes, and art forms that launched its economic success. Hegemonic leadership becomes a matter of defining modern society in a way that provokes a hybrid politics of envy, resistance, and emulation. Some countries, such as North Korea, reject it outright. Other countries, such as France, attempt official policies to challenge prime modernity’s cultural supremacy that are

limited in their success by popular tastes. The prime modernity is always translated through geographically specific cultural lenses and practices, so that the “template” offered by the hegemonic power is always contextualized. The greater the extent of emulation of the prime modernity the less the hegemonic power need resort to force.

Prime modernity rests on an extraterritorial assertion: The form of society epitomized by the hegemonic power is argued to be universally applicable, beneficial, and desired. In reality, the universalism of prime modernity is challenged constantly. The exercise of hegemonic authority is not bound by the territorial assumptions that were the foundations of just-war theory. Instead, the authority of the hegemonic power is challenged whenever a country resists the appeals of the prime modernity; the integrating strength of the prime modernity rests on the fiction of its universalism. Indeed, the war on terrorism has been portrayed in this way. Jean Bethke Elshstain’s (2003, 6) argument is as simple as it is stunning: “We must and will fight—not in order to conquer any countries or to destroy peoples or religions, but to defend who we are and what we, at our best, represent.”

The prime modernity of the United States, in its contemporary role as hegemonic power, is expressed by Elshstain as a moral imperative that is globally desired and universally applicable. Defending the prime modernity, the American “way of life,” becomes the basis for a just war, a moral response, because of the extraterritorial assumptions of hegemonic universality. Threat is not just military challenge but political challenge to hegemonic authority, as can be seen in the identification of the “axis of evil” (Agnew 2003b; Dijkink 2003).

The construction of the U.S. war against Iraq as just rested on both the territorial and extraterritorial elements of a just war. The issues of weapons of mass destruction (WMD) and Saddam Hussein’s alleged linkages to Al-Qaeda were grounded in the established tenets of defense of sovereign territory to justify a preemptive war. But lack of evidence of both WMD and terrorist sponsorship has meant that authority to wage war rested on the moral power of the United States. Belief in the evil of Hussein’s regime and his aggressive plans depended on a moral authority, a morality that was hegemonic in its ability to define the proprietary of acts conducted in other sovereign spaces.

Moral authority to wage war has been a central component of just-war theory. Violation of sovereign territory was portrayed as a violation of the norms of the international state system; morally, a response was expected (Orend 2000b; Walzer 2000). Hegemonic power rests on a different morality, the assumed right to judge

acts within other sovereign territories and deem some incidents as demanding a response (such as Iraq) while other acts are ignored (the Congo, for example). If moral power is necessary in constructing a hegemonic war as just, it follows that resistance to war conducted by the hegemonic power may be centered on allegations of hegemonic immorality. Indeed, given the military superiority of the hegemonic power, and its economic muscle, no other course of resistance may be available. Extraterritorial morality is used by the hegemonic power to justify its acts of war. In turn, this moral mission is used by those resisting the military force of the hegemonic power to question its extraterritorial intervention. Morality and immorality, extraterritorial reach and national sovereignty, are juxtaposed in a rhetorical battle over the justness of exercising military might.

An Unjust War: The View from Iraq

When extraterritorial military expeditions are deemed necessary to maintain hegemonic authority, they face the opposition of rhetorical strategies emphasizing the just defense of national territory from hegemonic intervention. Understandably, Iraqi newspapers and the Iraqi news agency were all preoccupied with preparations for war, and focused on showing that the imminent war against them was unjust and that the American administration was the aggressor. Table 1 indicates the kind of topics addressed in the headings of the surveyed five Iraqi media sources in mid-March 2003,⁴ at the time of the 16 March 2003 summit meeting of the three leaders of the United States, Britain, and Spain in the Azores. The meeting closed an attempt by the United States, Britain, and Spain to pass a second UN Security Council resolution to give the rubber stamp to an American "preemptive" war against Iraq. Many commentators interpreted the Azores summit as a staged "go-to-war"

decision.⁵ Consistent with this interpretation, and the growing global protest against a war against Iraq, many cities witnessed demonstrations at the time of the summit.

In the Iraqi media, there was extensive reporting on the support of various organizations, world leaders, and professional and religious groups for Iraq in its current crisis. Headings in the newspapers highlighted phrases such as "American's threat," "American colonial aggression," or "American and British aggression." Such rhetoric gives extraterritoriality an immoral edge. The Iraqi media put great emphasis on the activity of European countries, especially France, Germany, and Russia, who did not see war as a solution to the Iraqi question. Any country, no matter how large, small, poor, or rich, that opposed the U.S. "solution" was considered a source of moral aid to Iraq. For example, the newspaper *Al-Iraq* (2003a) reported on the statement by the president of Angola that his country opposed the use of force against Iraq and that any solution to the problem should be consistent with UN decisions. The newspaper *Al-Jumhuriya* (2003c) reported that forty-two female members of the Japanese Parliament demanded that their prime minister use his diplomatic skills to stop an American attack on Iraq. *Al-Thawra* (2003b) reported that seventy-eight unions in Europe went on strike for fifteen minutes to protest the American threat to Iraq. This concentration on opposition to war signified the U.S. actions as against the will of the international community and, by extension, immoral.

Significantly, the Iraqi newspapers also pointed to the opposition to war in the "enemy's camp." For example, *Al-Jumhuriya* (2003b) reported on the resignation of political advisor John Brady Kiesling at the U.S. Embassy in Athens as a protest against the belligerency of the Bush Administration against Iraq. In the same article *Al-Jumhuriya* mentioned a threatened resignation in Tony

Table 1. Reports pertaining to the imminent war, published in selected Iraqi newspapers, mid-March 2003

Source	Iraqi News Agency 15 March	<i>Al-Jumhuriya</i> 15 March	<i>Al-Thawra</i> 14-15 March	<i>Al-Iraq</i> 15 March	<i>Alef-Ba</i> ^a 12 March
Total no. of headlines	28	6	30	13	5
Items related to current crises	28	6	26	12	5
Topics:					
Demonstrations in support of Iraq	2	—	1	3	1
People/Organizations and States support Iraq/against U.S.	15	4	18	3	—
Confrontation with U.S./British airplanes	2	—	1	1	—
UN Inspectors in Iraq	3	1	—	1	—
Issues related to the UN	1	—	—	1	1
Issues related to President Saddam Hussein	3	1	1	1	1
Other	2	—	5	2	2

^aA weekly magazine.

Blair's government. In another piece, entitled "British people support France's stance opposing American aggression," *Al-Jumhuriya* (2003a) reported that the French Embassy had received hundreds of letters and e-mails from British citizens who supported France's objection to an American attack on Iraq. *Al-Thawra* (2003a) reported on the resignation of senior intelligence analyst Andrew Wilky, who was quoted as saying that "all available evidence indicates that Iraq does not threaten any country and does not constitute a threat to anyone." Kuwait was also included in reports of countries where voices of opposition to an attack on Iraq could be heard and demonstrations had taken place. The Iraqi Foreign Minister, Dr. Naji Sabri, praised recent attacks carried out in Kuwait against American troops in Kuwait, blaming the Kuwaiti leaders for going against popular public sentiment (*Al-Iraq* 2003b). The reports of antiwar protest across the globe allowed the Iraqi leadership to show that it was not alone in resisting the universality of American ideas and practices. In other words, it could claim that its defense was a just conflict against an extraterritorial force, a force that was being rejected in other parts of the globe too.

On the internal Iraqi front, there was extensive reporting of a purported exchange of messages between the Iraqi people and the leadership. At a time of international conflict, the goal of Hussein's regime was to portray an internal unity, which denied regional, religious, and ethnic differences under the cloak of a just war against aggression. Commonly, in the event of impending war, leaders address their people, assuring them they should have faith in their country's ability to defeat any possible aggressor. At the same time, messages are sent from the population in a manner that serves to renew the social contract with their leadership, namely that the loyalty of the citizenship will be rewarded by leadership providing security and prosperity for the nation. The Iraqi newspapers portrayed a situation in which an Iraqi people and their leadership did not have any illusions about the fact that the United States was stronger in military and technological terms. However, that disparity did not prevent them from defining themselves as a nation united by its confrontation with American military might in the defense of national territory. Iraqi leaders put great emphasis on the popular will to fight, especially because the war was readily portrayed as an unjust invasion of Iraqi sovereign space.

Iraqi newspapers defined American aggression as unjust, and Iraqi defensive actions as a just response. The *Alef-ba* weekly magazine (2003) published a piece entitled "With resolve we will win the war," citing Saddam Hussein as saying to an Iraqi fighter—after the recent dis-

mantling of Sumud missiles at the request of UN inspectors (in Arabic, *sumud* means steadfastness) that "You are the Sumud Missiles, and if it was not for you and for the Iraqi spirit and determination to defend our homeland, these missiles would not have been found." In this piece, the American army is metaphorically described as the new Mongols of the twenty-first century (in analogy to the first Mongols who conquered Baghdad in 1258 CE in an earlier manifestation of extraterritoriality; Khashan 2003). Again the justness of defense and the unjustness of invasion were portrayed by reference to the geographic axiom of territorial sovereignty.

The notion of fighting a just war was a strand of rhetoric found in almost every speech by any Iraqi leader, part of the ideology conveyed to the masses to bolster their support. And there was reportage on what the people were doing to prepare for war. *Al-Iraq* (2003c, 2003d) carried two reports on activities in mosques in Baghdad and other cities on 14 March 2003 when Moslems prayed for the war to be avoided if possible, and if not, for Iraqi Moslems to fight a holy war. It also was reported that mosques would become centers for the defensive struggle against the invaders; they would recruit fighters for *jihad* and would help the people with food and medicine. In the same vein, *Al-Thawra* (2003c) published a *fatwa* (legal opinion), issued by Moslem religion scholars calling for *jihad* against any possible aggression aimed at Iraq and to block any sort of help for the aggressor from land, space, or sea. By including a religious message, the Iraqi leadership was able to lay claim to just authority at the "cosmic" scale of divine will (Juergensmeyer 2001). In this instance, justness was being sought at a higher geographic scale, the divine, and not in the realm of territory.

Unsurprisingly, in our analysis of Iraqi news media under Hussein's dictatorial regime, no opposing views supporting the U.S. intervention were found.⁶ The war was constructed as unjust by arguing both against U.S. intervention as trespassing international rules about the territorial sovereignty of states and against American moral grounds as hegemonic power: the U.S. emphasis on democracy and human rights was constructed as an excuse for intervention into the region.

Just or Unjust War? The View from the Rest of the Arab World

We interpreted the commentary, headlines, and imagery of sixty Arab newspapers collected across seventeen different Arab countries seven times in a twenty-one-day period during November 2002.⁷ This period was chosen because it was the time when debate regarding a

Table 2. Arabic newspapers' reportage of U.S. relations, November 2002

Country name	Newspaper	No. of newspapers	Total reports pertaining to the U.S.	Average
Morocco	<i>Al-alam, Al Anbaa, Attajdid, Assabah</i>	4	3	0.75
Algeria	<i>al-Fadjr, El Khabar, El Youm, An-Nasr, Al-Shaab</i>	5	8	1.6
Libya	<i>Al-Fajralijadeed, Al-Jamahiria, Al-Shames, Al-Zahfalakhder</i>	4	6	1.5
Egypt	<i>Alquds alarabi, Al Wafd, Al-Osboa, Akhbar Aladab, Alhram, Al-Sharq, Al-Awsat, Al-Shaab, Akhbar-al-Yom</i>	8	34	4.25
Sudan	<i>Alray Alaa'm</i>	1	6	6
Palestine	<i>Al-Ayyam, Al-hayat Al-Jadeeda, Al-Quds, Falasteen, Fasl-Almaqal</i>	5	17	3.4
Lebanon	<i>As-Safir, An Nahar, An Nabaa, Al-Mustaqbal, Al-Aman, Al Intiqad</i>	6	16	2.67
Syria	<i>Al Baath, Al-Thawra, Teshreen</i>	3	8	2.67
Jordan	<i>Assabeel Weekly, Alarab-Alyawm, Ad-Dustour, Al-Rai</i>	4	51	12.75
Kuwait	<i>Al Watan, Al Seyassah</i>	2	13	6.5
Bahrain	<i>Alayam, Alwasat, Akhbar Alkhaleej</i>	3	13	4.33
Qatar	<i>Al-Rayah, Al-Sharq, Al-Watan</i>	3	45	15
United Arab Emirates	<i>Al Bayan, Al-Ittihad, Al-Khaleej</i>	3	57	19
Oman	<i>Oman Daily, Al Watan</i>	2	4	2
Saudi Arabia	<i>Al-Jazirah, Okaz, Al-Riyadh</i>	3	32	10.66
Yemen	<i>Al-Ayyam, Al-Gumhuryah, Al-Sahwa, Al Thawra</i>	4	43	10.75
Total		60	356	5.93

UN resolution to disarm Iraq was gaining momentum, and so it reflects the politics of Arab reactions to the construction of just causes of war by the United States. This period of posing justifications for war and challenging them culminated in UN Security Council Resolution 1441, passed on 8 November 2002.⁸

Newspapers from the following countries were analyzed: Morocco, Algeria, Libya, Egypt, Sudan, Palestine, Lebanon, Syria, Jordan, Kuwait, Bahrain, Qatar, United Arab Emirates, Oman, Saudi Arabia, and Yemen (see Table 2). This collection of newspapers, searched on the Internet, produced 356 reports related to the United States in the two surveyed sections of the newspapers: the front page and the "international affairs" section. The number of reports obtained from this survey has to be taken with some caution. Access difficulties to certain newspapers' websites at the time of the survey, plus failures of some newspapers to update their website on a daily basis, suggest that these reports are at the bare minimum. (Iraqi newspapers were also surveyed but have been discussed earlier.)

It is not the purpose of this article to focus on quantitative analysis; that is, the purpose is not to determine which of the countries of the Arab world produced more reports on America than others. Frequency of reporting on American affairs does not help much in exploring the rhetoric used to challenge or support the United States' claims of mobilizing for a just war. However, this analysis can reveal variations among Arab states in their level of geopolitical toleration of American hegemony through

their approach to informing readers about American affairs, culturally, economically, and politically.

Nevertheless, Table 2 provides partial support for the notion that countries bordering Iraq or adjacent to Iraq were of great asset to the United States. It can be argued that these countries received more attention from the United States than did others during the period of war preparations. U.S. contact with these countries, including constant visits by diplomats and politicians, was intensified, placing them at the center of their local news. Hence, their respective newspapers were very keen to report on such American diplomacy. Five of the six countries that exceeded the average score of six articles per paper border with Iraq. Yemen, with a total of eleven, does not have a common border with Iraq but received great attention from the United States because at the time of the survey Yemen's government was cooperating with the United States in a military operation inside its own territory, portrayed as an episode in the "war on terrorism." Understandably, this event became major news for Yemeni newspapers.⁹

Although the United States sought worldwide legitimacy for the war in Iraq, the Arab world was a region of crucial importance. Yet U.S. relations with the states of the Arab world were diverse and constantly changing. As a hegemonic power, the United States had at its disposition a wide range of institutional instruments to engage governments in the Arab world in order to advance its interests in the region. As the United States prepared for war, contact with Arab countries was in-

Table 3. The bilateral relation of U.S. and Arab states: The view from the U.S.A.

	General relations	Iraq 1990-1	11 September 2001	Economic relations	Military relations
Morocco	'Longest unbroken treaty relationship in U.S. history.' 'Mutual respect and friendship.'	'First Arab country to condemn Iraq's invasion of Kuwait and sent troops to help defend Saudi Arabia.'	'First among Arab and Muslim states to denounce 911.'	Economic assistance: \$42.1 million U.S. aid since 1953 Peace Corps since 1963	U.S. Navy port visits American forces enter Moroccan Air space NASA, IBB transmitter
Algeria	'Growing relationship between the U.S. and Algeria.'		'Strongly supportive of the international war against terrorism.'	TIFA (2001) USNAEP MEPI	U.S. naval ships Joint naval exercises Staff exchanges
Tunisia	'The U.S. has very good relations with Tunisia, which date back more than 200 years.'	'Tunisia objected to U.S. intervention following Iraq's invasion of Kuwait.'		U.S. aid MEPI, USNAEP, TIFA (2002) American private assistance	Joint military exercises U.S.-Tunisia Joint Military Commission
Libya	'The U.S. government declared Libya a 'state sponsor of terrorism' on December 29, 1979.'		'Due to Libya's continuing support for terrorism, ...'	Iran and Libya Sanction Act (ILSA) (1996). ILSA renewed in 2001	
Egypt	'The U.S. and Egypt enjoy a strong and friendly relationship.'			U.S. aid \$25 billion (1975-2002)	U.S. military aid: \$1.3 bil/yr. Military exercises (Operation Bright Star) U.S. troops stationed U.S. 6 th fleet regular visits
Syria	'U.S.-Syrian relations, severed in 1967, were resumed in June 1974...' 'The U.S. continues to have serious differences with Syria.'	'Member of multinational coalition of forces in Gulf War.'			
Lebanon	'The U.S. seeks to maintain its traditionally close ties with Lebanon and to help preserve its independence, sovereignty, national unity, and territorial integrity'			Program of relief, rehabilitation and recovery (since 1975 \$400 million) Assistance to American University of Beirut, Lebanese American University, Lebanese American Community School, International College	Military Education and Training program resumed in 1993
Jordan	'Relations between the U.S. and Jordan have been close for four decades.'		'Jordan's opposition to terrorism.'	Economic assistance totaling more than \$2 billion (since 1952)	U.S. military assistance (materiel and training)

Table 3. (Contd.)

	Iraq 1990-1	11 September 2001	Economic relations	Military relations
	General relations			
Saudi Arabia	'SA's unique role in the Arab and Islamic worlds, its possession of the world's largest reserves of oil, and its strategic location make its friendship important to the U.S.' 'close consultations'	'Saudi cooperation in Gulf War.' September 11.' (15 of the suicide bombers were Saudi citizens). 'Important partner in the campaign against terrorism.'		U.S. military mission at Dhahran (1953) U.S. troops stationed
Yemen	'As a result of Yemen's actions in the Security Council following the Iraq invasion of Kuwait the U.S. drastically reduced its presence in Yemen, including canceling all military cooperation, non humanitarian assistance and the Peace Corps program'	'Currently Yemen is an important partner in global war on terrorism.'	U.S. dropped in 1991 to \$2.9 mio U.S. AID health MEPI funding	'Defense relations are improving' Recent resumption of International Military Education and Training Assistance \$1.9 mio foreign military financing in 2003
Oman	'U.S.-Omani relations were deepened in 1980'		Joint Commission for Eco and Technical Cooperation US eco assistance Peace Corps program	Access to Oman military facilities
UAE	'The U.S. has enjoyed friendly relations with the UAE since 1971.'	'The breadth, depth and quality of U.S.-U.A.E. relations increased dramatically as a result of the U.S.-led coalition's campaign'	Private commercial ties, especially in petroleum	
Qatar	'Bilateral relations are strong and expanded.' 'Ties between the U.S. and Qatar are excellent.'		Extensive economic ties, esp. in the hydrocarbons sector	
Bahrain	'The U.S. designated Bahrain a major Non-NATO ally in October 2001'	'Base for military operations in the Gulf.'	U.S.-Bahrain economic ties have grown steadily since 1932; oil. Free Trade Agreement due early 2004.	Base for U.S. naval activity since 1947. Defense Cooperation Agreement (Oct 1991). Access to military facilities.
Kuwait	'Strategic cooperation.'	'Important partner in the ongoing U.S.-led campaign against international terrorism.'		'U.S.-Kuwait strategic partnership intensified dramatically after Iraq's invasion in 1990.'

Source: Quotes and factual information from Country Information file, U.S. Department of State at <http://www.state.gov/p/nea> (last consulted December 2003). Postwar developments disregarded, such as troops withdrawn from Saudi Arabia (August 2003), training of 30,000 Iraqi police cadets in Jordan for the Coalition Provisional Authority in Iraq, and radical change of relations with Libya (December 2003).
MEPI = Middle East Partnership Initiative to support civil society. TIFA = Trade and investment Framework Agreement; USAID = U.S. Agency for International Development; USNAEP = U.S.-North African Economic Partnership.

tensified to secure public support and logistical cooperation. From the American perspective, support from countries that border on Iraq or those located in the immediate region was a war asset. But the support of other Arab states was also sought by the United States as desirable for politically isolating Iraq and portraying America's war as just. Moreover, support from all Arab states and their populations was necessary to convey the geopolitical belief that the prime modernity of the United State was universally desired and applicable.

Table 3 portrays the geographic variation in U.S. military presence and economic assistance across the Arab world. Until the dramatic change of geopolitical stance in January 2004, Libya–U.S. relations from the 1960s were defined by diplomatic and military conflict in the form of state-sponsored terrorism, retaliatory air-strikes, and diplomatic and economic sanctions. In marked contrast was the U.S.–Jordan relationship, with diplomatic relations and economic and military assistance appearing as the obvious manifestations. With the exception of Libya, the countries of North Africa had a tradition of relatively close ties with the United States. Most of the Gulf States also displayed a history of cooperative ties with the United States, but the relationship with Yemen became strained after the Gulf War of 1991. Saudi Arabia's position in the American geopolitical vision was more complex, and contested within American policymaking circles. Geopolitical cooperation beyond the restricted scope of diplomacy and military aid was evident, as was its geographic variation in the establishment of American universities, as well as social ties founded on immigration patterns.

The three spheres of hegemonic power and extraterritoriality are evident from Table 3. Political authority was manifest in diplomacy and military contacts, economic strength was translated into either integrative power (aid) or coercive power (sanctions), and cultural power was extended through education (Boulding 1990). But the differences across the columns of Table 3 also indicate that the hegemonic pursuit of extraterritoriality and a universal prime modernity was sometimes resisted (e.g., Libya, Syria, Yemen). Resistance may, ultimately, provoke war. Prevalent was the continued need of the hegemonic power to extend and maintain its extraterritorial reach. Especially at a time of approaching war, the hegemonic power needed to establish political and military support in the region of conflict. Moreover, Arab countries had to portray such contacts to their public in a way that strengthened the legitimacy of ruling elites through their use of the hegemonic power's extraterritorial demands. The media became a very important venue for portraying, on the one hand,

hegemonic influence as something that was desirable and, on the other hand, resistance to hegemony as a logical and successful act.

How was the impending war constructed in the Arab media? Looking at the issue from an Arab government perspective, although the general mood changed from day to day depending on how these governments perceived the current U.S. determination to go to war, they all, except Kuwait, told their populations that they generally did not support a war against Iraq, and sought in the time period studied to find a diplomatic and peaceful solution to accommodate the American demands on Iraq. Given this situation, the geopolitical reality reflected variations in the levels of tolerance to this war among Arab regimes. Underlying this toleration was the balance by which each state tried to compromise between its own national political interest, which sought to preserve the status quo of bilateral relations with the United States (and even improve it, as Qatar did), and the need to satisfy the call of Arab populations who, on the whole, opposed any hostile act by the United States against a fellow Arab or Moslem country.¹⁰

The remainder of this section depicts the view of the Kuwaiti newspapers. Kuwait's position was clear from the beginning. Kuwait described this war as one between the United States and Iraq, having nothing to do with its own state (*Al Seyassah* 2003a), and making the distinction between its own opposition to the Iraqi regime and its sympathy for the Iraqi people. Kuwait, having recovered its independence after an international U.S.-led intervention in 1991, was of all Arab states the most concerned by the new war on Saddam Hussein's regime. The invasion of Kuwait by Iraq and the subsequent war in 1991 resulted in expectations that the Kuwait authorities would be most supportive of American military efforts to destroy Saddam Hussein's regime. In mid-March 2003, at the time of the meeting in the Azores, the rhetoric of war in Kuwaiti newspapers¹¹ focused on the inevitability of armed conflict and the coming "moment of truth" for the Iraqi regime. In an article by Nasser Al-'Utaybi (16 March 2003), the Kuwaiti daily *Al-Qabas* cited Shaikh Sabah al-Ahmad, Deputy Prime Minister and Foreign Minister, as saying that the U.S. strike against Iraq was coming soon and only a miracle would stop it. Another Kuwaiti figure, Shaikh Dr. Mohammad al-Sabah, Minister of State for Foreign Affairs, was cited as saying that the military strike against Iraq was imminent: "it is now midnight minus one minute and counting. The Iraqi regime has one minute to prove it will cooperate with the inspectors and completely implement [all] decisions if it wishes to avoid the horrible consequences of war" (Al-'Utaybi 2003).

The underlying theme of these media reports is the immorality of the Iraqi regime in its lack of acquiescence to U.S. demands. By default, the argument is that the United States was waging a just war against an international pariah. In other words, the official Kuwaiti line followed an established theme of just-war theory: extraterritorial military reach is just if waged against an immoral regime. The contrast between American justness and Iraqi illegitimacy was strengthened by the *Al Seyassah* newspaper's publication of three reports pertaining to the upcoming war against Iraq (Al Hassan 2003; Al-Jarallah 2003; *Al-Seyassah* 2003b). One of these reports opened with the subtitle "What if we defeated America?!" The author of this piece, Ahmad Al-Jarallah (2003), launched a severe attack on the Arab satellite TV stations that were repeatedly broadcasting pictures of mass demonstrations against the war on Iraq. In doing so, the author claimed these satellite stations, backed by their respective governments, did not serve the interest of the Iraqi people. Nor would it change Western opinion, but would only obfuscate the other voices in the Arab masses who were not against the war. The author concluded, among other things, that this war was imminent and was not meant to be against the Iraqi people but against its regime, and that such demonstrations would not stop it.

Although the general view, both officially and expressed publicly in Kuwait newspapers, was that a U.S.-led war against Iraq resulting in regime change would be just and in the best interests of Kuwait, these newspapers also published reports inconsistent with the national viewpoint. For example, *Al-Qabas* carried a piece by Faleh Al-Fadily (2003) in which he cited the criticism by a Kuwaiti parliament member, Mubarak Al-Duwaila, of the Kuwaiti policy in handling the Iraqi file in the current crisis with the Americans; Al-Duwaila was quoted as saying Kuwaitis should not speak with an "American tongue" in the UN, as that would provoke Arabs and Moslems. Furthermore, he is credited with saying that "we should remember that Kuwaitis are an integral part of this Umma: though the Americans are here today, it does not mean they will stay in the region tomorrow, and Iraq and Iran, as neighbors, share destiny." Al-Duwaila also criticized Kuwait's acquiescence in hosting American troops at a time when Turkey rejected such a move. According to Al-Duwaila, basing American troops was a clear challenge to the Islamic and Arab Umma, and his dissent is based on the territoriality of the Arab people, and the immorality of the temporary and self-serving extraterritorial encroachment of U.S. hegemony. Despite this critique, the dominant view in Kuwait was to acknowledge both the extraterritorial reach of the United

States and its moral right to intervene against Hussein's regime.

It is clear from this overview that the Arab region was entangled in a web of economic and military cooperation with the United States, and Washington was able to exploit such bilateral relations to its own advantage. By and large, the United States dealt with many Arab countries through a strategy of carrot and stick. Those who did not follow and accept American hegemonic aspirations in the region were likely to be punished while other states were rewarded. The strategy of carrot and stick was evident across different manifestations of hegemony. In the economic sphere, both the promise of financial aid and the threat of sanctions were on the agenda. Similarly, in the political sphere, political cooperation, within the rubric of "democratization," and military alliances operated in tandem with diplomatic isolation and the use of military force. The specter of morality hovered over all these relationships, as the United States' economic and political agenda was couched within the rhetoric of "civilization" and "human dignity."

Arab Constructions of an Unjust War Against Iraq: Political Cartoons

Since most newspapers in the Arab world are under direct government surveillance and censorship, they represent mainly the regime view of the American interventions and tend to refrain from publicizing dissident opinions. The differential level of toleration of American policy is a carefully calibrated balance between a country's own national political interest and its geopolitical and geoeconomic ties to Washington on the one hand, and the pressure emanating from Arab political and civil society, both national and transnational, on the other. The Arab regimes have diverse (geo)political interests and agendas vis-à-vis the United States, but most of them are anxious to maintain good relations with the hegemonic power. On the other hand, public opinion was unanimous throughout the Arab world in opposing any aggression against a fellow Arab or Moslem country. This public opinion was doubtless also shaped by the mediascape. Because of limited press freedom in the Arab world, our second analysis focuses on political cartoons, which offer an outlet for critique of official views.

The U.S. hegemonic power and its plans to launch a "just war" against Iraq and its leadership were challenged by the rhetoric emanating from the Arab world, a rhetoric that filters through the local Arabic newspapers. In newspapers, political images express prevalent fears

about geopolitical threats and dangers; satirical cartoons question dominant power relations and knowledge; cartoonists even contest dominant representations of political affairs, such as Steve Bell's antigeopolitical eye in his depiction of the Falklands War (Dodds 1996). Cartoons are expected to be effective tools to challenge both the foreign policy of a regime and the hegemonic position of the United States.

This section shows the kind of arguments adopted by Arab cartoonists in their opposition to the 2003 U.S.-led war against Iraq. Political cartoons are probably the best medium by which Arab newspapers could express their protest and disseminate a viewpoint on this thorny and highly political issue. If, for reasons of censorship and political pressure, an editor or a reporter of a newspaper was unable to explicitly convey in writing an anti-U.S. message, the same sentiment could be effectively presented in a cartoon illustration. A cartoon presents a highly subjective opinion, but it can be understood as constituting certain elements of "truth" prevailing among the population. Indeed, a cartoon can enforce such a "truth" and become an effective tool of persuasion. A cartoon that has the element of amusing the reader can appeal to all readers and not only to the sophisticated ones. As such it spreads from mouth to ear and eventually reaches many people.

The Arabic newspapers surveyed provided rich geopolitical arguments that the United States was unjust and unfair in launching a war against a fellow Arab country. At the center of these geopolitical messages was the issue of immorality, especially the ability to resist the American war by portraying the United States as immoral and, therefore, unable to wage a war for just reasons and with just authority. From an interpretation of the following selection of cartoons, we are able to show that Arab commentators were adept in identifying the geopolitical foundations necessary for the United States to argue that its actions were "just," and then undermining the hegemonic power's actions by attacking these self-same moral foundations: the disrespect of Iraqi sovereignty and by extension that of other Arab states. Discussion of the cartoons is organized around five themes: imperialist intent; the arrogance of power; realist power politics; double standards toward the UN and international law; and support of Israel portrayed as immoral.

Imperialist intent. In the Arab newspapers we surveyed on twenty-one days during November 2002, the U.S. call for war against Iraq was seen as serving only America's grand policy in the region, and thus was depicted as an immoral use of military might for selfish



Figure 1. Cartoon from the Egyptian newspaper *Al-Osboa* (6 November 2002).

reasons rather than an act of preemptive self-defense or humanitarian intervention. Figure 1 is a cartoon that appeared in the Egyptian newspaper *Al-Osboa* (6 November 2002) and is a drawing of an Uncle Sam-like character standing before a list of Arabic countries ordered from top-down: Palestine, Iraq, Syria, Saudi Arabia, and Egypt. Uncle Sam has already drawn lines through Palestine and Iraq and is ready to continue marking off the remaining Arabic nations, as though it were a to do list. This image depicts the Arab feeling that the United States is against all Arabs (and not only Iraq) and will eventually attempt to dominate them all. Significantly, Uncle Sam is seen writing with his left hand and has what looks like a cigarette in his right hand, but this adds to a careless ever-consuming (and over-consuming) image of the United States and how they treat the Arabs in a disrespectful way.¹² The countries at the top of the list are more readily identified as current "enemies" of the United States, but those nearer the bottom are generally understood to be on friendlier terms. The geopolitical message of the list is twofold: The United States is not to be trusted as it may turn on its friends, and the unity of the Arab people is the only meaningful geopolitical response. The cartoon speaks to the Arab population to be unified not only against the hegemonic power but also Arab leaders who make alliances with the hegemonic power. The cartoon identifies the universal pretensions of prime modernity, but portrays their manifestation as coercive or imperialistic rather than integrative or welcomed. The cartoon identifies American hegemony as an act of immoral territorial subjugation, and so undermines the geopolitical assumption that the United States is a "just authority," a prerequisite for waging a just war.

Figure 2 represents the same Uncle Sam-like character, this time in the Jordanian *Ad-Dustour* daily (15



Figure 2. Cartoon from the Jordanian *Ad-Dustour* daily (15 November 2002).

November 2002). The sinister figure is peering downward through a telescope at what is labeled "Middle East" in Arabic. The telescope is made up of a series of three progressively bigger oil drums, the largest of which is labeled "Oil" in English. Significantly, this cartoon was published in Jordan, not an oil-producing country. Though, as hegemonic power, the United States has many interests in the Middle East, the cartoon makes an argument that oil is the primary geopolitical interest. This cartoon disseminates a sense of Arab unity in response to the U.S. hegemonic project, resting on an ingredient of popular Arab politics: Oil is seen as belonging to the Arab nation as a whole rather than a single country. Oil, in the rhetoric of Arab nationalism, is a means to construct a regional political identity that could challenge the universal influence of the hegemonic power. The cartoon depicts the United States as motivated only by material greed, destroying countries for access to oil, rather than using its hegemonic power to construct political projects. This "Uncle Sam" sees the Middle East only through the lens of oil. U.S. military intervention is portrayed as a policy of using military might to gain access to strategic resources; an immoral geopolitical act *par excellence*.

The arrogance of power. The next cartoon argues that the United States not only does not care about the interests of the Arab people but does not even bother to listen to their appeals and requests. Figure 3 shows a cartoon that appeared in the Jordanian *Assabeel* weekly on 3 November 2002. An Arab man is shown laying down papers with Arabic writing on them to form a road

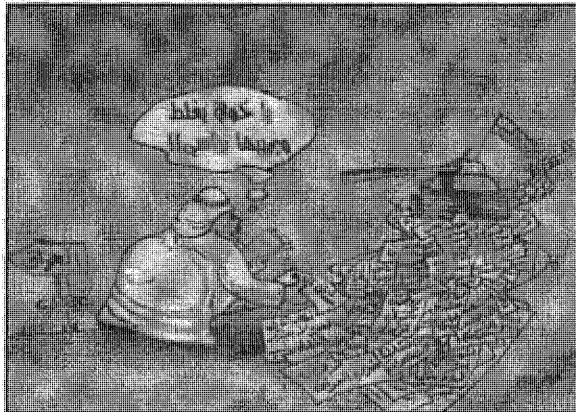


Figure 3. Cartoon from the Jordanian *Assabeel* weekly (3 November 2002).

or pathway along which an American tank is proceeding. A sign points to Iraq. Written on the papers are statements such as, "We oppose attacking Iraq," "For God's sake," "We will not participate in attacks on Iraq," and "We will not stand with hands tied." The antihegemonic message of the cartoon is clear: contrary to U.S. statements that military action against Iraq was in the name of the Arab people and their human rights, this appeal was not universal. In other words, the integrative appeals to the Arab people were resisted.

The cartoon tells a story of the necessity and brutality of U.S. coercive power, literally "rolling over" or crushing dissent. The tank has an American flag on it and a long gun barrel pointed at the Arab man's head. The man is thinking "I'm afraid that he may mistakenly hit me on the head." The reader wonders who commits the mistake; the tank seems to be guided by the path of Arabic protests leading straight to Iraq. Moreover, and in contrast to the hegemonic message of human rights, the message is of disrespect for human life, especially Arab. In this image, American hegemony is not spread via the willing emulation of the prime modernity. Instead, American hegemony is identified as coercive military might. The spread of the prime modernity is against Arab wishes, and its adoption is an unwilling one, brought about through the exercise of violence. The coercive means of disseminating the hegemonic message is shown to contradict its key ingredient of human rights.

A popular saying in the Arab world is "*Mosadurat al Qarar el 'Arabi*" ("The confiscation of the Arab decision"). Figure 4, published in the United Arab Emirates *Itihad* (21 November 2002) propagates this feeling. The cartoon shows two debaters, one Arab and the other our old friend "Uncle Sam." Both men are speaking, the

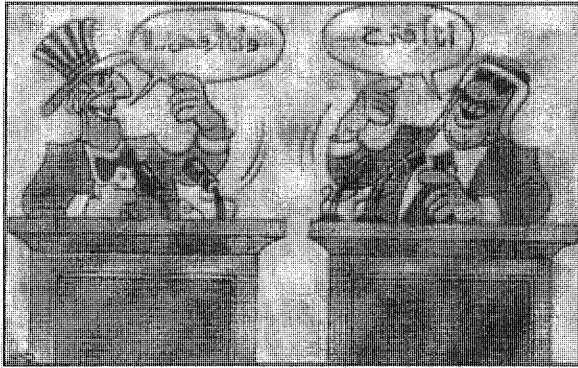


Figure 4. Cartoon from the UAE *Ittihad* daily (21 November 2002).

Arab saying, "I propose," and the American saying, "I reject." The cartoon shows Arabic frustration with the unfeeling and selfish U.S. attitudes toward the Arabs. Again, the message of the cartoon is that U.S. prime modernity is not wanted nor is it applicable to the Arab world; nevertheless the United States intends to diffuse it against the wishes of Arab nations. In this sense, the United States is acting immorally by transgressing the sovereignty of Arab nation-states and the ability of their populations to make their own political decisions.

The Arab countries and their leaders were not only having their appeals disrespected by the United States, but also found themselves at a low point of power, unable to change the course of events as the American hegemony prevailed. Furthermore, a portrayal of the impotency of Arab leaders evident in Figure 4 is continued in Figure 5, a cartoon by Ommaya Jeha that appeared in the Palestinian daily *Al hayat Al Jadeeda* (12 November



Figure 5. Cartoon from the Palestinian daily *Al hayat Al Jadeeda* (12 November 2002).

2002). The right panel shows the face of an Arab leader in front of his people, looking ominous and slyly powerful. The left panel shows the same man in front of America. The once proud moustache is drooping down and the right hand is raised in submission. The image is one of the Arab people's frustrations with their own leaders.

The universal appeal of the hegemonic power is exposed, in the cartoons shown in Figures 3, 4, and 5, as a political tactic requiring the assistance of weak and complicit Arab leaders, rather than an integrative project adopted by the general public. This message reinforces the argument of Arab commentators that the United States uses its hegemonic power in an immoral manner by running roughshod over the wishes of Arab citizens; self-determination within sovereign spaces, the moral foundation of the international state system, is seen to be violated by America.

Realist power politics. The cartoons in the Arab newspapers surveyed depict the United States as behaving in a manner akin to realist doctrines, often opting for the use of excessive power to achieve its political goal. Hegemonic power in the modern world system operates in a system of sovereign states, a system that the United States advocates in its own rhetoric. However, the means and benefits of hegemonic rule require the constant extraterritorial violation of territorial sovereignty. In other words, the hegemonic power simultaneously maintains a moral political geography of sovereign states, but must violate this very morality in order to manage the system. The Arab cartoons we analyzed were successful in not only identifying this paradox but using the inconsistency to generate political opposition against the "immorality" of hegemonic practice.

Figure 6 illustrates this notion. This cartoon, from the Qatari daily *Al-Watan* (12 November 2002), shows a

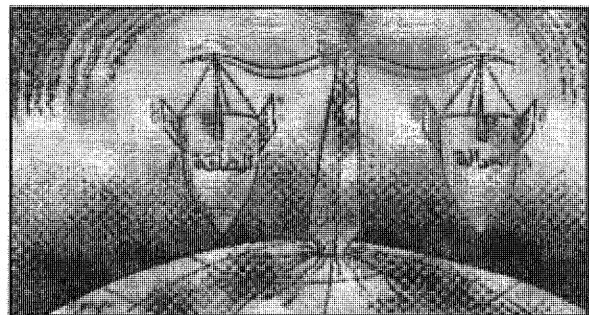


Figure 6. Cartoon from the Qatari daily *Al-Watan* (12 November 2002).

sword labeled "U.S.A." sticking in the surface of the world. The yolk of a scale is mounted to the hilt of the sword and the weights balancing the scale are large bombs. "Absolute Justice" is written in Arabic on the bombs. The cartoon depicts the perception (and reality) of the military power the United States wields over the rest of the world. Both justice and dignity were key themes in President George W. Bush's geopolitical code (National Security Strategy 2002). By using themes of injustice and indiscriminate violence, this cartoon represents a strong challenge to the rhetorical foundations of the current foreign policy of the United States. The cartoon suggests that the United States is willing to use its coercive power across the whole globe, and not just in the Middle East, and so illuminates the hypocrisy of resorting to coercive power to enforce what are portrayed as integrative political projects.

The language of power was reaffirmed in sketched representations of the president of the United States. The Bahraini daily *Al-Ayam* (21 November 2002) attached a sketch of President Bush to a report on Afghanistan (Figure 7). The picture is a caricature showing him as a bare-chested cowboy, posed to show a powerfully muscled upper body, complete with flexed biceps.



Figure 7. Cartoon from the Bahraini daily *Al-Ayam* (21 November 2002).

He is wearing a cowboy hat and cowboy boots with spurs, and has a large revolver holstered on each hip. On his face is a serious "Don't mess with me" look. The image is a cross between a wild-west movie cowboy and a modern-day body builder, not that fearful by itself but effective in catalyzing political resistance when we think that this person is the leader of the most powerful nation on Earth. In other words, the cartoon argues that the hegemonic power's authority rests on violence; it is an immoral coercive power and not an integrative power resting on shared morals (Boulding 1990).

Arab commentators argued that even when the President is extending a "Happy New Year" he does not miss the opportunity to show dominance and the language of threat. The cartoon in Figure 8 appeared in the Moroccan newspaper *Attajdid* (18 November 2002), showing a simple rocket poised for take-off. Arabic text next to it says, "Happy New Year, on the imminent air attack on Iraq," and the message is signed W. (the letter W in English) Bush (in Arabic). The cartoon shows the resignation of the Arab world to the probability of an American attack and their disdain for Mr. Bush. Significantly, the scale of the imminent military action is identified by the cartoon as the sovereign nation-state of Iraq, and not Saddam Hussein or his regime. In combination with the "Happy New Year" theme the scale of the attack suggests that the hegemonic power somehow "enjoys" the widespread use of coercive power. The use of U.S. military might is portrayed as an unjust violation



Figure 8. Cartoon from the Moroccan newspaper *Attajdid* (18 November 2002).

of territorial sovereignty, rather than a moral mission to defend universal definitions of “civilization” and “human dignity.”

Double standards toward the UN and international law. Cartoons and commentary in the Arabic newspapers surveyed portrayed the United States behaving as if it were above arguably the world’s most accepted organization, the UN. Moreover, it is important to remember that the UN was a creation of the United States at the peak of its hegemonic power. Hence, portrayals of the United States ignoring the wishes of the UN are commentaries not only on U.S. immorality but also its political desperation. We found in Arabic newspapers a fairly large number of cartoons that portrayed the U.S. “treatment” of the UN and the team of WMD inspectors who were dispatched to Iraq. The cartoons shared the notion that the United States was not interested in the “truth” of Iraq’s military capabilities and intentions, and was, hence, acting immorally.

Figure 9 is a cartoon that appeared in the Jordanian daily *Alarab-Alyawm* (9 November 2002). The cartoon pictures President George W. Bush, again dressed as a cowboy, winding a key in the back of a figure labeled “UN” that is wearing a work jacket and cap with a pouch on the side holding what looks like some type of reporting device with an antenna (probably a weapons inspector). The image identifies the United States as a dominant controlling power over the inspection team. Another cartoon with a similar message appeared in the



Figure 9. Cartoon from the Jordanian daily *Alarab-Alyawm* (9 November 2002).



Figure 10. Cartoon from the Palestinian daily *Al-Quds* (21 November 2002).

Palestinian daily *Al-Quds* (21 November 2002). This cartoon (Figure 10) shows the UN inspection team (four helmeted figures) along with their van (the helmets and the van are all labeled “UN”). Arabic writing in the background says, “Search for weapons of mass destruction.” The men seem to be serious in their attitude. Inset in the cartoon, however, is a caricature of George W. Bush. The reader sees his thoughts that vow to smash Iraq—the message being that the inspectors will have little effect on the will of Mr. Bush.

The Jordanian daily *Alarab-Alyawm* (21 November 2002) published another cartoon in which the American president is seen as impinging on the work of the inspectors of WMD sent to Iraq. This image (Figure 11) is of a large bomb suspended by a rope. Above the bomb,

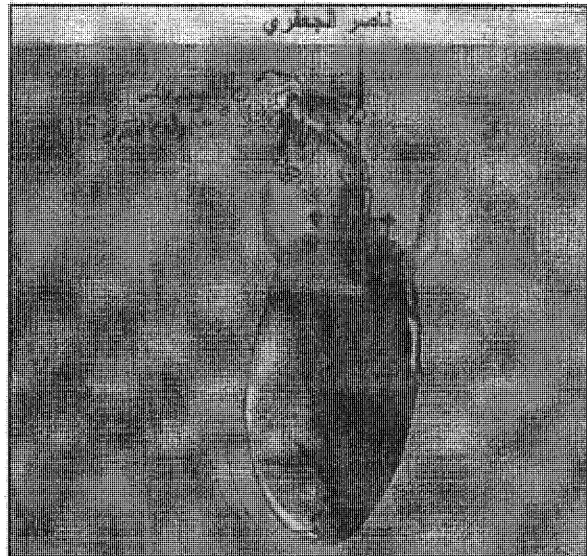


Figure 11. Cartoon from the Jordanian daily *Alarab-Alyawm* (21 November 2002).

holding the rope in one hand and a pair of scissors in the other, is a small image of President Bush. Bush is imploring Mr. Blix to hurry up and finish his work. The cartoon implies that President Bush wants to cut this rope holding the bomb and is getting impatient with the inspectors. The Syrian daily *Teshreen* (20 November 2002) published a cartoon with a similar message, but with a more elaborative structure (Figure 12). In this cartoon a gaudily dressed Uncle Sam-type figure is holding handguns with hammers drawn on two other figures. On his left, dressed in what looks like a lab coat, searching through a garbage can, is the UN weapons inspector. On his right, in Arab headdress, long moustache, and necktie, but barefoot and in ragged clothing, is a man representing the Iraqi people. Uncle Sam is telling the UN inspector to immediately find an excuse for military action. The message is of an impatient and demanding United States that cares as little for the UN representatives as it does for the poor people of Iraq.

In these cartoons, the United Nations, the contemporary guardian of just war philosophy and political territorial sovereignty, is widely seen as a respected but toothless institution. In other words, it is the protector of the widely held moral assumptions underlying the world political map: portraying the United States as trampling over the authority of the UN is to paint the hegemonic power as an immoral political force. It is the UN that is portrayed as the purveyor of integrative power, based on territorial sovereignty and due process, whereas coercive power, with no respect for the sovereignty of nation-states, is firmly attached to the actions of the United States.



Figure 12. Cartoon from the Syrian daily *Teshreen* (20 November 2002).

Support of Israel portrayed as immoral. Probably more than any other justification in the Arab newspapers, American support for Israel, and especially the support of President Bush to the government of Prime Minister Sharon, makes the United States an immoral state, disclosing earlier misuse of the United States' hegemonic power. If the support for Israel and the veto of almost any UN resolution that condemns Israel's atrocity against the Palestinians was not enough, the United States is portrayed as reaching the point where it supports the Israeli killing of Palestinians under the guise of the "war against terror." Such actions make it very difficult for even the friendliest of Arab leaders and people not to see the United States except with a double standard of policy and vision: terror is identified as actions against Israeli and American power and not the reciprocal action of those states. The following cartoons provide only a narrow snapshot into what this U.S.-Israel bond means to the people of the Arab world.

Figure 13 is a cartoon that was published in the Qatari daily *Al-Watan* (6 November 2002). This is a picture of a dove with an olive branch in its mouth, signifying peace, and labeled "UN." The wings of the dove of peace are held down by the heads of Mr. Bush and Mr. Sharon. This cartoon shows the frustration of the Arab people with UN peace efforts and also with the reasons for their failure. It also signifies the partnership between two warring states that are seen as disrupting UN efforts to make peace in this world. In these cartoons, it is the UN, as the voice of the world, that is portrayed as the defender of "civilization," with the United States acting immorally against its wishes.

The U.S. president's appreciation of Israel's Prime Minister, praising him as a man of peace, was illustrated in a cartoon that appeared in the Yemeni daily *Al-Sahwa* (18 November 2002). This cartoon (Figure 14) is an image of Uncle Sam reading from a document entitled "The American Vision." Here, Uncle Sam has an un-



Figure 13. Cartoon from the Qatari daily *Al-Watan* (6 November 2002).



Figure 14. Cartoon from the Yemeni daily *Al-Sahwa* (18 November 2002).

sual top hat with a very large nose and facial features and bowtie that resemble pictures of the “Mad Hatter” from Lewis Carroll’s *Alice in Wonderland*. He reads from the document through lenses that are cut in the shape of the Star of David, and proclaims, “Sharon is a hero of peace, Jerusalem is Israel’s eternal capital, the [Palestinian] resistance are terrorists, and, and, and . . .” The use of *Alice in Wonderland* displaces the policies of the United States from the real world of national interests and the circumstances of people into a world of fantasy. The integrative power of the United States is seen to be a fiction, unaware of the actualities of human rights and injustice in relation to the Palestinians. In other words, any integrative and universal message of human rights is portrayed as inapplicable, but belongs instead to a dysfunctional world of misperceptions and falsehood. Rather than a universal prime modernity based on rational foundations, the United States is portrayed as the epitome of madness; a detachment from civilization rather than the author of the modern definition of “civilization.”

The U.S. policy of a double standard toward UN resolutions was depicted in a cartoon that appeared in the Palestinian daily *Al-Quds* (9 November 2002). Here (Figure 15) one sees a cowboy portrayal of President Bush standing on a box among many other boxes in a warehouse-like place. He is holding up a UN resolution against Iraq, and strewn around him among the boxes



Figure 15. Cartoon from the Palestinian daily *Al-Quds* (9 November 2002).

are various UN resolutions against Israel: 242, 338, 191, and so forth. President Bush is saying that “this resolution [i.e., 1441] must be implemented now, immediately, without dragging, without excuses, without delay, without discussion . . .” The cartoon is saying that the UN should forget about all the resolutions against Israel, and instead drop everything and implement this one against Iraq. In the view of this cartoon, the United States is violating one of its own institutional foundations of hegemonic power to support what is seen as the illegal actions of Israel. In this instance, the immoral acts of the United States are twofold: supporting Israel and misusing the authority of the UN.

The last cartoon we include in this article shows commonalities between the United States and Israel when it comes to the treatment of Arab citizens residing in the two countries. Figure 16 is a cartoon that was published in the Palestinian daily *Al-Quds* (9 November 2002). This cartoon compares the deportation of Arab-Americans, post 11 September 2001, with the deportation of Palestinians by the Jews in 1948 in Israel. The

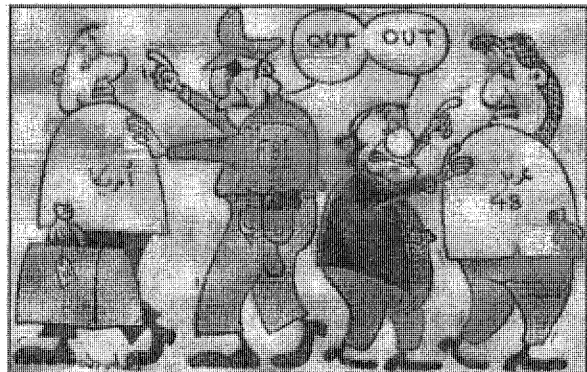


Figure 16. Cartoon from the Palestinian daily *Al-Quds* (9 November 2002).

writing on the T-shirt of one of the Arabs says "Arab 48," a common term in the Arab world referring to those Palestinians who remained in Palestine post-1948 and became Israeli citizens. The cartoon paints a picture of the United States and Israel as comparable states in terms of their willingness and ability to disregard human rights and deport their own citizens. There are four figures in the cartoon. The two in the middle are both pushing out the ones outside of them, physically, with hands on, and both are saying "OUT!" On the left is the Arab-American being kicked out by the American policeman with "FBI" emblazoned on his shoulder. This figure is dressed more like a Gestapo officer than a Federal Agent. The Arab-American seems confused but is moving along. On the right is the Palestinian, wearing a headdress that looks like the one worn by President Arafat. The Palestinian is being thrown out by a much smaller Jewish man, complete with yarmulke. The underlying allusion to a comparison between the policies of Israel/United States today and Nazi Germany, widely acknowledged as the most immoral regime of the twentieth century, is evident.

The prime modernity of the United States conveys itself as the moral basis of hegemonic power by championing human dignity and human rights, within a world political map of territorially sovereign states. In opposition to these claims, the cartoons referring to the role of Israel, Palestine, and the United States in the occupation presented the hegemonic power as abetting the violation of human rights and not respecting the drive toward Palestinian national self-determination. In the context of contemporary geopolitics, perhaps nothing could be more effective in portraying the United States as an immoral force in the Arab world, rather than a hegemonic power whose legitimacy is founded on a generally accepted moral agenda. The important consequence is that an immoral power is not a "just authority" and, therefore, is unable to wage a just war (Walzer 2000).

Conclusion

Media representation is part of geopolitical conflicts, not an external matter of objective reporting. The hypermedia environment in which the drama leading up to this war occurred, including instant information over the Internet and online newspapers, was unique in modern history and geopolitics: it helped fuel popular and governmental opposition on an unprecedented global basis, providing information on protests virtually as they occurred. The media thus helped to bolster morale within Iraq and significantly strengthened the

"rhetorical position" of the Iraqi leadership both internally and abroad. As the Iraqi vice-president said on CNN on the evening of 17 March 2003, the "whole world" was opposed to Bush and his war in light of the rhetoric disseminated by those opposed to military invasion and the number of media reports contradicting the statements made by the U.S. government.

Never in modern times has there been such a global and well-organized welling up of opposition to a war even before the first shots were fired. Such global dissent is an example of the "network society" (Castells 1996) and the impact of cyberspace and the mediascape in shaping geopolitical resistance. In the infoscapes and mediascapes of postmodern informational capitalism, groups of people began to join global webs, and the quickening pace of flows upended traditional divisions between local, national, and global. This created what Luke (1995) termed a scalar dynamic of "neo-world orders": a new connectivity of resistance and outrage manifested in a "glocalized" tidal wave of opposition that was indeed a kind of "glocal space." The reality that "America's leadership had lost touch with reality," as Krugman (2003) put it, was in part a function of geopolitics being played out in cyberspace, at multiple levels, in a wired age; an extraordinary challenge to hegemonic power and its manipulations. Debates within the infoscapes were grounded, however, within the usage of a territorial political geography to make moral claims.

The cartoons and newspaper stories we analyzed were published in newspapers with ties to a particular state or territory, but simultaneously accessible throughout the world in electronic format. In other words, they were popular geopolitical representations, disseminated through the global electronic mediascape and, simultaneously, newspapers centered on a territorially defined readership—representations that negotiated the watching eye of authority, both government and civil society, while reaching a mass audience. The commentaries were, therefore, able to make a global appeal but the message was grounded in the political geography of sovereignty. The target of the cartoons was clear: the United States. Equally clear was the cartoonists' understanding of the political geography of hegemony: the extraterritorial reach into sovereign spaces. In addition, the rhetoric of the hegemonic power, the calls of "justice" and "moral right," is pervasive in the visual commentaries. Perhaps most telling is the ease with which the cartoonists were able to ridicule the moral foundations of American hegemonic prime modernity. The case for the United States being an unjust authority waging an unjust war was as easy to make as it was passionately felt. The impression taken from the cartoons is that resistance to an unjust war was not mo-

bilized by these cartoons; there was no need as such opposition already existed.

In the Iraqi media in March 2003 the war was expectedly constructed as unjust military aggression requiring a just responsive war against U.S. imperialism and infringement of Iraqi sovereignty. In the rest of the Arab world, the media delivered mixed signals, echoing the reservations of but lack of opposition by most Arab regimes. Reportage of the intercourse between Arab and American political elites reflected the acknowledged hegemony of American policies, and the need to define national policies within the context of American demands and interests. Media reports friendly to Arab elites helped justify national policies constrained by U.S. extraterritoriality. On the other hand, political cartoonists were more adept at offering a critical interpretation of hegemonic reach into sovereign Arab territory. Arab political cartoonists did not need to resort to regionally or culturally specific discourses. Instead, the universal political ideal of territorial sovereignty was the basis for their argument.

The interpretation of the cartoons and reportage allows some reflections upon our theoretical understandings of popular geopolitics and the process of hegemony. First, a connection between Ó Tuathail and Dalby's (1998) "high" and "low" sites of geopolitical production is evident. Newspapers targeting a mass audience were critically engaging the ideas of geopolitical mission, morality, and universal human rights generated by elites. Second, the common criticism of the geopolitical morals of the United States was found across different states; this may be a function of the "network society" (Castells 1996) and its mediascapes of resistance (Luke 1995) diffusing common language, or a commonality in the experience of U.S. hegemony, or, most likely, a combination of the two. Thus, the rhetoric within Arab newspapers is a good illustration of how political agency is framed by geopolitical contexts or structures.

A second set of insights relates to the process of hegemony, or at least the particular case of the United States. First, the geography of flows and networks at this particular moment of hegemony facilitates conduits of resistance that transcend sovereign states (Agnew 1993; P. Taylor 1993). Second, the legitimization of the hegemonic power's military actions using the language of morality and just wars is a significant component of its integrative power (Boulding 1990). Third, and this is the conundrum the hegemonic power must negotiate, the universal hegemonic morality must be grounded within the world political map of sovereign nation-states (Flint 2004b). In other words, the hegemonic power must define its universal morality in an extraterritorial fashion

while transgressing the common-sense principles of the world political map, as it is commonly understood.

Without a geopolitical crystal ball no predictions about the trajectory of American power can be made. But the hegemonic conundrum outlined above suggests that the integrative power of the United States is in decline: popular Arab sentiment regards the world political map as a defense against the imposition of U.S. power rather than the arena for the construction of society modeled on American modernity (P. Taylor 1999). The Arab media we surveyed viewed the world political map of nation-states not as a field for the diffusion of U.S. prime modernity, but as an arena for resistance. Though the hegemonic power still possesses unquestionable military might, this article provides evidence that its integrative capacity is weak across the Arab world. Without integrative power, coercive capabilities become too costly to maintain and a decline in power follows (Boulding 1990).

The impotency of Arab people and their leaders prior to the war has important implications in the aftermath of the invasion of Iraq, the occupation. The embarrassing absence of WMD, acknowledged disconnects between al Qaeda and Hussein's Baa'thist party, and the reactionary show of a return to militarized violence in the face of armed resistance have maintained and reinforced popular Arab sentiment that the United States' military expedition in Iraq was an unjust, selfish, and materially motivated geopolitical enterprise. In other words, the themes of the cartoons are, perhaps, more important in the aftermath of the invasion. The success of the U.S.'s project in Iraq depends on integrative power, but the cartoons provide ample evidence that the Arab people were not compliant with this strategy. The result is that coercive power of invasion begets the coercive power of occupation. What has changed is the strategic situation: the practice of unjust war has placed American soldiers in situations where they are vulnerable to attack. It is no longer a war of drones, stealth bombers, and cruise missiles, but rather of rocket launchers, mortars, and improvised "roadside devices." Perhaps the saddest irony for all involved, those suffering loss in Iraq and the United States, is that the continued military presence is necessary to maintain the lie that the war was "just." A withdrawal now would be an admission of geopolitical adventure. The nightmare of occupation is necessary in order to maintain the dream that "democracy" and "development," the twin pillars of prime modernity, are being delivered. People are continuing to die in an unjust war, so that the fiction of a U.S. hegemony curing the ailments of a grateful Arab world can still be written as part of a universal story.

Acknowledgments

We thank Audrey Kobayashi for her encouragement and careful editing. We are also grateful for the comments provided by the reviewers and their role in improving our initial submission. Finally, thank you to the editors of the Arab newspapers who granted permission for the reproduction of the cartoons.

Notes

1. Our task is not to provide an evaluation of whether the war on Iraq met criteria that would classify it as just or unjust. For such a debate see Carter (2003) and Kaplan and Kristol (2003), and special issues of *Antipode* (vol. 35, no. 5, 2003) and *Arab World Geographer* (vol. 6, no. 1, 2003). For academic considerations regarding the United States' actions in the post-9/11 world see Ó Tuathail (2003) and a special issue of the journal *Geopolitics* (vol. 8, no. 3, 2003). Instead, our focus is the rhetoric used by policymakers and opinion makers in their efforts to persuade the public of the justness or unjustness of the war.
2. For a full discussion of just war theory see Walzer's (2000) seminal text, plus its use by Johnson (1984), Orend (2000a, 2000b), and Temes (2003).
3. See Orend (2000c) for the same assumptions in Kantian just war theory.
4. The survey includes four Iraqi newspapers online (*Al-Thawra*, *Al-Iraq*, *Al-Jumhuriya*, and *Alef-ba* [a weekly magazine]) and the Iraqi News Agency report of 15 March 2003.
5. The *Al-Jazeera* website (www.aljazeera.net) poll survey of 16 March 2003 (2:42 p.m. EST), involving 41,571 respondents, provides instructive results on how the Arabic-speaking readers of their website viewed this summit: 68.9 percent saw it as council for declaration of war; 25.8 percent as an attempt for finding an alternative to war; 3.2 percent as retreat from the war; with 2.2 percent undecided. Indeed the bombing started 19 March 2003. To relate the protests and bombing to just war theory, see Johnson (1984), Orend (2000a, 2000b, 2000c), Rengger (2002), and Temes (2003).
6. The four Iraqi newspapers surveyed were the most popular Iraqi newspapers at the time and all were placed under the surveillance of the government. *Al-Jumhuriya* (English translation: Republic) was the official Iraqi government newspaper and speaks on its behalf; *Al-Thawra* (English translation: Revolution) was the official newspaper of the Al Ba'ith party; *Al-Iraq* (after the name of the country) belonged to a Kurdish group loyal to Saddam Hussein's government; and *Alef-ba* (after the first two letters in the Arabic language alphabet) weekly was issued by the Iraqi Ministry of Information. There were two other popular Iraqi newspapers at the time but they were not included in our survey: *Al-Qadisiyya* was the Iraqi Army newspaper and *Babel* was managed by Uday Saddam Hussein, the oldest son of Iraqi President Saddam Hussein.
7. In our survey, websites of newspapers were consulted every third day beginning on 3 November for newspapers in Arabic, on 4 November for newspapers in French, and on 5 November for newspapers in English.

8. On 8 November, Syria, holding a temporary seat in the Council, surprised many by voting in favor of a UN Security Council resolution and making the UN vote unanimous.
9. Similarly, though Sudan does not have a border with Iraq, the country scores above average, probably a reflection of U.S. involvement in the conclusion of the Sudanese civil war.
10. The two characteristics are often conflated in Arabo-centric perceptions of the Moslem world.
11. Three major Kuwait newspaper websites (*Al-Qabas*, *Al-Watan*, and *Al Seyassa*) were surveyed on 16 March 2003.
12. In the Qura'an, references imply that the left-hand is associated with evil, and in Arab culture use of the left-hand is often related to carelessness.

References

- Agnew, John. 1993. The United States and American hegemony. In *Political geography of the twentieth century: A global analysis*, ed. Peter J. Taylor, 207–38. London: Belhaven Press.
- . 2003a. American hegemony into American empire? Lessons from the invasion of Iraq. *Antipode* 35 (5): 871–85.
- . 2003b. Commentary III: Learning from the war on Iraq. *The Arab World Geographer* 6 (1): 58–60.
- Agnew, John A., and Joanne P. Sharpe. 2002. America, frontier nation: From abstract space to worldly place. In *American space/American place: Geographies of the contemporary United States*, ed. J. A. Agnew and J. M. Smith, 79–107. New York: Routledge.
- Alef-ba*. 2003. With resolve we will win the war. 12 March.
- Al-Fadily, Faleh. 2003. Al-Duwaila criticizes Kuwaiti policy. *Al-Qabas*, 15 March.
- Al-Hassan, Abu. 2003. We are living in the last hours of diplomacy and those who oppose the war are trying to protect their interests. *Al-Seyassah*, 16 March.
- Al-Iraq*. 2003a. The Angolan president rejects the use of force against Iraq and suggests dialogue. 15 March.
- . 2003b. Foreign minister: Iraqis decide to fight in defense of their homeland against aggression. 15 March.
- . 2003c. Mosque sermons in Baghdad and other provinces plead to God to protect Iraq. 15 March.
- . 2003d. President of the religious committee prays to God to support Iraq and its leaders. 15 March.
- Al-Jarallah. 2003. Opening: What if we defeated America? *Al-Seyassah*, 16 March.
- Al-Jumhuriya*. 2003a. The British support French opposition to American aggression. 15 March.
- . 2003b. Continual resignations of American officials from their posts. 15 March.
- . 2003c. Peace delegates opposing American aggression renew solidarity with Iraq. 15 March.
- Al-Seyassah*. 2003a. Closure of northern region and the deployment of security forces in sensitive locations. 11 March.
- . 2003b. The Iraqi army builds "burning belt" around Baghdad to confront American tanks. 16 March.
- Al-Thawra*. 2003a. Australian protests are escalating against Howard. 14–15 March.
- . 2003b. European trade unions went on strike today in protest against American threats. 14–15 March.
- . 2003c. Moslem Ulama call for Jihad against American threat. 14–15 March.

- Al-'Utaybi, Nasser. 2003. Mohammad Al-Sabah talks to Al-Qabas. *Al-Qabas*, 16 March.
- Anderson, James. 2003. American hegemony after 11 September: Allies, rivals and contradictions. *Geopolitics* 8 (3): 35–60.
- Arrighi, Giovanni. 1994. *The long twentieth century: Money, power, and the origins of our times*. New York: Verso.
- Bacevich, Andrew J. 2002. *American empire: The realities and consequences of U.S. diplomacy*. Cambridge, MA: Harvard University Press.
- Boulding, Kenneth. 1990. *Three faces of power*. Newbury Park, CA: Sage.
- Carter, Jimmy. 2003. Just war, or just a war? *New York Times*, 9 March.
- Castells, Manuel. 1996. *The rise of the network society*. Oxford, U.K.: Blackwell.
- Dalby, Simon. 2003a. Calling 911: Geopolitics, security and America's new war. *Geopolitics* 8 (3): 61–86.
- . 2003b. Geopolitics, the Bush doctrine, and war on Iraq. *The Arab World Geographer* 6 (1): 7–18.
- Dijkink, Gertjan. 2003. World police, unilateralism, and the future of a country. *The Arab World Geographer* 6 (1): 19–23.
- Dodds, Klaus. 1996. The 1982 Falklands War and a critical geopolitical eye: Steve Bell and the If... cartoons. *Political Geography* 15 (6-7): 571–92.
- . 2000. *Geopolitics in a changing world*. Harlow, Essex: Prentice Hall (Pearson Education).
- Elshtain, Jean Bethke. 2003. *Just war against terror: The burden of American power in a violent world*. New York: Basic Books.
- Flint, Colin. 2003. Geographies of inclusion/exclusion. In *The geographical dimensions of terrorism*, ed. S. L. Cutter, D. B. Richardson, and T. J. Wilbanks, 53–58. New York: Routledge.
- . ed. 2004a. *The geography of war and peace: From deathcamps to diplomats*. Oxford, U.K.: Oxford University Press.
- . 2004b. The “war on terrorism” and the “hegemonic dilemma”: Extraterritoriality, reterritorialization, and the implications for globalization. In *Globalization and its outcomes*, ed. J. O'Loughlin, L. Staeheli, and E. Greenberg, 361–85. New York: Guilford.
- Flint, Colin, and Ghazi-Walid Falah. 2004. How the United States justified its war on terrorism: Prime morality and the construction of a “just war.” *Third World Quarterly* 25 (8): 1379–99.
- Ginsberg, Robert. 1969. Introduction: Philosophy vs. war. In *The critique of war*, ed. R. Ginsberg, ix–xxiv. Chicago: Henry Regnery.
- Hardt, Michael, and Antonio Negri. 2000. *Empire*. Cambridge, MA: Harvard University Press.
- Johnson, James T. 1984. *Can modern war be just?* New Haven, CT: Yale University Press.
- Juergensmeyer, Mark. 2001. *Terror in the mind of God: The global rise of religious violence*. Berkeley: University of California Press.
- Kaplan, Lawrence F., and William Kristol. 2003. *The war over Iraq: Saddam's tyranny and America's mission*. San Francisco: Encounter Books.
- Khashan, Hilal. 2003. Baghdad's new Mongols. *The Arab World Geographer* 6 (1): 39–43.
- Krugman, Paul. 2003. George W. Queeg. *New York Times*, 14 March.
- Long, Jerry M. 2004. *Saddam's war of words: Politics, religion, and the Iraqi invasion of Kuwait*. Austin: University of Texas Press.
- Luke, Timothy. 1995. New world order or neo-world orders: Power, politics and ideology in informationalizing localities. In *Global modernities*, ed. M. Featherstone, S. Lash, and R. Robertson, 91–107. London: Sage.
- Lustick, Ian. 2003. America's war and Osama's script. *The Arab World Geographer* 6 (1): 24–26.
- Mamadouh, Virginie. 2003. 11 September and popular geopolitics: A study of websites run for and by Dutch Moroccans. *Geopolitics* 8 (3): 191–216.
- Mitchell, Timothy. 2002. McJihad: Islam and the U.S. global order. *Social Text* 20 (4) [73]: 1–18.
- Murphy, Alexander B. 1996. The sovereign state system as political-territorial ideal: Historical and contemporary considerations. In *State sovereignty as social construct*, ed. T. J. Biersteker and C. Weber, 81–120. Cambridge, U.K. Cambridge University Press.
- . 2003. Commentary II: Geopolitical perspectives on the Iraq war: Emerging insights and remaining challenges. *The Arab World Geographer* 6 (1): 54–57.
- National Security Strategy of the United States of America. September 2002. Washington, DC: The White House.
- O'Loughlin, John. 2004. The political geography of conflict: Civil wars in the hegemonic shadow. In *The geography of war and peace*, ed. C. Flint, 85–110. New York: Oxford University Press.
- Orend, Brian. 2000a. Michael Walzer on resorting to force. *Canadian Journal of Political Science* 33 (3): 523–47.
- . 2000b. *Michael Walzer on war and justice*. Montreal: McGill-Queen's University Press.
- . 2000c. *War and international justice: A Kantian perspective*. Waterloo, Ontario: Wilfred Laurier University Press.
- Ó Tuathail, Gearóid. 1996. *Critical geopolitics: The politics of writing global space*. Minneapolis: University of Minnesota Press.
- . 2003. “Just out looking for a fight”: American affect and the invasion of Iraq. *Antipode* 35 (5): 856–70.
- Ó Tuathail, Gearóid, and Simon Dalby. 1998. Introduction: Rethinking geopolitics, towards a critical geopolitics. In *Rethinking geopolitics*, ed. G. Ó Tuathail and S. Dalby, 1–15. New York: Routledge.
- Rengger, Nicholas. 2002. On the just war tradition in the twenty-first century. *International Affairs* 78 (2): 353–63.
- Sidaway, James D. 2003. Decentering political geographies. *The Arab World Geographer* 6 (1): 47–48.
- Smith, Neil. 2003. *American empire: Roosevelt's geographer and the prelude to globalization*. Berkeley: University of California Press.
- Taylor, Jonathan, and Chris Jasparo. 2003. Editorials and geopolitical explanations for 11 September. *Geopolitics* 8 (3): 217–52.
- Taylor, Peter J. 1993. The last of the hegemons: British impasse, American impasse, world impasse. *Southeastern Geographer* 33 (2): 1–22.
- . 1999. *Modernities: A geohistorical interpretation*. Minneapolis: University of Minnesota Press.
- Taylor, Peter J., and Colin Flint. 2000. *Political geography: World-economy, nation-state and locality*, 4th ed. Harlow, U.K.: Pearson Education.
- Temes, Peter S. 2003. *The just war: An American reflection on the morality of war in our times*. Chicago: Ivan R. Dee.

Wallerstein, Immanuel. 1984. *The politics of the world-economy*. New York: Cambridge University Press.

———. 2002. America and the world: The twin towers as metaphor. In *Understanding September 11*, ed. C. Calhoun, P. Price, and A. Timmer, 345–60. New York: Social Science Research Council.

———. 2003. *The decline of American power: The U.S. in a chaotic world*. New York: The New Press.

Walzer, Michael. 2000. *Just and unjust wars: A moral argument with historical illustrations*, 3rd ed. New York: Basic Books.

Correspondence: Dr. Ghazi-Walid Falah, Department of Geography and Planning, University of Akron, Akron, OH 44325-5005, e-mail: falah@uakron.edu (Falah); Department of Geography, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3671, e-mail: flint@uiuc.edu (Flint); AMIDSt, Department of Geography, Planning and International Development Studies, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands, e-mail: v.d.mamadouh@frw.uva.nl (Mamadouh).

Session 3: The West and the Far East – Past and Present

CHAPTER XIII

JUST WAR: CHINESE AND WESTERN PERSPECTIVES

EDMUND RYDEN, S.J.

Traditional Chinese Philosophy has a theory of Just War that is to be compared with Christian theories of just war. The present paper is part of a longer work discussing the subject. Its focus, on the Christian side, is the just war theory of Thomas Aquinas and Alexander of Hales, on the Chinese side, it is the *Huainanzi* as well as pre-Han such texts as the *Mozi* and *Laozi*. Comparison is also made with the Geneva Conventions regarding conduct in war as these illustrate the norms of *ius in bello*. In both fields, *ius ad bellum* and *ius in bello*, the Chinese texts display concerns and norms similar to those evidenced in the later Western texts.

Warfare is a matter of great importance for the State. It is the ground of life and death, the way to success and failure. It may not be studied. (The Art of War)¹

The sage's use of warfare is like using a comb on hair or a hoe on crops. What is removed is slight but its benefits are great. (Huainanzi)²

In previous works I have discussed the notion of "just war" in early Chinese philosophy (6th Century B.C.– 1st Century A.D.).³ That work was carried out from the viewpoint of Chinese philosophy and with no relation to the Western treatment of the same issue. This work takes a different tack, presenting the Chinese material within the framework of the Western discussion, noting differences and similarities. From the Western tradition I will focus on the works of Alexander of Hales and Thomas Aquinas. For a more complete discussion of the topic in the Western tradition, I refer the reader to my work, *Nonviolent Resistance*.

The present essay presents a preliminary study of the context of just war discourse in Christian philosophy and Chinese culture and then examines the respective just war traditions point by point, firstly with respect to two medieval Christian philosophers and then with respect to the 1949 Geneva Conventions.

THE CONTEXT OF THE DEBATE ON WAR

Aquinas and the scholastics addressed the question of war under the heading, *Utrum bellum sit semper peccatum*, Whether war is always a sin? (Aquinas) *Utrum bellum sit intrinsece malum*. Whether war is intrinsically evil? (Suarez). This negative formulation of the question is not without its modern critics. What it does do, however, is face the reality of sin in the world. The world is not yet paradise; sin is present. In the Chinese tradition, we obviously do not find the Hebrew notion of sin, but we find that the context for the discussion of war is one that takes account of conflict present in the world.

Chinese texts present a legendary history of the world typically centered around the virtuous figure of the Yellow Emperor.⁴ The Yellow Emperor is a paragon of virtue and yet he is credited with inventing or at least using weapons. This creates a paradox and thus initiates a discussion as to whether it would be possible to eliminate all warfare. The answer, based on this use of the legendary origin, is that in the real world in which we live, this would not be possible. In an ideal world, it may not be so.

Warfare is never judged as a good in itself: Weapons are instruments of misfortune and not the instruments of a gentleman.⁵

Three evils: The first is liking evil instruments; the second is promoting a rebellious spirit; the third is being led astray by the lusts of the heart. These are called "the three evils."⁶

Thus while the language is not theological in the Judaeo-Christian sense, the parallel is very strong. War, warfare, weapons (the same Chinese word *bing* covers all these uses) are intrinsically evil. Yet their use cannot be completely forbidden:

The Yellow Emperors, Tang and Yu, were the most glorious emperors. They possessed the entire realm and its control lay in the hands of a single man. Yet [even] during this time armed forces were not dispensed with.⁷

This situates the question in a legendary past, which we may take as a philosophical device for discussing the abstract proposition of how these evil weapons relate to justice. The same idea is presented in the *Lushi Chunqiu* in a discussion of King Hui's desire to disband the army:

King Hui of Zhao asked Gongsun Long, 'I have wanted to disband the army for over ten years without success. Is it that the army cannot be disbanded?' Gongsun Long replied, 'the idea of disbanding troops comes from a heart which seeks to love the world impartially. Loving the whole world impartially cannot just be empty words. It must have practical effects.'⁸

Gongsun Long argues that the disbanding the army is an unrealistic ideal; war is a fact of human life.

At this point, it is worth noting that many contemporary discussions of war and justice do not even mention the reality of sin or evil. There seems to be a tendency to think in terms of ideals without sufficient awareness of practical realities. The ancient Chinese and the medieval people shared our desire for a peaceful world without war, but they were also aware of the reality of evil. Indeed, this element is even further emphasized in Chinese discourse, as will be pointed out.

MEDIEVAL CHRISTIAN PHILOSOPHY AND ANCIENT CHINESE TEXTS ON WAR

Alexander of Hales' *Summa* is more detailed than the corresponding one of Aquinas. It seems to have attained its final form some 10 years or more after Alexander's death.⁹ This would bring it to 1255, a time when Aquinas was teaching in Paris. Alexander's work may be seen much more as a textbook of knowledge, rather than an attempt to think creatively with regard to the tradition.

Alexander's work is notable for its elaboration of six conditions for assessing the justice of a war. The first three are objective: (1) The first considered here is *meritum* (desert), which is concerned with evaluating the way in which an injury should be dealt with.¹⁰ (2) Authority is the question of who can declare war. (3) Alexander demands that the soldiers be secular, not clerical, and uses the Latin term *conditio* to refer to this. There are three subjective grounds for war which present both the purpose of the war and the way in which war is conducted: (4) *causa*, the purpose of the war, (5) *affectus*, the way in which the soldiers fight, and (6) *intentio*, the way in which the war is conducted.

Meritum. The Western tradition justified war on the grounds that it upheld the civilised order. In the post-medieval period this became equated with national sovereignty. The ancient Chinese tradition arises in a different political context. Nominally the King, or later the Emperor, claimed authority over the whole known world. This authority was thus greater than that assumed even by the Roman Emperor, who was always aware of other civilizations, for instance, Greece, Persia, and Egypt. The claim to universal authority by the Chinese monarch meant that war could always be seen as a form of rebellion and, hence, war carried out by the monarch would be seen in terms of punishment. This motive was present in Alexander but has been gradually eliminated from the Western tradition, perhaps because no nation-state could ever pretend to universal authority.

War is a form of punishment and the removal of evil. At least eight early texts repeat this message in words similar to those given below:

Punish tyranny and disorder and prevent unrighteousness.¹¹

Hence, the values that are being upheld are those of an ordered and just society.

Even when the context is not that of the universal monarch, the idea is still that what exists should be preserved. The *Heguanzi* describes the role of the just minister as preserving the dying and maintaining the line of rulers with no issue, to save the weak and punish tyranny.¹²

This idea is taken up in earnest by the Mohists. They are generally credited with being anti-war, but this is to misread their position. The relevant chapter in the *Mozi* is entitled *Against Aggression*.¹³ The opposition is not to war as such, but to aggressive warfare. Indeed, the Mohists were organized so as to defend the small states that existed in 6th to 4th century China under the nominal auspices of the Zhou King. A large part of the *Mozi* is devoted to the topic of making ballistic devices that can be used to defend towns under attack. The *Mozi* presents attacks on these small states as a case of theft or murder. Just as these states merit punishment as crimes, so too does war, which causes even more people to suffer. The following passage comes in the *Mozi*, but is perhaps not to be ascribed to the direct Mohist tradition:

To kill one person is said to be unjust and for this the murderer must be executed. Extrapolating from this, to kill ten persons is ten times more unjust and demands that the ten murderers be executed. To kill a hundred persons is a hundred times more serious and requires that the hundred murderers be executed. Now, when all reasonable men know something to be wrong they say it is unjust. Yet when it comes to the great injustice of attacking states they do not acknowledge it as unjust.¹⁴

One of the three schools of Mohism applies this idea of punishment to the wars, which it approves. In these, heaven is said to give signs and speak clearly ordering an attack:

Three gods spoke to [King Wu] in a dream, saying, now that we have submerged Zhou of Yin in wine, you go and attack him. We will surely let you destroy him. . . . Speaking of the work of these three sages [Yu, Tang and Wu], it is not to be called attack but punishment.¹⁵

Thus, war of this kind is preserving the established order, whether on a large scale or on a small scale.

To this should be added another purpose of war, namely to stop war. This is set out in the opening paragraph of the *Sima Fa*. Government is seen as best achieved by the Confucian virtue of benevolence, *ren*, combined with justice *yi*:

In antiquity, taking benevolence as the foundation and employing justice to govern constituted 'uprightness.'¹⁶

When this combination, based on consensus and harmony, was unable to achieve its objective then resort to political authority based on military might may be employed. Thus, the text states:

For this reason, if one must kill to give peace to the people, then killing is permissible. If one must attack a state out of concern for the people, then attacking is permissible. If one must stop war with war, although it is war, it is permissible.¹⁷

We may also put the question of *meritum* in a negative way: If war is not waged, what will be lost? The *Huainanzi* is quite explicit on this point. Failure to go to war simply enables tyrants to go even further in their cruelty. It is better to nip the evil in the bud than allow it to proliferate:

There is no greater disaster than killing innocent people or sustaining an unjust ruler. There is nothing more abysmal than appropriating the wealth of all under heaven and apportioning it to the wants of one man.¹⁸

The passage goes on to relate how if Jie of the Xia and Zhou of the Yin Dynasties had been stopped, then they would not have barbecued people. A further two rulers who were also not stopped early on are mentioned.

These four princes all committed small faults but since none disputed them they went on to possess the whole realm and harm the common people. By giving free rein to the evil of one individual, they brought misfortune to all within the oceans. This is something that First Principles cannot allow.¹⁹

The term translated as "First Principles" literally translated is "Heaven's Reasoning" (*Tian lun*). It might well be considered an equivalent of the term "Natural law." According to the *Huainanzi*, the wrongdoer should be removed just as the otter must be taken out of the fish pond or the wolf out of the pens of chickens and flocks of animals.²⁰ The purpose of war is thus seen as protection of the innocent.

Authority. The *Mozi* is exceptional in Chinese tradition in speaking so approvingly of popular religion, but it shares with the whole tradition a belief that heaven alone had the authority to command war. The *Sima fa* describes the ceremony in the temple prior to going into war:

They then announced it to August Heaven and to the sun, moon, planets and constellations. They prayed to the gods of earth, the spirits of the four seasons, mountains and rivers and at the great altar [of state]. Then they offered sacrifice to the former kings. Only thereafter would the Prime Minister charge the army before the feudal lords, saying, a certain state has acted contrary to the Way. You will participate in the rectification campaign on such a year, month and day. On that date the army will reach the [offending] state and assemble with the Son of Heaven to apply the punishment of rectification.²¹

This formal ceremony indicates that the rulers sought celestial authority for their wars in combination with the Zhou King, the Son of Heaven.

The state of Yue, which engaged in war with Wu after having first been defeated, did so on the advice of Fan Li, who argued that the King of Yue needed to wait for the approval of heaven, in the form of famine in Wu, and the support of the local population of Wu in the form of resentment against their ruler.²² Thus authority is seen as coming from outside, whether from above or from the appeal of the people of the enemy state. Authority is never seen as proceeding legitimately from the ruler himself.

Conditio. Aquinas discusses the question whether clerics and bishops may fight. His conclusion is that they may not. For this he produces two arguments: one general and one particular.²³ The general argument states that clerics are ordained to a life of prayer for the people, and anything which would disturb this contemplative life is to be shunned, whether it be business affairs or war. The second reason is to do with killing. All clerics are ordained to the ministry of the altar where the passion of Christ is present in sacramental signs. Since on this altar they remember the blood Christ shed for us, it would not be fitting for them to kill others and shed others' blood, rather they themselves should be more ready to give their blood for Christ. In other words, they may be martyrs but not soldiers.

Causa. The question of just cause is one that also receives treatment in the Chinese tradition. It is found in the characterization of the possible motives for war, which are typically given as three or five possible descriptions of troops. Of these categories, only two out of five or one out of three are approved.²⁴ We may list them according to source as follows:

-
Category/Source *Wuzi Wenzi* and Han History *Jing*

1. Just War 1 1 2

2. Self-defense 2 2 -

3. War of anger 3 3 3
4. War for gain 4 4 1
5. War of rebellion 5 5 -

The last three categories in this table are all objectionable; only the first two merit approval. The second, for instance, is said to lead to victory by the *Wenzi*. It is the first, however, that gains the most attention. The tradition is that just wars are possible. Their purpose is stated negatively as preventing tyranny and disorder and removing evils, and positively, as rescuing suffering people, the weak, and rulers whose line is threatened with extinction. The negative role is summed up in the beautiful metaphor of the *Huainanzi* placed at the head of this chapter: war is like a comb or a hoe.

As I have shown elsewhere, the dominant image in Chinese thinking on just war is, however, that of water. War is designed to clear out the weeds in the pool and so to purify the water. Once this is done the pond will become peaceful and calm. Simply letting the water be is not sufficient; the weeds must be removed first.²⁵

While the military works are all unanimous on the role of the just war, there are other voices in Chinese thought. The *Mencius*, in reviewing Chinese history, decides that there have never been any just wars:

In the Spring and Autumn Annals there are no just wars. Instances there are only of one war better than another.²⁶

This statement should not be seen as denying the need to assess wars in terms of justice, because it supposes the concept in order to make the judgment. What it does do is use the criterion of just cause to examine historical examples.

The statement of just cause given by the *Huainanzi* deserves special mention for its similarity to that of Alexander:

support of the good, coercion of the bad and peace for all. (Alexander of Hales)²⁷

The sage emerged and rose up and disputed with the strong and tyrannous, pacified a confused world and leveled out dangers, removing dirt. (Huainanzi)²⁸

Affectus. *Affectus* refers to the state of mind of both the soldiers and the person responsible for declaring war. As Nordquist points out, Augustin required "a mournful mood" as a condition for war, but in modern discourse this mental prerequisite is omitted, or perhaps, better, is converted into a legal and objective assessment.²⁹ In ancient Chinese tradition it is an important aspect. As noted above, war motivated by anger is condemned in all military texts. This applies not only to the inception of the war, but even to its continuation. A war might have begun with good *affectus*, but if it is carried on too far then it may be self-defeating and unjust.

The *Daode Jing* describes the victorious general as weeping for the slaughter of so many persons.

When victorious, he does not think this a good thing
Because to do so would be to delight in slaughtering people.

At the slaughter of such masses of people,

He weeps for them, mourning and lamenting.

When victorious in battle,

He takes his stand as at a mourning ceremony.³⁰

This brief phrase is expanded in the *Huainanzi*, which describes the traditional way in which the Chinese general would go to war. He dresses as a corpse and leaves by the gate reserved for the dead. Whereas some earlier texts suggested that he returns in glory, this one declares the opposite. The general comes back full of apology, in the very mournful mood that will be described later by Augustin.³¹

Intentio. *Intentio* rules out fighting for booty or from cupidity. In the Chinese context, this is called fighting for gain or profit. The *Jing* gives a certain degree of approbation to this motive:

What is meant by profit is, the ruler sees [the land is suffering] famine, the state is not at rest, superiors and inferiors are at odds with one another, yet the ruler raises an army and thus causes misery. Although there is no great profit, there is no great harm either.³²

The *Huainanzi* does not share even this qualified support:

One who wages war for land cannot become king of that land; one who wages war for himself cannot establish his merit. If one undertakes affairs for others then all will help one; if one undertakes affairs for oneself then all will desert one. When all come to help then even what is weak will be strong; when all desert then even what is large will be lost.³³

The one who acts for profit, thus, will ultimately lose whatever gains are made. The Purist version of the *Mozzi* holds that aggressive warfare aroused enmity in surrounding states and, thus, was ultimately responsible for the destruction, sometimes total, of the original aggressor:

It was due to offensive wars that Ju perished between two great states. And it was due to aggressive wars that in the south Chan and Cai were extinguished by Wu and Yue. And it was due to aggressive wars that in the north Bu Tu He perished among Yan, Dai, Hu and Mo.³⁴

The passage goes on to relate how King Fu Chai of Wu attacked Qi and then Yue and won control over all of China, but because of his pride, Yue was able to invade and totally wipe out the kingdom of Wu.

If war for profit and out of anger are to be ruled out, the Chinese tradition reserves its greatest disapproval for wars of rebellion. In the Western tradition, rebellion and regicide are also discussed with similar disapproval in the majority of cases. However, since these kinds of armed conflict do not fit the narrower definition of war that has been chosen as the object of this study, I will not enter into greater detail. In fact, from the point of view of philosophical reasoning about war, rebellion may be seen as a particular application of the norms of "just war theory."

Conclusion: Last Resort. This criterion is upheld consistently in all military texts. Only when it cannot be avoided is warfare to be engaged in:

He [the gentleman] uses them [weapons] only when unavoidable.³⁵

This same statement is repeated in other texts. For instance, the *Jing* after listing the three different causes of war adds a fourth paragraph explaining the circumstances in which war is to be engaged in:

The use of them [weapons] in conformity with the

Way is only when unavoidable.

If used when unavoidable

Then success is unlimited.³⁶

The belief that success will be achieved when warfare is consistent with the Way reveals a basic metaphysical assumption of most Chinese reasoning: Only conformity to the Way brings success. In this sense, the military formulation is simply the application of a general rule.

IUS IN BELLO

Under this heading, we shall look at the way in which warfare is conducted and the limits imposed. Before considering equivalents of the principles of discrimination and proportionality, it will be good to draw attention to a constant theme in all Chinese military texts, namely, that the best kind of war is one in which there is no warfare:

A hundred engagements, a hundred battles, is not the best. To overcome the others' troops without any fighting, this is the best.³⁷

The truly just army acts in the same way, not shedding any blood:

The coming of the just army is to manage to reach a state of not fighting and then stop.³⁸

Thus, in all these texts, we find the essential norms of the Geneva Conventions that govern international norms for conduct in time of war. To bring out this point in greater detail, we will examine the Conventions alongside the relevant Chinese texts.

Treatment of Victims and Civilians. There are six documents relating to norms for the conduct in war. The first two deal with the wounded, sick, and shipwrecked members of the armed forces. The third and fourth deal with prisoners of war and civilians, respectively, whilst the last two are additional protocols relating to the protection of victims in international and noninternational armed conflicts.³⁹ Naturally, the details of a modern legal document go far beyond what can be found in any of the Chinese philosophical texts, but the essential principles are the same.

The Four Geneva Conventions state how noncombatants and those who have laid down arms are to be treated:

Persons taking no active part in the hostilities, including members of armed forces who have laid down their arms and those placed *hors de combat* by sickness, wounds, detention, or any other cause, shall in all circumstances be treated humanely, without any adverse distinction founded on race, colour, religion or faith, sex, birth or wealth, or any other similar criteria.⁴⁰

The Second Protocol states that civilians and objects indispensable to their survival are not to be attacked:

The civilian population and individual civilians shall enjoy general protection against the dangers arising from military operations. . . . It is . . . prohibited to attack, destroy, remove or render useless . . . objects indispensable to the survival of the civilian population, such as foodstuffs, agricultural areas for the production of foodstuffs, crops, livestock, drinking water installations and supplies and irrigation works.⁴¹

Turning to Chinese philosophical texts, we find that these same principles are all observed. Thus, the Prime Minister and other officials deliver the following command to the army before it invades an adversary's state:

When you enter the offender's territory, do not do violence to his gods; do not hunt his wild animals; do not destroy earthworks; do not set fire to buildings; do not cut down forests; do not take the six domesticated animals, grains, or implements. When you see their elderly or very young, return them without harming them. Even if you encounter adults, unless they engage you in combat, do not treat them as enemies. If an enemy has been wounded, provide medical attention and return him.⁴²

The *Huainanzi* repeats this section adding prohibitions against the burning of stores and the digging up of tombs. The *Liu Tao* adds the further stipulation that captives are not to be slain whilst the *Xunzi* forbids making prisoners of any who willingly surrender.⁴³

Respect for tombs is also set out in the Geneva Conventions:

They shall further ensure that the dead are honorably interred, if possible according to the rites of the religion to which they belonged, that these graves are respected, grouped if possible according to the nationality of the deceased, properly maintained and marked so that they may always be found.⁴⁴

Regarding the treatment of civilians, the *Huainanzi* explicitly mentions care for orphans, widows, and the poor:

They [the invading army] support orphans and widows, cherish the poor, empty the prisons and reward the virtuous.⁴⁵

In the Fourth Geneva Convention, we find:

the Parties to the conflict shall endeavor to conclude local agreements for the removal from besieged or encircled areas, of wounded, sick, infirm, and aged

persons, children and maternity cases, and for the passage of ministers of all religions, medical personnel and medical equipment on their way to such areas.⁴⁶

Ruses and Perfidy

One principle that is common to all Chinese military texts is that ruses should be employed. The *Sunzi* puts it bluntly: "Warfare is the way of deception."⁴⁷ The hero of the novel *The Three Kingdoms*, Zhuge Liang, is famed for his ability to use ruses.⁴⁸ Aquinas also permits the use of lying in war. However, perfidy is to be condemned. The First Protocol to the Geneva Conventions makes a distinction between the two:

It is prohibited to kill, injure or capture an adversary by resort to perfidy. Acts inviting the confidence of an adversary to lead him to believe that he is entitled to, or is obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence, shall constitute perfidy.

Ruses of war are not prohibited. Such ruses are acts which are intended to mislead an adversary or to induce him to act recklessly, but which infringe no rule of international law applicable in armed conflict and which are not perfidious because they do not invite the confidence of an adversary with respect to protection under that law.⁴⁹

An example of perfidy would be to pretend to surrender, but once the opposing troops were close by to then open fire, or again, to kill prisoners of war. Ferguson shows how during the First World War, atrocities of this kind happened on both sides partly owing to a lack of trust.⁵⁰ The holding power might fear that prisoners would break trust and open fire and hence, they would kill the prisoners first. Tension and resentment could also lead to such acts of perfidy. Ruses, the use of moving machines, dummies, camouflage, and other activities are what make Zhuge Liang such an exciting general. Here it is a question of skill and intelligence, rather than of trust.

Hence, from this brief study it is clear that the basic principles enshrined in the Geneva Conventions were already present some 2,000 years previously in ancient China.

THE SORROW OF WAR

While there is no basis in the Chinese tradition for pacifism as a counterweight to the theory of just war: there is, nonetheless, a constant poetic theme of regret for war and the sorrows it brings. The theme of being parted from friends and from loved ones is particularly poignant. For much of Chinese history, the major danger of war has always come from the north where the nomadic peoples known as Huns, Mongol, or Manchus

were adept at riding in the saddle and sweeping down on the civilised areas of China. War was thus a disruption of normal life as the following poem makes clear:

The Ballad of Longxi

Oath-bound to sweep aside the Huns, without caring for their own lives,

Five thousand soldiers with fur-trimmed coats died in the dust of the frontier.

O, the pity of it! These bones by the banks of the inconstant river are yet men

in the Spring dreams of their wives.⁵¹

Liu An, the Prince of Huainan appeals to these kinds of sentiment when he condemns a proposal to send an army to subdue barbarians in southeast China. He points out that the climate of the region is such as to engender cholera, yellow fever, and other epidemics. Geographical obstacles such as mountains and rivers all add to the difficulties, just as the vast deserts of the north make combat difficult. In a previous study of Liu An's argument I also noted that he has other reasons, related to his theory of humanitarian intervention, which rule out attack.⁵²

CONCLUSION

The Chinese texts discussed here show remarkable parallels to the Western texts. There is a similar assessment of the evil of war, a recognition, though, that in an imperfect world, it may be used but only subject to moral conditions that bear striking resemblance to those in the Western tradition. Norms of proportionality and discrimination are designed to guarantee the safety of civilians in the same way as do the Geneva Conventions.

Naturally, the comparison is not perfect, especially because the Chinese texts considered are all over 2,000 years old while in the West we were looking at a much more recent history, noting how ethical discussion developed over the centuries. But these historical developments were not seen as engendering a fundamental shift in the way of thinking. Even the extreme case of applying the proportionality or discrimination arguments to rule out war is by definition precisely that: a use of these arguments.

Director

John Paul II Institute for Research into Dialogue for Peace

Fu Jen Catholic University

Taipei, Taiwan

China's Assertive Behavior

Part One: On "Core Interests"

Michael D. Swaine

(I am deeply indebted to Rachel Odell and Tan Weilu, Carnegie Junior Fellows, for their invaluable research assistance in the preparation of this article.)

Among both casual observers and experts alike, the single most dominant theme in Sino-U.S. relations of the past year or more has been the emergence of a more "assertive China." In *CLM 32*, we examined how both Chinese and outside observers look at China's growing assertiveness on the international stage, that is, the purely perceptual dimensions of the issue. In this and several subsequent CLMs, we intend to assess whether, to what extent, and in what manner, the Chinese government is becoming more assertive in several major areas of relevance to the United States: First, in defining and promoting the concept of "core interests"; second, with regard to U.S. political and military behavior along China's maritime periphery; third, concerning a variety of economic, trade, and finance issues, from so-called indigenous innovation to global standards regarding reserve currencies; and fourth, with regard to several issues related to international security, from counter-proliferation to climate change.

In each of these four areas, we shall to varying degrees attempt to answer several basic questions regarding Chinese assertiveness that build on those addressed in *CLM 32*: In what ways are Chinese leaders becoming more assertive, employing what methods, and to what apparent ends? Is Chinese assertiveness a "new" and highly significant phenomenon for U.S. interests, and if so, in what manner? What misconceptions, if any, exist about China's assertiveness? What internal and external forces are driving China's assertive behavior? In particular, is Chinese assertiveness associated with particular interest groups or factions within Chinese state and society? How is China's assertiveness evolving in response to both inside and outside pressures? And finally, what do the answers to the foregoing questions tell us about the likely future direction and strength of China's assertiveness over the next several years?

What Kind of Assertive Behavior?

As indicated in *CLM 32*, China's assertiveness means different things to different people. As a result, the concept, in describing Chinese behavior, is somewhat vague and ambiguous, potentially encompassing everything from attempts to play a more active role

in a wide variety of international regimes, to deliberate efforts to alter basic international norms and challenge the fundamental national interests or policies of the United States. In addition, there are many forms of assertiveness, from mere verbal statements or comments, to concerted official actions that appear designed to intimidate or even to force other nations or foreign entities to change their behavior. As this typology suggests, some forms of Chinese assertiveness are probably beneficial to the workings of the international system and U.S. interests while others are not. Indeed, U.S. officials welcome a more active, engaged China that seeks both to strengthen and to shape international institutions and norms in ways that advance prosperity, stability, and the peaceful resolution of problems. They presumably do not welcome a China that desires or appears to do otherwise.

In addition, not all indications of Chinese assertiveness (whether “good” or “bad” for the United States and other Western powers) are sanctioned or supported by the Chinese government. Indeed, as we have seen in *CLM* 32, many unofficial Chinese observers and pundits express or advocate various levels and types of assertiveness that are not reflected in official Chinese statements or documents.

Thus, any assessment of Chinese assertiveness must distinguish between official and unofficial actions or utterances, productive or creative assertiveness (what one might call “positive activism”) and confrontational, destabilizing, or threatening (from a Western or U.S. perspective) assertiveness. This essay, and those that follow, focuses primarily on identifying, measuring, and assessing official or governmental forms of negative or potentially threatening Chinese assertiveness, given its clear significance for future Sino-American relations and the obvious attention that it has received among outside observers.

Why “Core Interests”?

Since at least November of 2009, when it was inserted in the U.S.-China Joint Statement between Hu Jintao and Barack Obama during the latter’s state visit to China,¹ the notion of China’s “core interests” (核心利益) has received enormous attention among both media pundits and experts alike. Many observers interpret the use of this concept by the PRC government as an indication of strong (and growing) Chinese assertiveness in the international arena, for three apparent reasons: first, because in recent years the concept has been more formally defined and included in official PRC (and at least one bilateral U.S.-PRC) statements and documents to a greater extent than in the past; second, because some Chinese officials and unofficial observers have apparently asserted that China’s “core interests” are essentially nonnegotiable in nature, thus conveying a level of rigidity and perhaps militancy toward whatever issue might be defined as a core interest; and third, because China is allegedly steadily defining more and more controversial international issues as affecting its “core interests,” including U.S. arms sales to Taiwan, meetings between foreign leaders and the Dalai Lama, and disputed territories in the South China Sea, thus by implication challenging an array of foreign activities relating to such issues. In the remainder of this essay, we shall examine these observations in turn.

Increasing Usage and an Official Definition

An examination of the historical record indicates that the Chinese government has indeed in recent years invoked China's "core interests" far more frequently, and presented publicly a more explicit definition of the term, than it has done in the past.² In fact, official Chinese sources only began referring to China's "core interests" on a fairly frequent basis in 2003–2004. The term was initially used in Chinese official media during the 1980s and '90s only in reference to the interests of *other* nations.³

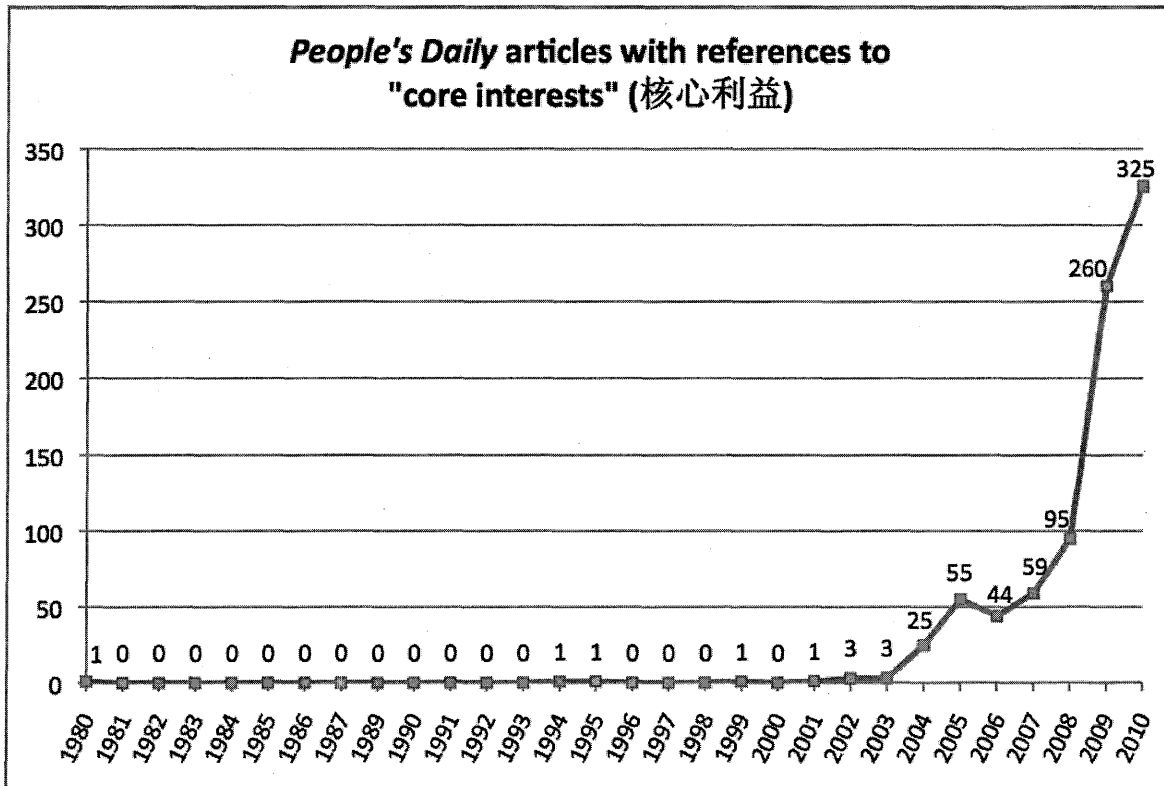
It was first used with reference to China in the mid-'90s and in the first years of the new century, but primarily in a domestic context. At that time, the term was closely associated with and seemed to emerge from the term "fundamental interests" (根本利益) as applied to China's economic- and social-reform policies and the general maintenance of domestic order and stability.⁴

The term "core interests" has also been used in official PRC media alongside the term "major concerns" (重大关切).⁵ The latter term was in fact employed earlier than "core interests" in official PRC media and at times was used in joint statements between China and foreign governments, for example, in a report on a meeting between Jiang Zemin and President Chirac of France in 2000. It has also been used to refer to the Taiwan issue and the one-China principle.⁶

The term "core interests" was apparently first applied to China in a foreign context in PRC media in early 2002, but in an unofficial capacity, in an article written by a Chinese scholar.⁷ The first official foreign-oriented reference to the term "core interests" appeared in the report of a meeting between Foreign Minister Tang Jiaxuan and Secretary of State Colin Powell on January 19, 2003, in which Tang identified Taiwan as among China's "core interests." (See below for more on the link between Taiwan and the emergence of the official PRC use of "core interests").⁸

The first apparent official identification of the oft-espoused concepts of "sovereignty and territorial integrity" (主权和领土完整) as a Chinese "core interest" occurred in April 2004, again in the context of a discussion of Taiwan.⁹ During the remainder of 2004, both official and unofficial Chinese usage of the term "core interests" in reference to sovereignty and territorial interests (and Taiwan in particular) increased significantly.¹⁰ The related issue of "national security" (国家安全) was apparently first explicitly identified officially as a core interest in a speech given by then Foreign Minister Li Zhaoxing in September 2006 and reported in the *People's Daily*.¹¹

As one might surmise from the above references, major official and unofficial PRC media mentions of China's core interests in a foreign-policy context increased notably beginning in the early 2000s, from a mention in one *People's Daily* article in 2001 to 260 articles in 2009 and 325 articles in 2010.



Moreover, by 2004, Chinese officials had begun routinely mentioning the need for countries to respect and accommodate one another's "core interests" in speeches with foreign officials and dignitaries, thus indicating that the term had not only entered the official lexicon but also become an important element of PRC diplomacy.¹²

It is therefore not surprising that the senior Chinese official responsible for PRC foreign policy (State Councilor Dai Bingguo) publicly defined the general elements of China's core interests in July 2009, during a session of the U.S.-China Strategic and Economic Dialogue (S&ED). Dai stated in his closing remarks at the S&ED that the term includes three components: 1) preserving China's basic state system and national security (维护基本制度和国家安全); 2) national sovereignty and territorial integrity (国家主权和领土完整); and 3) the continued stable development of China's economy and society (经济社会的持续稳定发展).¹³ Variations of this multi-part definition have occurred officially since that time, and have been repeated by unofficial Chinese sources as well.¹⁴

In addition, Chinese officials have also at times identified "national unity" or "reunification" (国家统一/两岸统一/统一大业) as a Chinese core interest, as well as "independence" (独立), in some instances alongside the three elements listed above. However, the former references were almost invariably intended to buttress the Chinese position regarding issues associated with territorial integrity, such as Taiwan, and hence can be taken as largely duplicative of an element contained in Dai Bingguo's list.¹⁵ In the case of "independence," references have been very few in number and have always occurred in a bilateral or multilateral context (with regard to the "core interests" of both

countries or of countries in general); in some cases the word was inserted within the phrase “sovereignty and territorial integrity,” as in: “The defense of sovereignty, independence and territorial integrity is the core interest of every country.”¹⁶ Thus, the concept is, as with “national unity” or “reunification,” most closely associated with the general category of sovereignty and territorial integrity as a core interest.

Chinese officials have also identified “human rights” as a Chinese core interest. However, this has only occurred nine times in official Foreign Ministry sources. Hu Jintao used it twice, in November 2006, during state visits to Laos and Pakistan. In both instances, he was expressing China’s appreciation for the support the two countries have extended to Beijing’s position on “. . . Taiwan, Tibet, human rights and other major questions involving China’s state sovereignty and core interests.”¹⁷ The seven other references include two from Yang Jiechi in 2008 and five from various ambassadors, including statements by Zhou Wenzhong, ambassador to the United States, in November 2009, and Song Zhe, ambassador to the EU, in December 2008.¹⁸ Moreover, overall, this context seems to suggest that human rights as a core interest refers primarily to Beijing’s right to determine how the lives of China’s citizenry will be promoted, especially in contested regions such as Taiwan and Tibet, for example, via the advancement of local economic and social conditions. In other words, the issue is again associated with domestic interests or other core interests involving sovereignty and territorial integrity.

Therefore, the most common and authoritative formulation of the general concepts comprising China’s core interests appears to remain that of Dai Bingguo, presented in July 2009. That said, since Dai articulated his definition, Chinese officials have continued to place the most emphasis on “sovereignty and territorial integrity” as the most important characteristic of China’s core interests. The first and third elements of Dai’s definition—“basic state system and national security,” and “continued stable social and economic development”—are still only infrequently mentioned in the context of China’s “core interests.”

Motivated by the Taiwan Issue?

It is not entirely clear what prompted official Chinese sources to begin employing the term “core interests” to such a degree and in this manner. Of course, the defense or protection of China’s national security, the PRC system or regime, and Chinese sovereignty and territorial integrity, as well as the protection and advancement of China’s economy and society, has been a staple of PRC foreign policy for decades. Indeed, they are basic to any nation’s definition of its national interests.

That said, it is quite likely that Beijing began to sharpen and promote vigorously the concept of “core interests” in response to growing concerns over the Taiwan issue. By 2004–2005, Beijing had become extremely worried about what it regarded as the efforts of former Taiwan president Chen Shuibian to achieve *de jure* Taiwan independence, possibly with U.S. backing. In response, during that time, the National People’s Congress promulgated the so-called Anti-Secession Law (ASL, in March 2005), and PRC officials began pressing (or warning) the United States and other countries to reject Chen’s efforts

and to recognize China's vital interests on the issue.¹⁹ As indicated above, it is precisely at this time that Chinese officials began to emphasize China's core interests, and to specify Taiwan as a primary example.²⁰ Indeed, for some unofficial Chinese observers, Beijing's "core interests" are primarily about sovereignty and territorial integrity.²¹

Used as a Warning and for Diplomatic Leverage

As suggested above, China's relatively recent and repeated invocation of the phrase "core interests" generates concern among both foreigners and some Chinese in large part because of: 1) Beijing's efforts to pressure foreign governments (and especially the United States) to officially acknowledge acceptance of the general concept and the specific policy issues to which it applies (such as Taiwan—discussed below); and, more importantly, 2) its apparent association with a rigid, uncompromising diplomatic or military stance. In other words, the appearance of the term appears to signal a more vigorous attempt to lay down a marker, or type of warning, regarding the need for the United States and other countries to respect (indeed, accept with little if any negotiation) China's position on certain issues.

Regarding the first point, beginning in the early 2000s, Chinese officials increasingly pressed the United States to issue formal statements indicating a willingness to respect one another's core interests (as indicated above), and even, in recent years, to explicitly and formally recognize the category of "core interests and major concerns" in general, as a necessary basis for the advancement of the bilateral relationship.²² This pressure campaign culminated in the inclusion of the term in the November 2009 U.S.-China Joint Statement. This was the first time that it had been used in an official, high-level Sino-American statement or communiqué. In fact, even in past meetings where senior Chinese officials were pressing their U.S. counterparts to respect China's "core interests," U.S. officials never repeated the phrase, but instead merely conveyed support for various long-standing U.S. policies, such as the "one China" principle and the three joint communiqués.²³

The 1972 and 1982 Sino-U.S. joint communiqués do affirm "respect for the sovereignty and territorial integrity of all states" and "respect for each other's sovereignty and territorial integrity" (respectively), which Beijing has since identified as one of its core interests.²⁴ However, by November 2009 Beijing had significantly expanded its definition of core interests to include several other general categories, as indicated in the July 2009 statement by Dai Bingguo, discussed above. Therefore, such past U.S. acknowledgments (of respect for China's sovereignty and territorial integrity) could not be viewed as a precedent for the U.S. acceptance of Beijing's "core interests" in the Hu-Obama joint statement.

Since the signing of the November 2009 Joint Statement, Beijing has repeatedly and emphatically cited the mutual commitment to respect one another's "core interests" contained in that document as a basis for its demands that Washington alter its behavior in a variety of areas, from arms sales to Taiwan to presidential meetings with the Dalai Lama. Indeed, the Chinese have branded the joint statement as an "important consensus"

that is a major step in the development of a “new era” (新时期) in U.S.-China relations. It is often mentioned in official Chinese sources alongside the three Sino-U.S. joint communiqués.²⁵ However, the reference to “core interests” was *not* included in the joint statement issued after Hu Jintao’s state visit to Washington in January 2011.²⁶ The reason for this omission is not entirely clear, but most likely reflects, at least partly, a U.S. desire to avoid the controversy that followed the inclusion of the term in the 2009 joint statement. At that time, some observers argued that the Obama administration had shown undue weakness in allegedly acceding to a Chinese demand to include a phrase closely associated with Beijing’s claim to sovereignty over Taiwan, and other supposed territorial ambitions.²⁷

Regarding the association of “core interests” with an uncompromising official PRC stance, the historical record suggests that, although Chinese officials have not to our knowledge used the phrase “non-negotiable” (不能谈判, 非流通, 不可转让, 免谈的) to describe China’s stance toward its “core interests,” they have certainly employed similar terms on many occasions. For example, officials have stated that China will “never waver, compromise, or yield” (决不动摇, 决不妥协, 绝不会让步), will not haggle or bargain (付价还价), and “must stand firm, be clear-cut, have courage to fight, and never trade away principles” (必须立场坚定, 旗帜鲜明, 敢于斗争, 善于斗争, 决不拿原则作交易) when dealing with its core interests, and with issues involving sovereignty and territorial integrity in particular.²⁸

Moreover, Chinese officials and official media sources have at times separately used the term “non-negotiable” to refer to issues that Beijing has described as a “core interest,” notably, sovereignty and territorial integrity, involving, for example, Hong Kong, Taiwan, and Tibet.²⁹ And in some cases, Chinese officials have explicitly connected the defense of specific core interests (again, most notably sovereignty and territorial integrity) with the possible use of force. Such language has been used quite consistently for years.³⁰ Of course, even official (and unofficial) statements of strong resolve and a refusal to compromise do not necessarily guarantee that Beijing would in reality in every instance employ such a rigid approach (including, perhaps the use of force) to defend what it has defined as its core interests. Yet, at least with regard to sovereignty and territorial issues, the historical record of China’s behavior suggests that such a possibility would be extremely high, and certainly cannot be dismissed.³¹

Hence, what China labels as its “core interest” is certainly significant. And so, perhaps the most important issue becomes, what specific policy areas does Beijing include among its core interests?

Territorial Issues (and Especially Taiwan) are at the Core

A large number of unofficial Chinese and foreign observers have identified a range of issues as being among China’s “core interests.” These include Taiwan, Tibet and Xinjiang-related issues; territories in the South China Seas; the defense of the Yellow Sea; the Diaoyu (Senkaku) Islands; bilateral trade; and the value of China’s currency. Some of these issues (such as Taiwan and access to international waters near China)

directly concern critical U.S. security interests.³² In truth, much of the unofficial commentary contains inaccuracies, distortions, and misconceptions. A close examination of the historical record, along with personal conversations with knowledgeable senior U.S. officials, confirms that thus far the Chinese government has officially, and repeatedly, identified only three closely related issues as specific core interests: the defense of China's sovereignty claims regarding Taiwan, Tibet, and Xinjiang.

As indicated above, the Taiwan issue probably originally triggered official use of the term "core interests" in the realm of foreign affairs and has clearly been most often associated with its official use.³³ On the subject of Tibet, Chinese officials have of course for many years referred to Tibet's status as part of China as an important or fundamental sovereignty issue. For example, the 1992 PRC white paper on Tibet stated that "there is no room for haggling" (讨价还价) on the fundamental principle (根本原则) that "Tibet is an inalienable part of China."³⁴ However, perhaps the first (or at least an early) occurrence of an official, explicit reference to Tibet as a Chinese "core interest" occurred in April 2006, in a meeting between PRC Vice President Zeng Qinghong and the prime minister of Sri Lanka.³⁵

Similarly, Chinese officials have often referred to Xinjiang's sovereign status as an important Chinese interest. The earliest use of the term "core interest" in reference to that Chinese region apparently also dates from 2006. In November of that year, in a speech in Pakistan, Hu Jintao first identified "the fight against East Turkestan" terrorist forces as a Chinese "core interest," alongside Taiwan, Tibet, and human rights. This clearly implies that the defense of China's sovereignty over Xinjiang (which the East Turkestan terrorist forces violently contest) is a PRC core interest.³⁶ On subsequent occasions (beginning largely in 2009, it seems), Chinese officials have referred simply to "Xinjiang" as being among China's core interests.³⁷

As far as we can surmise from the official PRC sources used in this study, references to the defense of the Yellow Sea, Diaoyu (Senkaku) Islands, bilateral trade, and the value of China's currency as Chinese core interests are entirely unofficial. In other words, we can find no official source stating that such concepts are among China's "core interests."³⁸

The reference to the South China Sea as a Chinese core interest is a more complex matter. The *New York Times* apparently first reported that Chinese officials had identified the defense of China's territorial claims to the South China Sea as a "core interest" in a private meeting held in Beijing in March 2010 with two senior U.S. officials, NSC Asia Director Jeffrey Bader and Deputy Secretary of State James Steinberg.³⁹ Many other media sources (and other *New York Times* reporters) subsequently repeated this initial *New York Times* story, often without citing it as the original source (or perhaps in some cases merely duplicating the story by interviewing the same U.S. official paraphrased in the *New York Times* story),⁴⁰ thus creating the impression that the report came from multiple sources. At least one media source subsequently asserted that Dai Bingguo had also identified the South China Sea in this manner to Hillary Clinton, at the May 2010

meeting of the S&ED in Washington.⁴¹ Clinton herself repeated this during a recent press interview in Australia.⁴²

However, a close examination of the official Chinese sources consulted for this study failed to unearth a single example of a PRC official or an official PRC document or media source that publicly and explicitly identifies the South China Sea as a PRC “core interest.” In fact, when given the opportunity to clarify the official record on this issue, Chinese officials have avoided doing so.⁴³ During their October 11, 2010, meeting in Hanoi, Chinese Defense Minister Liang Guanglie apparently did not mention the issue of the South China Seas to U.S. Defense Secretary Robert Gates.⁴⁴ And when President Hu Jintao traveled to Washington for his state visit in January 2011, he explicitly identified only Taiwan and Tibet as core interests.⁴⁵

In addition, personal communications with very knowledgeable U.S. officials confirm that Chinese officials did *not* explicitly identify China’s territorial claims to the South China Sea as a “core interest” in the March 2010 meeting with Steinberg and Bader. In that meeting, the PRC officials listed the issue as one among several about which they were attempting to elicit U.S. understanding and deference for Beijing’s position. Although this effort was viewed as a clear attempt to raise the importance of the South China Sea issue in China’s overall hierarchy of concerns, it did not explicitly constitute an effort to brand it as a “core interest.”⁴⁶

Regarding Dai Bingguo’s comment to Clinton at the May 2010 S&ED, a review of Dai’s remarks at the opening and closing sessions reveals no reference to the South China Sea issue as a PRC “core interest.”⁴⁷ Moreover, one very well-placed U.S. official confirmed to the author in a personal correspondence (conveyed prior to Clinton’s recent remark in Australia) that Dai indeed did not describe the South China Sea issue in this manner. In fact, all Chinese remarks regarding the South China Sea made at the S&ED were presented spontaneously by a lower-level official, and thus should not be regarded as authoritative, according to the U.S. official.⁴⁸

What then about Clinton’s recent remark, noted above? It is possible that Dai actually made the remark to Clinton in a private, offline, and unofficial conversation, or that Clinton: a) did not accurately recall what Dai said; b) mistook the abovementioned lower-level official for Dai Bingguo; or c) made the remark, knowing it was not true, to add to existing U.S. efforts to deter China from attempting to add the South China Sea to its list of core interests.

In any event, the foregoing information strongly suggests three conclusions: first, at the very least, Beijing has not unambiguously identified the South China Sea issue as one of its core interests, as it has done with Taiwan, Tibet, and Xinjiang. Second, even if Beijing did identify the issue as a core interest on one occasion (at the May 2010 S&ED), this was done in a decidedly unofficial manner. Third, although Beijing originally attempted in the March 2010 Steinberg/Bader meeting to raise the importance of the South China Sea issue as a Chinese interest in U.S. eyes, it has deliberately avoided

clarifying its stance on the matter since that time, thus creating the impression that it is backing away from the controversy.

Perhaps for some observers, the issue of whether or not Beijing has identified the South China Sea as a core interest is a purely semantic one, of little real significance, especially given China's apparent attempt to raise its relevance in March 2010. However, as noted above, the Chinese application of the term "core interest" to an issue is intended to convey a very high level of commitment to managing or resolving that issue on Chinese terms, without much if any discussion or negotiation (at least regarding basic questions such as China's ultimate sovereign authority, as in the case of Taiwan, Tibet, and Xinjiang). In other words, it conveys a high level of resolve, and to some extent a warning of sorts to other powers. In this particular instance, labeling China's claims to the South China Sea as a core interest would have signaled a significant, and alarming, shift in China's historical stance toward the issue. That stance not only recognizes the multinational nature of the South China Sea issue as a sovereignty dispute among several countries (albeit one that Beijing wishes to handle on a bilateral basis, with each claimant), but also seeks to convey Beijing's willingness to negotiate the ultimate nature and extent of Chinese sovereignty over the region. In contrast, Taiwan, Tibet, and Xinjiang are described as purely Chinese internal affairs not subject to dispute or negotiation.⁴⁹

But if the term "core interests" has such significance, why has Beijing avoided clarifying whether or not it applies to the South China Sea issue? Although it is impossible to say with certainty, it is probably because confirming the association would signal a clear shift in position that would likely provoke an even stronger international reaction than has occurred thus far (as indicated above), while an official denial of the association might convey an impression of weakness and retreat from China's basic stance on sovereignty and territorial issues, thus inviting domestic attack.⁵⁰ Moreover, in reality, Beijing has not clearly confirmed the precise nature and extent of its sovereignty claims to the South China Sea; hence, clarifying its stance on whether the issue constitutes a core interest could generate confusion and thereby force China to make such a clarification.⁵¹

Some unofficial Chinese observers have also argued that Beijing should not officially confirm that the South China Sea is a core interest because to do so would not only sow confusion among other nations, but also "... be used by unfriendly forces in the international community in a bid to contain China."⁵² Some Chinese academics even suggest that the United States was falsely accusing China of elevating the South China Sea to the level of a "core interest" in order to hype the China threat among China's neighbors, culminating in Secretary Clinton's orchestrated pushback against the Chinese at the ASEAN Regional Forum in late July 2010.⁵³ Other Chinese observers more broadly argue that China should be extremely cautious in describing any specific issues (including the South China Sea) as a "core interest," given the potentially provocative nature of the term to other nations.⁵⁴ However, as suggested in endnote 32, many unofficial Chinese observers argue that the South China Sea issue is or should be declared a core Chinese interest.

The preceding unofficial differences in viewpoint, along with the likely dilemma involved in confirming whether the South China Sea is a core interest, together suggest the possibility of disagreement among the Chinese leadership on this matter. If such disagreement exists, it is probably not along civil-military lines, however, since some PLA officers (such as Han Xudong and Yin Zhuo) oppose declaring the South China Sea a core interest, while others (such as Luo Yuan) support such a move.⁵⁵

Conclusion

As the foregoing analysis shows, Beijing's use of the term "core interest" with regard to issues involving the international community, and the United States in particular, is a relatively recent phenomenon. Its usage probably derives from growing Chinese concerns over the Taiwan issue in the early 2000s. However, the term has since been unambiguously applied to two other sovereignty-related issues (Tibet and Xinjiang), and its general coverage has been expanded to include three other general sets of state interests: the Chinese political system, national security, and socioeconomic development. Despite much reporting to the contrary, Beijing's territorial claims with regard to the South China Sea have not been clearly identified officially and publicly as a "core interest." Moreover, the application of the term by senior PRC officials to other general issues such as "independence," "human rights," and "national unity" or "reunification" seem intended to reinforce the existing primary emphasis placed on sovereignty and territorial issues as core Chinese interests.

The term "core interests" has its precursors and draws on long-held stances toward sovereignty and territorial issues. However, its increasing use in official statements and diplomatic documents, and its explicit application to specific contentious policy issues (most notably Taiwan) arguably signals an attempt by a stronger, more assertive Chinese leadership to elicit greater respect and deference from other nations for China's position on those issues. Equally important, as suggested in *CLM* 32, this effort is perhaps also motivated by a belief that the United States and other powers are increasingly challenging some of China's core interests, thus requiring a more assertive PRC response. In addition, Beijing's apparent refusal to "haggle" or compromise, and its stated willingness to employ extreme measures—including force—to defend its position with regard to China's core interests, arguably constitute a warning to other nations that should not be ignored.

Of course, every nation has its national interests, many of which are described as "vital" or "core."⁵⁶ China is obviously no exception. Nonetheless, Beijing's explicit and growing emphasis on the term, its adoption of a seemingly rigid negotiating stance on core interests, the application of the phrase to contentious issues such as Taiwan, and, perhaps most importantly, the possibility that a stronger China might expand the scope and sharpen the definition of its core interests further to include other issues of contention, together pose a significant challenge to U.S. (and Chinese) efforts to maintain a stable and mutually productive bilateral relationship.

Notes

¹ The Joint Statement included the following phrase: “The two sides agreed that respecting each other’s core interests is extremely important to ensure steady progress in China-US relations.” See “U.S.-China Joint Statement,” White House, Office of the Press Secretary, November 17, 2009, Beijing, China, available at <http://www.whitehouse.gov/the-press-office/us-china-joint-statement>; and “China-US Joint Statement,” November 17, 2009, Beijing, China, available at <http://www.fmprc.gov.cn/eng/wjzb/zjzg/bmdyys/xwlb/t629497.htm>. The language about core interests was absent from the January 2011 joint statement issued during President Hu Jintao’s state visit to Washington, as discussed further below. U.S.-China Joint Statement, January 19, 2011, Washington, D.C., available at <http://www.whitehouse.gov/the-press-office/2011/01/19/us-china-joint-statement>.

² The major primary sources employed in this study to chart the official use of the concept of “core interests” and related terms include: The official website of the Ministry of Foreign Affairs of the People’s Republic of China (中华人民共和国外交部), Chinese and English versions (<http://www.fmprc.gov.cn> and <http://www.fmprc.gov.cn/eng>, respectively); the archives of *People’s Daily* (人民日报), at PeopleData (<http://data.people.com.cn>, 人民数据库: 中国政府文献信息); the archives of *PLA Daily* (解放军报) at East View Information Services (<http://www.eastview.com>); and the databases of the Chinese Government and the Communist Party of China (CPC), both at PeopleData. We are also grateful to Professor Alastair Iain Johnston of Harvard University for providing his unpublished data on the PRC usage of the term “core interests.”

³ For example, *People’s Daily* apparently first employed the term “core interests” (核心利益) in June 1980, in discussing how the Soviet invasion of Afghanistan and Soviet support of Vietnam’s invasion of Cambodia threatened the “core interests” of the West (西方). See 象乾; 肖西, “不容漠视的绥靖暗流,” 人民日报 (Xiang Qian and Xiao Xi, “Critical Appeasement Undercurrent,” *People’s Daily*), June 21, 1980, <http://data.people.com.cn>.

The second reference to the phrase in the *People’s Daily* database discusses Vice President Al Gore’s five-part formulation of America’s national interests and describes how Gore explained that a technology policy would serve the U.S. “core interests.” See 徐勇, “美政府发表科技政策声明, 发展重点由军事技术转向民用技术,” 人民日报 (Xu Yong, “U.S. government declaration of scientific policy states the transfer of military technology to civilian use,” *People’s Daily*), August 6, 1994, 6th edition, <http://data.people.com.cn>.

For similar usages, see 黄泽全, “科技生财,” 人民日报 (Huang Zequang, “Technology Creates Wealth,” *People’s Daily*), July 14, 1995, 7th edition, <http://data.people.com.cn>; “美财长发表讲话指出, 美中关系事关全球经济繁荣,” 人民日报 (“U.S. Treasury Secretary indicates in his speech that U.S.-China relations are critical to global economic prosperity,” *People’s Daily*), October 14, 1999, 6th edition, available at <http://data.people.com.cn>; and 欧灿, “警告李登辉不要抱任何幻想,” 解放军报 (Ou Can, “Warning Lee Teng-hui not to have any illusions,” *PLA Daily*), August 29, 1999, available at <http://dlib.eastview.com/browse/doc/14557759>.

⁴ For example, see “驻美国大使馆召开胡锦涛总书记重要讲话座谈,” 中华人民共和国外交部, (“U.S. Embassy held important talks with General Secretary Hu Jintao,” Ministry of Foreign Affairs of the People’s Republic of China), January 11, 2009, available at <http://www.mfa.gov.cn/chn/gxh/tyb/zwbdt531443.htm>: “周大使 . . . 指出 . . . 把握民族根本利益和国家核心利益, 体现了尊重历史、尊重现实、尊重人民愿望的实事求是精神。” (“Ambassador Zhou . . . stated that . . . grasping fundamental interests and core national interests reflects the truth-seeking spirit of respect for history, respect for reality, and respect for the wishes of the people.”)

A 1995 article from the *PLA Daily* specifically discusses the adjustment from a focus on “fundamental interests” to “core interests” in the process of economic and social reform. However, it does not clearly define the difference between the two terms. Based on a reading of the article, it is possible that the former referred to the basic interest of promoting overall national economic development while the latter was intended to focus on the attainment of greater economic and social equality and common prosperity. Nonetheless, both concepts were at the time applied to domestic issues. See “图片 / 照片 / 其它,” 解放军报 (“Picture/Photo/Other,” *PLA Daily*), February 17, 1995, available at <http://dlib.eastview.com/browse/doc/14466790>. For a similar usage, also see “《求是》杂志发表署名文章《划清‘四个重大界限’的有关理论与实践问题》,” 解放军报 (“*Qiushi* published a signed article by Hua Qing: ‘Four Major Boundaries of Theoretical and Practical Issues,’” *PLA Daily*), August 17, 2001, available at <http://dlib.eastview.com/browse/doc/22327893>.

⁵ Apparently, the first time the two terms were used together (重大关切和核心利益) was in 2007, in a report on a meeting between Hu Jintao and German president Angela Merkel. See “胡锦涛会见德国总理默克尔,” 中华人民共和国外交部 (“Hu Jintao meets German Prime Minister Merkel,” Ministry of Foreign Affairs of the People’s Republic of China), August 27, 2007, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t355870.htm>.

⁶ The first reference to “major concerns” on the PRC Ministry of Foreign Affairs website is from 2000, in a report on a meeting between Jiang Zemin and President Chirac of France. See “江泽民主席与希拉克总统举行正式会,” 中华人民共和国外交部 (“Chairman Jiang Zemin and French President Chirac held meeting,” Ministry of Foreign Affairs of the People’s Republic of China), November 7, 2000, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t7275.htm>.

“中法之间有许多共同利益, 合作前景十分广阔。只要双方从战略的高度和两国关系长远发展的大局出发, 坚持相互尊重、平等互利, 特别是相互尊重对方的重大关切, 中法关系在 21 世纪里必将会提高到一个新的水平。” (“China and France share many common interests; the prospect for bilateral cooperation is strong. As long as the two countries start from the strategic perspective and the pattern of long-term development of the bilateral relationship, maintain mutual respect, equality, and mutual benefits, especially respecting each other’s major concerns, Sino-French relations can be raised to a new level in the 21st century.”) In a 2003 Policy Paper on the European Union, “major concerns” is used to directly refer to the Taiwan issue in a paragraph on the one-China policy. See “中国对欧盟政策文件,” 中华人民共和国外交部, 二〇〇三年十月 (“White Paper on Chinese policy toward the EU,” Ministry of Foreign Affairs of the People’s Republic of China, October 2003), available at <http://www.fmprc.gov.cn/chn/pds/ziliao/tytj/zcwj/t27700.htm>.

⁷ Wang Jisi, a well-known scholar of U.S.-China relations, wrote an article in *People’s Daily* in February 2002 suggesting that both sides seek common ground in order to prevent differences from developing into serious crises or conflicts that would damage the “core interests” (核心利益) of either party. This is the first time that a *People’s Daily* article referred to the “core interests” of China (in a foreign policy context) and not only the “core interests” of other countries. See 王缉思, “求同存异 稳定大局 (专论)—纪念中美《上海公报》发表 30 周年,” 人民日报 (Wang Jisi, “Seek common ground and stability (monograph)—celebrating the 30th anniversary of the China-U.S. ‘Shanghai Communiqué’,” *People’s Daily*), February 28, 2002, 7th edition, available at <http://data.people.com.cn>).

A second similar reference in the same source occurred in December 2002. See 尹承德, “大国关系的改善与演变,” 人民日报 (Yin Chengde, “Progress and Transformation of Great Power Relations,” *People’s Daily*), November 19, 2002, 7th edition, available at <http://data.people.com.cn>. In this article, Yin, a regular contributor to *People’s Daily* and possibly a Foreign Ministry official, referred to China in discussing the “core interests” of great powers.

⁸ See 丁刚, “唐家璇会见美国务卿,” 人民日报 (Ding Gang, “Tang Jiaxuan Meets U.S. Secretary of State,” *People’s Daily*), January 21, 2003, 3rd edition, available at <http://data.people.com.cn>. At that time, “Tang Jiaxuan said, the Taiwan issue concerns China’s core interests” (“唐家璇说, 台湾问题事关中国的核心利益.”) A month later, Colin Powell visited China and again held talks with Tang Jiaxuan, who again reiterated the point. See “Foreign Minister Tang Jiaxuan Held Talks With Powell,” Ministry of Foreign Affairs of the People’s Republic of China, February 25, 2003, available at <http://www.fmprc.gov.cn/eng/zxxx/t19583.htm>.

⁹ See “外交部发言人孔泉就美决定向台湾出售远程预警雷达系统答记者问,” 中华人民共和国外交部 (Foreign Ministry spokesperson Kong Quan warns the U.S. for its decision to sell Taiwan the long-range radar system at a press conference), January 1, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/fyrbt/dhdw/t82122.htm>.

¹⁰ Most notably, on the official level, such usage occurred at several press conferences held by Foreign Ministry spokespersons Zhang Qiyue and Kong Quan, in speeches by such senior PRC diplomats as Yang Jiechi, Zhang Yesui, and Zhong Jianhua, in a meeting between Tang Jiaxuan and then-National Security Advisor Condoleezza Rice in July 2004, in meetings between Li Zhaoxing and Colin Powell in October 2004, by Wen Jiabao in a speech at the October 2004 Asia-Europe Meeting and in remarks to Prime Minister Shaukat Aziz of Pakistan in December 2004, and in remarks by Hu Jintao at meetings with President Bush and Australian prime minister John Howard in November 2004. “2004 年 2 月 19 日外交部发言人章启月向记者招待会上答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Zhang Qiyue’s regular press conference on February 19, 2004,” Ministry of Foreign Affairs of the People’s Republic of China), February 19, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/fyrbt/dhdw/t66764.htm>;

“2004年4月1日外交部发言人孔泉在例行记者会上答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Kong Quan’s regular press conference on February 19, 2004,” Ministry of Foreign Affairs of the People’s Republic of China), April 1, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/fyrbt/jzhsl/t82128.htm>; “杨洁篪大使访问洛杉矶并接受杰出外交官奖,” 驻洛杉矶总领馆供稿 (“Ambassador Yang Jiechi visits Los Angeles and receives the award for Distinguished Diplomat,” Los Angeles General Consulate press release), April 5, 2004, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t441607.htm; “2004年4月8日外交部发言人孔泉在记者招待会上答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Kong Quan’s Regular Press Conference on April 8, 2004,” Ministry of Foreign Affairs of the People’s Republic of China), April 8, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/fyrbt/dhdw/t83827.htm>; “2004年4月27日外交部发言人孔泉在记者招待会上答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Kong Quan’s regular press conference on April 27, 2004,” Ministry of Foreign Affairs of the People’s Republic of China), April 27, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/fyrbt/t93544.htm>; “杨洁篪大使赴美阿肯色州就当前中美关系发表演讲 (10/05/04),” 驻美国使馆供稿 (“Yang Jiechi, Ambassador to the United States, delivered a speech on current Sino-U.S. relations in Arkansas,” press release of the Chinese Embassy to the United States), May 10, 2004, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t431003.htm; “唐家璇会见赖斯,” 中华人民共和国外交部 (“Tang Jiakuan Meets with Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), July 9, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/wzb/zxxx/t142558.htm>; “张业遂副部长在第五届中美军控、裁军与防扩散研讨会上的讲话,” 二〇〇四年七月二十日, 北京 (“Vice Minister Zhang Yesui’s speech at the Fifth China-U.S. arms control, disarmament, and nonproliferation conference,” July 20, 2004, Beijing), available at <http://www.fmprc.gov.cn/chn/gxh/wzb/wjbxw/t143524.htm>; “外交部长李肇星会晤美国务卿鲍威尔,” 中华人民共和国外交部 (“Foreign Minister Li Zhaoxing meets with U.S. secretary of state Colin Powell,” Ministry of Foreign Affairs of the People’s Republic of China), October 11, 2004, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzq/mzblwelm/xgxw/t163077.htm>; 温家宝, 中华人民共和国国务院总理, “加强对话合作, 深化伙伴关系,” 在第五届亚欧首脑会议上的讲话, 二〇〇四年十月八日, 河内, (“Chinese Premier Wen Jiabao, ‘Strengthen dialogue and cooperation, deepen the partnership,’ speech at the fifth ASEM summit, Hanoi, October 8, 2004), available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t163675.htm>; “胡锦涛主席会见美国总统布什,” 中华人民共和国外交部 (“President Hu Jintao meets U.S. President Bush,” Ministry of Foreign Affairs of the People’s Republic of China), November 21, 2004, available at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/ywzt/zt2004/hjtlatinamerica/t171122.htm>; “胡锦涛主席会见澳大利亚总理霍华德,” 中华人民共和国外交部 (“President Hu Jintao meets with Australian prime minister John Howard,” Ministry of Foreign Affairs of the People’s Republic of China), November 22, 2004, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/dyz/1206/xgxw/t171174.htm>; “驻洛杉矶总领事钟建华在青岛-长滩友协发表演讲,” 驻洛杉矶总领馆供稿 (“Zhong Jianhua, Los Angeles consul general gives speech to Long Beach-China Friendship Association in Qingdao,” Los Angeles Consulate General Press Release), December 3, 2004, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t443959.htm; “温家宝总理与巴基斯坦总理阿齐兹举行会谈,” 中华人民共和国外交部 (“Premier Wen Jiabao holds meeting with Pakistani prime minister Shaukat Aziz,” Ministry of Foreign Affairs of the People’s Republic of China), December 16, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/gjldrhd/t175152.htm>.

In that same year, China also highlighted its “core interests” in several press conferences and speeches vis-à-vis Singapore, reacting against Deputy Prime Minister Lee Hsien Loong’s visit to Taiwan in July 2004. “外交部发言人章启月就新加坡副总理李显龙访台发表谈话,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Zhang Qiyue’s remarks on the visit of Singapore Deputy Prime Minister Lee Hsien Loong,” Ministry of Foreign Affairs of the People’s Republic of China), July 11, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/wzb/fyrbt/t142666.htm>; “2004年7月13日外交部发言人章启月在例行记者会上答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Zhang Qiyue’s regular press conference on July 13, 2004,” Ministry of Foreign Affairs of the People’s Republic of China), July 13, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/fyrbt/jzhsl/t142922.htm>; “外交部发言人孔泉就新加坡副总理李显龙近日就其访台一事接受新加坡媒体书面采访答记者问,” 中华人民共和国外交部 (“Foreign Affairs Ministry Spokesman Kong Quan’s statement on Singapore deputy prime minister Lee Hsien Loong’s recent visit to Taiwan with the Singapore media,” Ministry of Foreign Affairs of the People’s Republic of China), July 20, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/fyrbt/t143616.htm>; “外交部副部长助理沈国放 2004年7月22

日网上答公众问全文,” 括号内为网友登录“中国外交论坛”时注册的昵称, 中华人民共和国外交部 (“Assistant Foreign Minister Shen Guofang puts the full text answer online on ‘Chinese Foreign Ministry Forum,’ Ministry of Foreign Affairs of the People’s Republic of China, July 22, 2004, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t143741.htm>; “胡锦涛主席会见新加坡总理李显龙,” 中华人民共和国外交部 (“President Hu Jintao Meets Singapore prime minister Lee Hsien Loong,” Ministry of Foreign Affairs of the People’s Republic of China), November 20, 2004, available at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/ywzt/2005year/hjzt/hjt2004apec/t170958.htm>; “驻新加坡大使张云谈孙中山精神,” 驻新加坡使馆 (“Ambassador to Singapore Zhang Yun discusses the spirit of Sun Yat-sen,” Embassy in Singapore), November 28, 2004, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwbdt/t172364.htm>. The term was also used by Wu Hongbo, ambassador to the Philippines: “吴红波大使出席让德吴氏宗亲总会全菲联谊大会,” 驻菲律宾使馆 (“Ambassador Wu Hongbo attends the Filipino-Chinese Grand Family Association sponsored by Rang Dewu,” Embassy in the Philippines, October 19, 2004, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwbdt/t166879.htm>. Such references to core interests continued in 2005 and beyond.

¹¹ Li stated in a speech that sovereignty and security are Chinese core interests “涉及国家主权、安全等核心利益。” See 李肇星, “新时期外交工作的宝贵精神财富——学习江泽民同志外交思想的体会,” 人民日报 (Li Zhaoxing, “Valuable spiritual wealth of new China’s diplomatic work—studying Jiang Zemin’s diplomatic thought,” *People’s Daily*), September 30, 2006, 2nd edition, available at <http://data.people.com.cn>.

The first reference on the PRC Ministry of Foreign Affairs website to “security” (安全) as a “core interest” was in a December 2008 Xinhua interview with China’s ambassador to the EU, Song Zhe. See “驻欧盟使团团长宋哲大使接受新华社《参考消息》年终专访,” 驻欧盟使团供稿. (“Mission to the EU, Ambassador ‘Song Zhe accepts Xinhua News Agency’s ‘Reference News’ year-end interview,” Mission to the EU Press Release), December 26, 2008, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zwbdt/wshd/t529299.htm>.

Hu Jintao also referred to “national security” (国家安全) as a “core interest” in a statement commenting on Sino-Russian bilateral relations, in June 2009: “国家主席胡锦涛接受中外媒体联合采访,” 中华人民共和国外交部 (“President Hu Jintao accepts the joint interview from foreign media,” Ministry of Foreign Affairs of the People’s Republic of China), July 18, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t568406.htm>.

¹² The first reference on the PRC Foreign Ministry website to respecting each other’s core interests is in a speech Yang Jiechi gave as ambassador to the United States in May 2004 in Arkansas. See “杨洁篪大使赴美阿肯色州就当前中美关系发表演讲 (10/05/04),” 驻美国使馆供稿 (“Yang Jiechi, ambassador to the United States, delivered a speech on current Sino-U.S. relations in Arkansas,” press release of the Chinese Embassy to the United States), May 10, 2004, available at

http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t431003.htm. The next reference occurred later that same year in remarks from Hu Jintao during a November 2004 meeting with Australian prime minister John Howard. See “胡锦涛主席会见澳大利亚总理霍华德,” 中华人民共和国外交部 (“President Hu Jintao meets with Australian prime minister John Howard,” Ministry of Foreign Affairs of the People’s Republic of China), November 22, 2004, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/dyz/1206/xgxw/t171174.htm>. Also see the record of a Sino-Peruvian meeting between Wen Jiabao and Peru’s president Alejandro Toledo in June 2005 (“吴邦国、温家宝分别会见秘鲁总统 托莱多,” 中华人民共和国外交部 [“Wu Bangguo, Wen Jiabao each met with Peruvian President Alejandro Toledo,” Ministry of Foreign Affairs of the People’s Republic of China], June 3, 2005, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t198568.htm>), and remarks given by Hu Jintao in a meeting with Vladimir Putin of Russia in March 2006 (“胡锦涛与俄罗斯联邦总统普京举行会谈,” 中华人民共和国外交部 [“Hu Jintao holds talk with Russian president Vladimir Putin,” Ministry of Foreign Affairs of the People’s Republic of China], March 21, 2006, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/oz/1206_13/xgxw/t241482.htm).

For more recent examples see “Chinese President Hu Jintao Meets with His U.S. Counterpart Obama,” Ministry of Foreign Affairs of the People’s Republic of China, April 2, 2009, available at <http://www.fmprc.gov.cn/eng/wjdt/wshd/t555783.htm>; “首轮中美经济对话: 除上月球外主要问题均已谈及,” 中国新闻网 (“First round of the Sino-U.S. Economic Dialogue: In addition to the main discussion of the Moon,” China News Online), July 29, 2009, available at <http://www.chinanews.com.cn/gn/news/2009/07-29/1794984.shtml>; “U.S. Defense Secretary Invited To Visit China Early Next Year,” Xinhua, October 11,

2010, Open Source Center (OSC) CPP20101011968207; President Hu Jintao, Remarks at Strategic and Economic Dialogue Opening Session, May 23, 2010 (China date: May 24, 2010), available at <http://www.state.gov/secretary/rm/2010/05/142134.htm>; and State Councilor Dai Bingguo, Remarks at Strategic and Economic Dialogue, “Concluding Joint Statements by Chinese Vice Premier Wang Qishan, Chinese State Councilor Dai Bingguo, Secretary of State Hillary Rodham Clinton, and Secretary of Treasury Timothy Geithner at the U.S.-China Strategic and Economic Dialogue (as released by the State Department),” Federal News Service, May 25, 2010.

¹³ See “首轮中美经济对话：除上月球外主要问题均已谈及，”中国新闻网 (“First round of the Sino-U.S. Economic Dialogue: In addition to the main discussion of the Moon,” China News Online), July 29, 2009, available at <http://www.chinanews.com.cn/gn/news/2009/07-29/1794984.shtml>. Also see State Councilor Dai Bingguo, Closing Remarks for U.S.-China Strategic and Economic Dialogue, Washington, DC, July 28, 2009, available at <http://www.state.gov/secretary/rm/2009a/july/126599.htm>.

Dai’s remarks were to some extent presaged (though not explicitly spelled out) by Foreign Minister Yang Jiechi, in remarks delivered in Washington, DC, on March 12, 2009. At that time, Yang stated: “Mutual respect means that we should respect international law and norms governing international relations, view each other in an objective and sensible way, respect each other’s choice of development path and *core interests*, and refrain from doing anything that may harm each other’s *sovereignty, security or development interests*.” Yang Jiechi, “Broaden China-U.S. Cooperation in the 21st Century,” Speech at the Center for Strategic and International Studies, Washington D.C., March 12, 2009, available at <http://www.china-embassy.org/eng/zmgx/t542231.htm>, emphasis added.

¹⁴ For example, see “Foreign Ministry Spokesperson Qin Gang’s Regular Press Conference on July 13, 2010,” Ministry of Foreign Affairs of the People’s Republic of China, July 14, 2010, available at <http://www.mfa.gov.cn/eng/xwfw/s2510/2511/t716745.htm>.

Da Wei, deputy director of the Department of American Studies at the China Institute of Contemporary International Relations [CICIR], cited Dai Bingguo’s definition in his discussion of how China should strategically utilize the term “core interests”; see Da Wei, “Why Should China Declare Its Core Interests?” *Huanqiu Shibao*, July 28, 2010, OSC CPP20100729788013.

Jin Canrong, vice director of the School of International Studies at China’s Renmin University, also discussed Dai’s definition, explaining that the phrase “basic system” referred to the system “of multiparty cooperation and political consultation led by the Communist Party of China.” See “China Denies Taking Tough Stance on International Affairs,” *Global Times Online*, March 8, 2010, OSC CPP20100308722006.

For other unofficial references to Dai Bingguo’s multi-part definition, see Wu Zhong, “China Cut To the Core,” *Asia Times*, August 19, 2009, OSC CPP20090819715002; and Cary Huang, “A Bolder China Asserts ‘Core’ Interests But Will It Act?” *South China Morning Post*, August 12, 2010, OSC CPP20100812715004.

¹⁵ See “驻新加坡大使张云谈孙中山精神，”驻新加坡使馆 (“Ambassador to Singapore Zhang Yun discusses the spirit of Sun Yat-sen,” Embassy in Singapore), November 28, 2004, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwbdt/172364.htm>.

This is the earliest reference to “national unity” (国家统一) or “reunification” (两岸统一) as core interests of China found on the PRC Foreign Ministry website. For similar later references, see “曾庆红会见匈牙利总理久尔查尼，”中华人民共和国外交部 (“Zeng Qinghong meets with Hungarian prime minister Gyurcsany, Ministry of Foreign Affairs of the People’s Republic of China), September 8, 2005, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t210951.htm>; “中国驻洛杉矶总领事张云出席南加州侨界反对台湾‘入联公投’大会，”驻洛杉矶总领馆 (“Consul General Zhang Yun in Los Angeles attends Southern California Overseas Chinese Bridge and opposes Taiwan’s ‘referendum to join the UN General Assembly’,” Consulate General in Los Angeles), September 16, 2007, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwbdt/441207.htm>; and “Getting to Know the Real China,” Statement by H.E. Wen Jiabao, Premier of the State Council of the People’s Republic of China, At the General Debate of the 65th Session of the UN General Assembly, New York, September 23, 2010, available at <http://www.fmprc.gov.cn/eng/topics/wenjibaochuxi65jieUNdh/t761353.htm>.

¹⁶ See “纳米比亚议会和政府支持我通过《反分裂国家法》，”中华人民共和国外交部 (“Namibian parliament and government to support the implementation of the ‘Anti-Secession Law,’” Ministry of Foreign Affairs of the People’s Republic of China), March 18, 2005, available at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/ywzt/2005year/fflgjf/t442120.htm>. This is the first reference to the phrase “independence” (独立) in the context of “core interests.” Also see “国家主席胡锦涛接受中外媒体联合

采访,” 中华人民共和国外交部 (“President Hu Jintao accepts joint interview from foreign media,” Ministry of Foreign Affairs of the People’s Republic of China), July 18, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t568406.htm>; here Hu Jintao referred to “independence” (独立) as a “core interest” in an interview commenting on Sino-Russian bilateral relations.

¹⁷ “Hu Jintao Holds Talks with General Secretary of the Central Committee of the Lao People’s Revolutionary Party (PRP) and Lao President Choummaly,” Embassy of the People’s Republic of China in the Republic of Lithuania, November 19, 2006, available at <http://lt.china-embassy.org/eng/xwtd/t281353.htm>; “Carry On Traditional Friendship And Deepen All-round Cooperation—Address by Hu Jintao, President of the People’s Republic of China, At Islamabad Convention Center, SRC-406,” November 24, 2006, available at <http://www.fmprc.gov.cn/eng/wjb/zzjg/yzs/gjlb/2757/2758/t285917.htm>. See also, “Hu Jintao Holds Talks with US President Obama,” Ministry of Foreign Affairs of the People’s Republic of China, November 17, 2009, available at <http://www.fmprc.gov.cn/eng/wjb/zzjg/bmdyys/xwlb/t628187.htm>, where Hu mentions “human rights” in the context of “core interests,” but does not explicitly label them in that category.

¹⁸ “杨洁篪会见马尔代夫外长沙希德,” 中华人民共和国外交部 (“Maldives Foreign Minister Shahid meets with Yang Jiechi,” Ministry of Foreign Affairs of the People’s Republic of China), May 26, 2008, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t458721.htm>; “杨洁篪部长接受巴基斯坦联合通讯社书面采访,” 中华人民共和国外交部 (“Foreign Minister Yang Jiechi accepts a written interview with the Associated Press of Pakistan,” Ministry of Foreign Affairs of the People’s Republic of China), April 25, 2008, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/wjbxw/t429027.htm>; “和谐相处 合作共赢—周文重大使就中美建交 30 周年为《人民日报》撰稿,” 驻美国使馆供稿 (“Harmonious and Win-win Cooperation—Ambassador Zhou Wenzhong at the 30th Anniversary of Sino-U.S. relations, writes ‘People’s Daily.’” U.S. Embassy Press Release, January 1, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/wzb/zwb/wshd/t530030.htm>; “驻欧盟使团团长宋哲大使接受新华社《参考消息》年终专访,” 驻欧盟使团供稿 (“Mission to the EU, Ambassador Song Zhe, head of the Xinhua News Agency accepts the end-of-year interview with ‘Cankao Xiaoxi,’” Mission to the EU Press Release), December 26, 2008, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zwb/wshd/t529299.htm>; “巴基斯坦成立首个巴中研究所,” 驻巴基斯坦使馆供稿 (“Pakistan establishes the first China-Pakistan Research Institute,” Embassy in Pakistan Press Release), October 2, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/cgb/zwb/wshd/t617868.htm>; “驻巴基斯坦大使罗照辉在伊斯兰堡发表演讲,” 驻巴基斯坦使馆供稿 (“Ambassador to Pakistan Luo Zhaohui gives speech at the World Affairs Council in Islamabad,” Embassy in Pakistan Press Release), May 23, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/cgb/zwb/wshd/t564112.htm>; “中国驻尼泊尔大使孙和平出席尼副首相兼 外交大臣奥利钱行晚宴,” 驻尼泊尔使馆供稿 (“Chinese Ambassador to Nepal Sun Heping attends the farewell dinner of Nepalese Deputy Prime Minister and Foreign Minister Oli,” Embassy in Nepal Press Release), March 7, 2007, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/yz/1206_24/1206x2/t301747.htm.

¹⁹ See “外交部发言人章启月就美国国务院表示陈水扁 ‘十点讲话’ 有积极内容答记者问,” 中华人民共和国外交部 (“Foreign Ministry Spokesperson Zhang Qiyue’s remarks on the positive elements in the U.S. State Department comments on Chen Shui-bian’s ‘ten-point statement,’” Ministry of Foreign Affairs of the People’s Republic of China), November 16, 2004, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/fyrbt/t170288.htm>; “李肇星与美国务卿举行会谈,” 中华人民共和国外交部 (“Li Zhaoxing held talks with U.S. Secretary of State,” Ministry of Foreign Affairs of the People’s Republic of China), October 25, 2004, available at

<http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t166903.htm>; “驻新加坡大使张云谈孙中山精神” 驻新加坡使馆 (“Ambassador to Singapore Zhang Yun discusses the spirit of Sun Yat-sen,” Embassy in Singapore), November 28, 2004, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwb/t172364.htm>.

²⁰ The ASL does not explicitly use the term “core interests.” However, it does state that “China’s sovereignty and territorial integrity ‘brook no division’ (不容分割). And “sovereignty and territorial integrity” (主权和领土完整) had been officially identified as Chinese core interests as early as April 2004, as indicated above. See “Anti-Secession Law, Adopted at the Third Session of the Tenth National People’s Congress on March 14, 2005,” available at <http://www.china.org.cn/english/2005lh/122724.htm>. For the original Chinese text, see “反分裂国家法,” 新华网 (“Anti-Secession Law,” Xinhuanet), March 14, 2005, available at http://news.xinhuanet.com/newscenter/2005-03/14/content_2694168.htm.

Moreover, Hu Jintao delivered a speech in the run-up to the passage of the Anti-Secession Law wherein he spelled out “Four Nevers” vis-à-vis Taiwan and declared that “safeguarding national sovereignty and territorial integrity is where a country’s core interest lies.” “President Hu sets forth guidelines on Taiwan,” Embassy of the People’s Republic of China in the Republic of Hungary, March 5, 2005, available at <http://www.chinaembassy.hu/hu/xwdt/t185890.htm>.

²¹ See Cary Huang, “A Bolder China Asserts ‘Core’ Interests But Will It Act?” *South China Morning Post*, August 12, 2010, OSC CPP20100812715004. This article cites Professor Jin Canrong, associate dean of Renmin University’s School of International Relations, as stating: “There is consensus within Chinese diplomatic circles and think-tank scholars that the term [core interests] will apply to only two categories—territorial integrity and sovereignty—at least for the moment.”

²² For recent examples, see Liao Lei and Hou Lijun, “(Foreign Affairs) Chinese Foreign Ministry Spokesman Zou Gang: We Expect More Positive Results from Obama’s China Visit,” *Xinhua*, November 10, 2009, OSC CPP20091110062018. The article states, “In response to a question about whether a specific statement will be made during this visit about meddling in Taiwan and Tibet, [Ministry of Foreign Affairs spokesperson] Qin Gang said that clear stipulations and promises are set out to address these issues in the three Sino-US joint communiqués and this is also an important political foundation for forward development of the Sino-US relationship. *To allow the Sino-US relations to move forward in a healthy and stable manner*, the United States must abide by the three joint communiqués and respect the core interests and major areas of concern on both sides” (emphasis added). Also see “Hu Jintao Holds Talks with US President Obama,” Ministry of Foreign Affairs of the People’s Republic of China, November 17, 2009, available at <http://www.fmprc.gov.cn/eng/wjb/zjzg/bmdyys/xwlb/t628187.htm>. This official PRC account of the Obama-Hu talks in November 2009 reported Hu’s stating that “to respect the core interest and major concern of each other” was necessary for strategic mutual trust.

For similar statements by senior Chinese officials made during the George W. Bush era, see “杨洁篪大使赴美阿肯色州就当前中美关系发表演讲 (10/05/04),” 驻美国使馆供稿 (“Yang Jiechi, ambassador to the United States, delivered a speech on current Sino-U.S. relations in Arkansas,” press release of the Chinese Embassy to the United States), May 10, 2004, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t431003.htm; “杨洁篪外长同美国国务卿赖斯会谈,” 中华人民共和国外交部 (“Foreign Minister Yang Jiechi talks with U.S. secretary of state Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), September 24, 2007, available at <http://www.fmprc.gov.cn/chn/pds/wjb/wjbz/zyhd/t365879.htm>; “中国驻洛杉矶总领事张云出席 ‘中美关系与中国和平统一论坛’ 开幕式,” 驻洛杉矶总领馆供稿 (“Consul General Zhang Yun attends the opening ceremony of the Forum on China-U.S. relations and on China’s Peaceful Reunification,” Consulate General in Los Angeles), November 25, 2007, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t434377.htm; and “国务委员戴秉国会见美国国务卿赖斯,” 中华人民共和国外交部 (“State Councilor Dai Bingguo meets U.S. secretary of state Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), June 30, 2008, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t470540.htm>.

²³ For examples, see “国务委员戴秉国会见美国国务卿赖斯,” 中华人民共和国外交部 (“State Councilor Dai Bingguo meets U.S. secretary of state Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), June 30, 2008, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t470540.htm>; “杨洁篪外长同美国国务卿赖斯会谈,” 中华人民共和国外交部 (“Foreign Minister Yang Jiechi talks with U.S. secretary of state Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), September 24, 2007, available at <http://www.fmprc.gov.cn/chn/pds/wjb/wjbz/zyhd/t365879.htm>; “胡锦涛主席会见美国总统布什,” 中华人民共和国外交部 (“President Hu Jintao meets U.S. president Bush,” Ministry of Foreign Affairs of the People’s Republic of China), November 21, 2004, available at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/ywzt/zt2004/hjlatinamerica/t171122.htm>; “外交部长李肇星会晤美国国务卿鲍威尔,” 中华人民共和国外交部 (“Foreign Minister Li Zhaoxing meets U.S. secretary of state Powell,” Ministry of Foreign Affairs of the People’s Republic of China), October 1, 2004, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t163077.htm>; “唐家璇会见赖斯,” 中华人民共和国外交部 (“Tang Jiaxuan Meets with Condoleezza Rice,” Ministry of Foreign Affairs of the People’s Republic of China), July 9, 2004, <http://www.fmprc.gov.cn/chn/gxh/wzb/zxxx/t142558.htm>; 丁刚, “唐家璇会见美国国务卿,” 人民日报 (Ding Gang, “Tang Jiaxuan Meets U.S. Secretary of State,” *People’s Daily*), January

21, 2003, 3rd edition, available at <http://data.people.com.cn>; and “Foreign Minister Tang Jiaxuan Held Talks With Powell,” Ministry of Foreign Affairs of the People’s Republic of China, February 25, 2003, available at <http://www.fmprc.gov.cn/eng/zxxx/t19583.htm>.

A PRC Ministry of Foreign Affairs report states that Obama “urged both sides to respect each other’s core interests” during the Hu-Obama meeting of April 2009 on the sidelines of the G20 Financial Summit in London. However, this report was not an official statement, and the official U.S. statement on the meeting found on the White House website does not mention any reference by Obama to “core interests.” For the Chinese Foreign Ministry report, see “Chinese President Hu Jintao Meets with His U.S. Counterpart Obama,” Ministry of Foreign Affairs of the People’s Republic of China, April 2, 2009, available at <http://www.fmprc.gov.cn/eng/wjdt/wshd/t555783.htm>. The Chinese version is available at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/dnzt/hjtzxdecfhsg/t555457.htm>. For the White House report, see “Statement On Bilateral Meeting With President Hu Of China,” The White House, Office of the Press Secretary, April 1, 2009, available at <http://www.whitehouse.gov/the-press-office/statement-bilateral-meeting-with-president-hu-china>.

²⁴ See Joint Communiqué of the People’s Republic of China and the United States of America, Issued in Shanghai, February 28, 1972, available at http://beijing.usembassy-china.org.cn/uploads/images/yJIG2IXkWU8gk1uq4Cx1tw/1972_Joint_Communique.pdf; and U.S.-PRC Joint Communiqué, August 17, 1982, available at http://beijing.usembassy-china.org.cn/uploads/images/QjyEUD4z2TG3AW9PIy5tRA/1982_Joint_Communique.pdf.

²⁵ See “外交部副部长何亚非就美国售台武器问题接受专访,” 中华人民共和国外交部 (“Vice Foreign Minister He Yafei answers questions on U.S. arms sales to Taiwan in an interview,” Ministry of Foreign Affairs of the People’s Republic of China), January 9, 2010, available at <http://www.fmprc.gov.cn/chn/pds/wjb/zygy/t650637.htm>; “U.S. Arms Sales To Taiwan Detrimental To Sino-U.S. Relations,” Commentary, Xinhua, January 11, 2010, OSC CPP20100111968162; “Foreign Ministry Spokesperson Ma Zhaoxu’s Regular Press Conference on January 26, 2010,” Ministry of Foreign Affairs of the People’s Republic of China, January 27, 2010, <http://www.fmprc.gov.cn/eng/xwfw/s2510/2511/t654019.htm>; “Foreign Ministry Spokesperson Ma Zhaoxu’s Regular Press Conference on February 4, 2010,” Ministry of Foreign Affairs of the People’s Republic of China, February 5, 2010, available at <http://www.fmprc.gov.cn/eng/xwfw/s2510/2511/t656333.htm>; “常驻联合国代表团与驻纽约总领馆联合举行新春招待会,” 驻纽约总领馆供稿 (“Permanent Mission to the United Nations and New York Consulate General held a joint spring conference,” Consulate General in New York Press Release), February 4, 2010, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/zwbj/jghd/t655935.htm>; “The Embassy Hosted Farewell Reception for Ambassador Zhou Wenzhong And Mme. Xie Shumin,” Embassy of the People’s Republic of China in the United States, February 19, 2010, available at <http://www.china-embassy.org/eng/zmgx/t659074.htm>; “FM: China Firmly Opposed To U.S. Moves That Undermine Ties,” Xinhua, March 7, 2010, OSC CPP20100307968053; “Premier Puts Forward Proposals for Closer China-U.S. Ties,” Xinhua, September 23, 2010, OSC CPP20100923968145; “Foreign Ministry urges U.S. to respect China’s core interests with credible steps,” Xinhua, March 7, 2010, available at <http://english.people.com.cn/90001/90776/90883/6911113.html>; and “China, U.S. agree to respect ‘core interests’: diplomat,” Reuters, April 6, 2010, available at <http://www.reuters.com/article/idUSTRE6360FT20100407>.

²⁶ U.S.-China Joint Statement, January 19, 2011, Washington, D.C., available at <http://www.whitehouse.gov/the-press-office/2011/01/19/us-china-joint-statement>.

²⁷ Thomas J. Christensen, “The Need to Pursue Mutual Interests in U.S.-PRC Relations,” United States Institute of Peace Special Report, forthcoming. See also “AIT Chairman Raymond Burghardt Press Roundtable Taipei,” American Institute in Taiwan, January 25, 2011, available at <http://www.ait.org.tw/en/officialtext-ot1102.html>; Susan V. Lawrence and Thomas Lum, “U.S.-China Relations: Policy Issues,” January 12, 2011, available at http://assets.opencrs.com/rpts/R41108_20110112.pdf; Bonnie Glaser, “U.S.-China Relations: Obama-Hu Summit: Success or Disappointment?” *Comparative Connections*, January 2010, available at http://csis.org/files/publication/0904qus_china.pdf; and Shih Hsiu-chuan, “Opinions differ on US-China statement,” *Taipei Times*, January 22, 2011, available at <http://www.taipetimes.com/News/taiwan/archives/2011/01/22/2003494135>.

²⁸ China uses the following terms and phrases in discussions of both its “core interests” as well as issues that it defines as its core interests, such as Taiwan, Tibet, and “sovereignty and territorial integrity” (主权和领土完整):

- 必须立场坚定，旗帜鲜明，敢于斗争，善于斗争，决不拿原则作交易 (must stand firm, be clear-cut, have courage to fight, and never trade away principles)
- 决不会妥协 (never compromise)
- 绝不会让步 (never yield)
- 决不退让 (not yield)
- 绝不会容忍 (never tolerate)
- 决不答应 (never agree)
- 绝不允许 (never allow)—in Anti-Secession Law
- 决不动摇 (never waver)—number 1 of Hu’s “Four Nevers”
- 决不放弃 (never give up)— number 2 of Hu’s “Four Nevers”
- 决不改变 (never change)— number 3 of Hu’s “Four Nevers”
- 决不妥协 (never compromise)— number 4 of Hu’s “Four Nevers”
- 没有任何回旋余地 (no room for maneuver)
- 没有任何妥协余地 (no room for compromise)
- 没有任何妥协和让步的余地 (no room for compromise or concession)
- 没有任何付价还价的余地 (no room for haggling)
- 决不会有丝毫犹豫、含糊和退让 (will never have the slightest hesitation, faltering, or concession)
- 在任何时候都决不会拿自己的主权和原则做交易 (will never at any time trade off its sovereignty and principles)
- 不容分割 (brooks no division)

For the most notable (among many) examples of official statements, see “President Hu sets forth guidelines on Taiwan,” Embassy of the People’s Republic of China in the Republic of Hungary, March 5, 2005, available at <http://www.chinaembassy.hu/hu/xwtd/t185890.htm>. This is Hu Jintao’s so-called “Four Nevers” speech. The Chinese version is: “胡锦涛就新形势下发展两岸关系提四点意见,” 中华人民共和国外交部, March 4, 2005, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t186259.htm>. Also see, “Speech at a Meeting Marking the 60th Anniversary of the Victory of the Chinese People’s War of Resistance Against Japanese Aggression and the World Anti-Fascist War,” September 3, 2005, available at <http://www.fmprc.gov.cn/eng/zxxx/t211664.htm>; “Getting to Know the Real China: Statement by H.E. Wen Jiabao, Premier of the State Council of the People’s Republic of China, at the General Debate of the 65th Session of the UN General Assembly,” September 23, 2010, available at <http://www.fmprc.gov.cn/eng/topics/wenjibaochuxi65jieUNdh/t761353.htm> (Chinese version: 中华人民共和国国务院总理 温家宝, “认识一个真实的中国,” 在第 65 届联大一般性辩论上的讲话, 2010 年 9 月 23 日 纽约, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/gjldrhd/t755356.htm>); and “Xu Caihou’s Visit To Promote Sino-U.S. Military Relationship,” *Jiefangjun Bao*, October 28, 2009, OSC CPP20091029705005 (in this article, Qian Lihua, director general of the Foreign Affairs Office of the Ministry of National Defense of the PRC, stated that “China will never compromise on issues involving such core interests as territorial sovereignty and national security.”)

²⁹ See Deng Xiaoping, “Speech at the Third Plenary Session of the Central Advisory Commission of the Communist Party of China,” October 22, 1984, *Selected Works of Deng Xiaoping*, volume 3, 1982–1992, ed. *People’s Daily Online*, available at <http://web.peopledaily.com.cn/english/dengxp/vol3/text/c1280.html>. In this source, Deng Xiaoping states that he told Prime Minister Margaret Thatcher in their 1984 meeting on the reversion of Hong Kong to Chinese rule that “sovereignty was not negotiable” (“主权问题是不能够谈判的”). This was said in the context of Hong Kong (and later in these remarks he compares Hong Kong to Taiwan). However, the comment is often taken to refer to sovereignty issues in general. Jiang Zemin also used the exact same phrase as Deng Xiaoping—“the sovereignty issue is non-negotiable” (“主权问题是不能够谈判的”), on this occasion in direct reference to Taiwan. See “江泽民主席会见美国总统国家安全事务助理伯杰,” 中华人民共和国外交部 (“President Jiang Zemin meets with U.S. Assistant to the President for National Security Affairs Samuel Berger,” Ministry of Foreign Affairs of the People’s Republic of China), November 7, 2000, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t7531.htm>. However, a 2008 article by current Foreign Minister Yang Jiechi is perhaps the most direct statement of the fundamental non-negotiability of

China's "core interests." See 杨洁篪, "改革开放以来的中国外交" (Yang Jiechi, "China's diplomacy since reform and opening"), September 16, 2008, available at <http://www.fmprc.gov.cn/chn/gxh/wzb/zxxx/t512782.htm>. Yang states: "In major issues involving core national interests, stand firm and clear-cut, dare to struggle, put up a good fight, and never trade principles. At the same time pay attention to strategy, adhere to a flexible and pragmatic approach, reasonable, beneficial, and restrained, maintaining the long-term and fundamental interests of our country." ("在涉及国家核心利益的重大问题上, 必须立场坚定, 旗帜鲜明, 敢于斗争, 善于斗争, 决不拿原则作交易。同时讲究策略, 灵活务实, 坚持有理、有利、有节, 维护我国的长远和根本利益。")

³⁰ For example, in a speech to the NPC on the draft Anti-Secession Law, Wang Zhaoguo asserts that "safeguarding sovereignty and territorial integrity" is China's core interest, and China has never forsworn the use of force to defend its "sovereignty and territorial integrity." See Wang Zhaoguo, "Draft Anti-Secession Law explained," speech at National People's Congress session, *China Daily*, March 8, 2005, available at http://www.chinadaily.com.cn/english/doc/2005-03/08/content_422875.htm. Chinese ambassador to the U.S. Zhou Wenzhong implied in 2007 that China would use non-peaceful means if necessary to prevent Taiwan independence in order to safeguard China's "core interest" of "national sovereignty and territorial integrity." "中国驻美国大使周文重在 '全球促进中国和平统一高峰论坛' 开幕式上发表讲话," 驻美国使馆供稿 ("Chinese ambassador to the U.S. Zhou Wenzhong gives speech at the opening ceremony of the 'Forum on the Global Promotion of the peaceful reunification of China,'" press release of the Chinese Embassy to the United States, November 17, 2007, available at http://www.fmprc.gov.cn/chn/pds/gjhdq/gj/bmz/1206_22/1206x2/t441126.htm).

A large number of unofficial Chinese media sources and personages convey a similar (and in some cases even more militant) resolve. See "American Shadow over South China Sea," editorial, *Global Times*, July 26, 2010, OSC CPP20100726722003; "Major General Luo Yuan Discusses the US-ROK Joint Military Exercise in the Yellow Sea," *Renmin Wang*, July 13, 2010, OSC CPP20100713787008; Li Yu-chi, "Vigorously Oppose US Meddling in South China Sea Issue," *Hsiang Kang Shang Pao*, August 11, 2010, OSC CPP20100811716003; Huang Xiangyang, "Commentary: Get Back When Attacked," *China Daily*, January 31, 2010, OSC CPP20100131968022; Yu Keli, director of the CASS Institute of Taiwan Studies, "The Two Sides of the Strait Should Look Squarely at the Issue of Ending the State of Hostility and Signing a Peace Agreement," *Zhongguo Pinglun*, August 2009, no. 140, pp. 4–9, OSC CPP20090805710007; and Zhou Zhihui, deputy director of the CASS Institute of Taiwan Studies, "A Programmatic Document for Advancing Peaceful Reunification," *Renmin Ribao* (Overseas Edition, Internet Version), October 19, 2007, p. 5, OSC CPP20071019704022.

³¹ Recently, former Secretary of State Henry Kissinger made the same observation regarding China's core interests, stating: "China has announced a number of 'core interests' which are, in essence, non-negotiable and for which China is prepared to fight, if necessary." See Henry A. Kissinger, "Power Shifts and Security," keynote address, 8th IISS Global Strategic Review, September 10, 2010, available at <http://www.iiss.org/conferences/global-strategic-review/global-strategic-review-2010/plenary-sessions-and-speeches-2010/keynote-address/henry-kissinger/>.

³² Adam Entous, "Tensions Ease in Asia, U.S. Says," *Wall Street Journal*, October 12, 2010; Abe Denmark, "China's Navy Gets Its Act Together, and Gets Aggressive," *Wired*, Danger Room, April 26, 2010, available at <http://www.wired.com/dangerroom/2010/04/chinas-navy-gets-its-act-together-and-gets-aggressive/>; Jiang Lifeng, "Some advice for Japan," *China Daily*, November 4, 2010, http://www.china.org.cn/opinion/2010-11/04/content_21269982.htm; Phoenix TV: China, Japan 'Likely To Engage in Fierce Confrontations Over Diaoyu'," *Feng Huang Wei Shih Chung Wen Tai*, interview with staff commentator Ho Liang-liang, October 22, 2010, OSC CPP20101025715031; "Disputed Island Tests Chinese Diplomacy," *Caijing*, September 29, 2010, OSC CPP20100929715030; Liu Huan, Meng Na, and Li Huizi, "Japan's Illegal Detention of Chinese Boat Captain Seriously Undermines People's Trust in Japan" (China Focus), Xinhua Asia-Pacific Service, September 20, 2010, OSC CPP20100920005007; Major General Luo Yuan, "US engaging in gunboat diplomacy," *People's Daily*, August 13, 2010, OSC CPP20100813787009; Cary Huang, "A Bolder China Asserts 'Core' Interests But Will It Act?" *South China Morning Post*, August 12, 2010, OSC CPP20100812715004; Wu Zhong, "A Daring Departure from Deng," *Asia Times*, August 6, 2010, available at <http://www.atimes.com/atimes/China/LH06Ad02.html>; John Pomfret, "U.S. takes a tougher tone with China," *Washington Post*, July 30, 2010, A01; Cary Huang, "War of Words Heats up With US Over South China Sea Disputes," *South China Morning Post*, July 29, 2010, OSC CPP20100729715012; "American Shadow over South China Sea," Editorial, *Global Times*, July

26, 2010, OSC CPP20100726722003; “We Would Like To Advise the United States Not To Make the South China Sea Troubled Waters,” *Ta Kung Pao*, July 26, 2010, OSC CPP20100726788010; “Major General Luo Yuan Discusses the US-ROK Joint Military Exercise in the Yellow Sea,” *Renmin Wang*, July 13, 2010, OSC CPP20100713787008; Wu Ting and Zhou Yun, “Sino-US-Japanese Sea Power Games in West Pacific,” *Dongfang Zaobao*, July 12, 2010, OSC CPP20100718138002; “China tells U.S. that S. China Sea is ‘core interest’ in new policy,” *Kyodo News*, July 3, 2010, OSC JPP20100703969016; Michael Wines, “Behind a Military Chill: A More Forceful China,” *New York Times*, June 8, 2010; Edward Wong, “Chinese Military Seeks to Extend Its Naval Power,” *New York Times*, April 23, 2010; and Wang Te-chun, “Strategic Changes Are Currently Taking Place in China-US Relations,” *Ta Kung Pao*, November 10, 2009, OSC CPP20091110710002.

³³ Some of the most significant official sources, as also cited above, include (in reverse chronological order): 丁刚, “唐家璇会见美国务卿,” 人民日报 (Ding Gang, “Tang Jiaxuan Meets U.S. Secretary of State,” *People’s Daily*), January 21, 2003, 3rd edition, available at <http://data.people.com.cn>; “Foreign Minister Tang Jiaxuan Held Talks With Powell,” Ministry of Foreign Affairs of the People’s Republic of China, February 25, 2003, available at <http://www.fmprc.gov.cn/eng/zxxx/t19583.htm>; “President Hu sets forth guidelines on Taiwan,” Embassy of the People’s Republic of China in the Republic of Hungary, March 5, 2005, available at <http://www.chinaembassy.hu/hu/xwtd/t185890.htm>; “Foreign Ministry Spokesperson Qin Gang’s Regular Press Conference on 27 November, 2007,” Ministry of Foreign Affairs of the People’s Republic of China, November 28, 2007, available at <http://www.fmprc.gov.cn/eng/xwfw/s2510/2511/t385091.htm>; “Foreign Ministry Spokesperson Qin Gang’s Regular Press Conference on June 22, 2010,” Ministry of Foreign Affairs of the People’s Republic of China, June 23, 2010, available at <http://www.mfa.gov.cn/eng/xwfw/s2510/2511/t711047.htm>.

³⁴ Information Office of the State Council of the People’s Republic of China, “Tibet—Its Ownership And Human Rights Situation,” government white paper, September 1992, Beijing, China, available in English at <http://www.fmprc.gov.cn/eng/topics/3711/t18996.htm#4>, available in Chinese at <http://www.fmprc.gov.cn/chn/pds/ziliao/zt/ywzt/wzwt/2296/t10532.htm>. The white paper also refers to Tibet as a “fundamental issue” (根本问题).

³⁵ At the meeting, Zeng stated: “The Chinese government appreciates the government of Sri Lanka for adhering to the one-China policy for a long period of time and offering precious support to China on Taiwan, Tibet and other questions involving China’s core interest.” See “Zeng Qinghong Meets with Wickremayake,” Ministry of Foreign Affairs of the People’s Republic of China, April 22, 2006, available at <http://www.fmprc.gov.cn/eng/wjb/zzjg/yzs/gjlb/2782/2784/t248856.htm>. For the Chinese version, see “曾庆红会见维克拉马纳亚克,” 中华人民共和国外交部 (“Zeng Qinghong meets with Wickremayake,” Ministry of Foreign Affairs of the People’s Republic of China), April 22, 2006, available at <http://www.fmprc.gov.cn/chn/pds/gjhdq/gjhdqzz/mzblwelm/xgxw/t248035.htm>. Also see “Foreign Ministry Spokesperson Qin Gang’s Regular Press Conference on November 27, 2008,” Permanent Mission of the People’s Republic Of China to the United Nations Office at Geneva and Other International Organizations in Switzerland, November 28, 2008, available at <http://www.china-un.ch/eng/fyrth/t524172.htm>.

³⁶ “Carry On Traditional Friendship And Deepen All-round Cooperation,” Address by Hu Jintao President of the People’s Republic of China At Islamabad Convention Center, SRC-406, November 24, 2006, available at <http://www.fmprc.gov.cn/eng/wjb/zzjg/yzs/gjlb/2757/2758/t285917.htm>.

³⁷ For example, see “驻密克罗尼西亚大使张卫东拜会密 外长罗伯特,” 驻密克罗尼西亚联邦使馆 (“Ambassador to the Federated States of Micronesia Zhang Weidong holds a secret meeting with Foreign Minister Robert [Luobote],” Embassy in Federated States of Micronesia), August 19, 2009, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/zwbdt/t579464.htm>; “杨洁篪外长会见美国常务副国务卿 斯坦伯格,” 中华人民共和国外交部 (“Foreign Minister Yang Jiechi Meets with U.S. Deputy Secretary of State James Steinberg,” Ministry of Foreign Affairs of the People’s Republic of China), September 30, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/mtb/bldhd/t616704.htm>; “习近平会见美国常务副国务卿斯坦 伯格,” 中华人民共和国外交部 (“Xi Jinping Meets with U.S. Deputy Secretary of State James Steinberg,” Ministry of Foreign Affairs of the People’s Republic of China), September 29, 2009, available at <http://www.fmprc.gov.cn/chn/pds/wjdt/gjldrhd/t615959.htm>; and Australia-China Joint Statement, 30 October 2009, available at <http://www.china.embassy.gov.au/bjng/statement.html>. The Chinese version is found here: “外交部副部长何亚非谈《中澳联合 声明》,” 中华人民共和国外交部 (“Vice Foreign Minister He Yafei

Talks about 'China-Australia Joint Statement', Ministry of Foreign Affairs of the People's Republic of China), September 31, 2009, available at <http://www.fmprc.gov.cn/chn/gxh/tyb/zyxw/t623887.htm>.

³⁸ On the subject of the Yellow Sea, PRC Foreign Ministry spokesperson Qin Gang neither confirmed nor denied its "core interest" status in response to direct questioning in a press conference. See "Foreign Ministry Spokesperson Qin Gang's Regular Press Conference on July 13, 2010," Ministry of Foreign Affairs of the People's Republic of China, July 14, 2010, available at <http://www.mfa.gov.cn/eng/xwfw/s2510/2511/t716745.htm>.

"Q: Recently, there are lots of reports on China's core interests in the Yellow Sea. Could you outline China's core interests? And in what way will the US-ROK joint naval exercise compromise China's core interests?"

A: China's core interests refer to national sovereignty, security, territorial integrity and development interests."

As for the Diaoyu (Senkaku) Islands, in response to a direct question on their "core interest" status, PRC Foreign Ministry spokesperson Hong Lei neither confirmed nor denied that the disputed islands were part of China's core interest. See "Foreign Ministry Spokesperson Hong Lei's Regular Press Conference on November 2, 2010," Ministry of Foreign Affairs of the People's Republic of China, November 3, 2010, available at <http://www.mfa.gov.cn/eng/xwfw/s2510/2511/t766610.htm>.

"Q: . . . China once said that issues such as Taiwan and Tibet bear on China's core interest. Is the Diaoyu island issue also part of China's core interest?"

A: The Diaoyu Island and its affiliated islands have been China's inherent territory since ancient times."

³⁹ Edward Wong, "Chinese Military Seeks to Extend Its Naval Power," *New York Times*, April 23, 2010. Wong cites "an American official involved in China policy" as follows: "In March, Chinese officials told two visiting senior Obama administration officials, Jeffrey A. Bader and James B. Steinberg, that China would not tolerate any interference in the South China Sea, now part of China's 'core interest' of sovereignty, said an American official involved in China policy. It was the first time the Chinese labeled the South China Sea a core interest, on par with Taiwan and Tibet, the official said."

⁴⁰ For example, see "China tells U.S. that S. China Sea is 'core interest' in new policy," Kyodo News, July 3, 2010, OSC JPP20100703969016; and John Pomfret, "U.S. takes a tougher tone with China," *Washington Post*, July 30, 2010, A01.

⁴¹ "China Retracts Policy on S. China Sea, Tells U.S.," Kyodo, October 22, 2010, OSC JPP20101022969102.

⁴² See Hillary Rodham Clinton, Secretary of State, Interview With Greg Sheridan of *The Australian*, Melbourne, Australia, November 8, 2010, available at <http://www.state.gov/secretary/rm/2010/11/150671.htm>. The transcript of the interview quotes Clinton as stating, "And when China first told us at a meeting of the Strategic and Economic Dialogue that they viewed the South China Sea as a core interest, I immediately responded and said we don't agree with that." The reporter then asked, "Was that Dai Bingguo that said that to you?" And Clinton replied, "Yes, yeah."

⁴³ For example, see "Foreign Ministry Spokesperson Jiang Yu's Regular Press Conference on September 21, 2010," Ministry of Foreign Affairs of the People's Republic of China, September 21, 2010, available at <http://www.fmprc.gov.cn/eng/xwfw/s2510/2511/t756092.htm>. In response to a specific question ("US officials said China once expressed that the South China Sea is its core interest. Please confirm."), Jiang replied: "Each country has its own core interests. Issues concerning state sovereignty, territorial integrity and major development interests are significant to all countries. China believes the South China Sea issue is only the dispute of territorial sovereignty and maritime rights and interests between relevant countries rather than an issue between China and ASEAN, let alone a regional or international issue. It can only and must be settled through friendly consultations between both parties in a peaceful manner. Adhering to 'putting aside disputes and seeking common development,' we are always committed to proper settlement through bilateral consultations with relevant countries. The channel of communication is smooth."

⁴⁴ John D. Banusiewicz, "Gates Receives, Accepts Invitation to Visit China," American Forces Press Service, October 11, 2010, available at <http://www.defense.gov/news/newsarticle.aspx?id=61218>; "China Retracts Policy on S. China Sea, Tells U.S.," Kyodo News, October 22, 2010, OSC JPP20101022969102.

⁴⁵ Hu stated, "Taiwan- and Tibet-related issues concern China's sovereignty and territorial integrity, and they represent China's core interests." "Address by Chinese President Hu Jintao at a U.S.-China Business Community Luncheon on U.S.-China Relations," Marriott Wardman Park Hotel, Washington, D.C., January 20, 2011, transcript by Federal News Service. See also Douglas H. Paal, "China and the United

States: Reining in Mutual Mistrust,” Asia Pacific Brief, Carnegie Endowment for International Peace, January 31, 2010, available at

<http://www.carnegieendowment.org/publications/index.cfm?fa=view&id=42422>.

⁴⁶ Personal communication with senior U.S. officials. One should add that Beijing probably emphasized the importance of the South China Sea territorial issue to Chinese interests at the March 2010 meeting with Steinberg and Bader in order to deter Washington from raising the issue at the Hanoi meeting of the ASEAN Regional Forum (ARF), scheduled for the following July. If so, the Chinese clearly failed in that attempt, since the ARF meeting witnessed a largely U.S.-orchestrated discussion of the issue, and a clear U.S. attempt to increase its overall involvement in the controversy. This event will be discussed in greater detail in an upcoming *CLM*, as part of an analysis of Chinese behavior regarding maritime issues.

⁴⁷ State Councilor Dai Bingguo, Remarks at Strategic and Economic Dialogue, “Concluding Joint Statements by Chinese Vice Premier Wang Qishan, Chinese State Councilor Dai Bingguo, Secretary of State Hillary Rodham Clinton, and Secretary of Treasury Timothy Geithner at the U.S.-China Strategic and Economic Dialogue (as released by the State Department),” Federal News Service, May 25, 2010; and Strategic and Economic Dialogue Opening Session, Remarks by Secretary of State Hillary Rodham Clinton; President Hu Jintao; Vice-Premier Wang Qishan; Treasury Secretary Timothy Geithner; State Councilor Dai, Great Hall of the People, Beijing, China, May 23, 2010, available at <http://www.state.gov/secretary/rm/2010/05/142134.htm>.

⁴⁸ Personal communication, senior U.S. official.

⁴⁹ For example, in the 2002 “Declaration on the Conduct of Parties in the South China Sea,” signed by all the claimants to South China Sea territories and waters (including China), the parties concerned undertook “to resolve their territorial and jurisdictional disputes by peaceful means, without resorting to the threat or use of force, through friendly consultations and negotiations by sovereign states directly concerned, in accordance with universally recognized principles of international law, including the 1982 UN Convention on the Law of the Sea” See “Declaration on the Conduct of Parties in the South China Sea,” November 4, 2002, available at <http://www.aseansec.org/13163.htm>. A similar commitment to “negotiation [of the South China Sea dispute] according to international law” is also contained in the joint statement between China and Vietnam signed in 2008. See China-Viet Nam Joint Statement, Beijing, China, October 25, 2008, available at <http://www.fmprc.gov.cn/eng/wjb/zzjg/yzs/gjlb/2792/2793/t520438.htm>.

For additional official affirmations of China’s desire to employ peaceful negotiations, talks, and consultations in addressing this issue, see “Foreign Ministry Spokesperson Jiang Yu’s Regular Press Conference on September 21, 2010,” Ministry of Foreign Affairs of the People’s Republic of China, September 21, 2010, available at <http://www.fmprc.gov.cn/eng/xwfw/s2510/2511/t756092.htm>; Yi Ling, “People’s Liberation Army Deputy Chief of General Staff Ma Xiaotian Reiterates China’s Stand on the South Sea Issue; China Protects Legitimate Navigation Freedom in the South Sea,” Xinhua, October 22, 2010, OSC CPP20101022163018; and “Chinese FM refutes fallacies on the South China Sea issue,” *China Daily*, July 25, 2010, available at http://www.chinadaily.com.cn/china/2010-07/25/content_11046054.htm.

⁵⁰ An October 2010 article in the *South China Morning Post* reported that the Chinese government had indeed privately included the Diaoyu Islands and South China Sea in its category of “core interests,” citing “diplomats familiar with the process.” PRC leaders allegedly “agreed in a meeting late last year to classify major foreign relations issues into two categories”: “core national interest” and “national interest.” See Cary Huang, “Diaoyus Row Marks Shift in Beijing’s Diplomatic Posture,” *South China Morning Post Online*, October 3, 2010, OSC CPP20101004715017. However, such a formulation, if it truly exists, has never been publicly articulated. More importantly, as explained in this *CLM* article, the specific issues of the Diaoyu Islands and the South China Sea have never been publicly identified by Chinese officials as core interests.

⁵¹ This issue will be discussed in greater detail in an upcoming *CLM*.

⁵² For example, see Wang Haiyun (vice president of Chinese Society for Study of the History of Sino-Russian relations), “China Does Not have an ‘Anti-Intervention Strategy,’” *Dongfang Zaobao*, October 12, 2010, OSC CPP20101018038001. Major General Wang (a former military attaché to Russia) largely agrees with the argument of these unofficial observers, stating: “The idea that ‘the South China Sea issue is a core interest for China’ has only been mentioned by individual Chinese scholars and has not been formally stated by high-level leaders; still less has it been written into national strategic reports or legal documents.” However, he then confuses the situation by adding: “The best explicit formulation is that ‘the question of South China Sea sovereignty involves China’s core interests,’ and to declare that even though it is an issue

involving China's core interests, China advocates resolving it through peaceful and friendly consultation." This seems like a distinction without much of a difference! In any event, Wang states that such a formulation "will not give the United States a pretext for attacking China, and [will] . . . reduce suspicion of China among countries around the South China Sea, and curb their irrational impulse to bring in external forces to put pressure on China." In contrast, at least one other Chinese military officer has made a distinction between "core interests" and "important maritime interests," placing Taiwan in the former category and the South China Sea in the latter. See "Major General Yin Zhuo, a Noted Military Expert, Comments on US-South Korean Military Exercises," *Renmin Wang*, July 29, 2010, OSC CPP20100730787001.

⁵³ See Chen Chenchen, "East Asian Unity Will Survive Rocky Waters," interview with Su Hao, director of the Asia-Pacific Research Center, China Foreign Affairs University, *Global Times Online*, October 25, 2010, OSC CPP20101026722011. Su Hao states, "China has never said publicly that sovereignty in the South China Sea is one of China's core interests. The concept has been hyped by the US to alienate China from its surrounding partners." See also Wang Haiyun, "China Does Not have an 'Anti-Intervention Strategy'," *Dongfang Zaobao*, October 12, 2010, OSC CPP20101018038001. Wang accuses the United States of promoting a "false proposition" about China's categorization of the South China Sea in order to "stir up anti-China feelings in the area and build a containment zone against China."

⁵⁴ For example, see Da Wei, "Why Should China Declare Its Core Interests?" *Huanqiu Shibao*, July 28, 2010, OSC CPP20100729788013. Da warns against overasserting "core interests" or misinterpreting the meaning of the term, especially since every state has core interests and asserting them can be not only defensive (when other countries are infringing on one's interests) but also offensive (when national strength is rising). Colonel Han Xudong of NDU argues that China should not declare its specific "core interests" at the present time because: a) China does not have the capability to militarily defend all of them; and b) it detracts from a focus on other non-core but still important interests. Han argues that China's core interests should be announced in "batches" as China's strength grows—hardly a reassuring thought to many foreign observers. See Han Xudong, "Prudent To Use Core National Interests," *Liaowang*, no. 30, July 26, 2010, p. 64, OSC CPP20100729788017.

⁵⁵ See endnotes 32, 52, and 54.

⁵⁶ For example, see "A National Security Strategy for A New Century," the White House, December 1999, pp. 1–2, available at <http://www.fas.org/man/docs/nssr-1299.pdf>. This Clinton administration document identified three levels of "national interest": "vital interests," "important national interests," and "humanitarian and other interests."

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

Regime Insecurity and International Cooperation

M. Taylor Fravel

Explaining China's Compromises
in Territorial Disputes

Following the 1995–96 Taiwan Strait crisis, scholars and policymakers alike have become increasingly concerned about the territorial ambitions of the People's Republic of China (PRC). Gerald Segal, for example, described “an irredentist China with a boulder rather than just a chip on its shoulder.”¹ Most recently, a 2005 Pentagon report on Chinese military power expressed concern that “conflicts to enforce China's [territorial] claims could erupt in the future with wide regional repercussions.”² Yet China has also frequently used cooperative means to manage its territorial conflicts, revealing a pattern of behavior far more complex than many portray. Since 1949, China has settled seventeen of its twenty-three territorial disputes. Moreover, it has offered substantial compromises in most of these settlements, usually receiving less than 50 percent of the contested land.

China's pattern of compromise in its territorial disputes presents several puzzles. For realists, China has not used its power advantages to bargain hard over contested land, especially with its weaker neighbors. Nor has it become

M. Taylor Fravel is Assistant Professor of Political Science and a member of the Security Studies Program at the Massachusetts Institute of Technology.

The author owes special thanks to Allen Carlson, Tanisha Fazal, George Gavrilis, Michael Glosny, Hein Goemans, Lei Guang, Ron Hassner, Alastair Iain Johnston, Andrew Kennedy, Alexander Montgomery, Nikolay Marinov, Kevin Narizny, Robert Ross, Holger Schmidt, Benjamin Valentino, Alan Wachman, Xu Xin, the anonymous reviewers, and especially Thomas Christensen, James Fearon, Jean Oi, and Scott Sagan, for helpful comments and suggestions. Seminar participants at the University of California, Berkeley, the University of Chicago, Duke University, Harvard University, the Institute on Qualitative Research Methods at Arizona State University, and Stanford University also provided useful feedback. Stanford's Center for International Security and Cooperation and Harvard's John M. Olin Institute for Strategic Studies offered valuable financial support for which the author is grateful. All interviews for this project were conducted on a not-for-attribution basis.

-
1. Gerald Segal, “East Asia and the ‘Constraint’ of China,” *International Security*, Vol. 20, No. 4 (Spring 1996), p. 110. See also Richard K. Betts, “Wealth, Power, and Instability: East Asia and the United States after the Cold War,” *International Security*, Vol. 18, No. 3 (Winter 1993/94), pp. 34–77; Aaron L. Friedberg, “Ripe for Rivalry: Prospects for Peace in Multipolar Asia,” *International Security*, Vol. 18, No. 3 (Winter 1993/94), pp. 5–33; John J. Mearsheimer, *The Tragedy of Great Power Politics* (New York: W.W. Norton, 2001), p. 375; and Denny Roy, “Hegemon on the Horizon? China's Threat to East Asian Security,” *International Security*, Vol. 19, No. 1 (Summer 1994), pp. 149–168.
 2. Office of the Secretary of Defense, *Annual Report to Congress: The Military Power of the People's Republic of China, 2005*, p. 9, <http://www.defenselink.mil/news/Jul2005/d20050719china.pdf>.

International Security, Vol. 30, No. 2 (Fall 2005), pp. 46–83
© 2005 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

less willing to offer concessions over disputed territory as its power has increased. Instead, China compromised in eight separate disputes as its power grew rapidly in the 1990s. For constructivists, the legacy of "unequal treaties" that ceded land to foreign powers in the nineteenth century and the central role of national unification in modern Chinese history suggest that conflicts over territory should be highly salient for China's leaders and basically nonnegotiable.³ In its many compromises, however, China has accepted the general boundaries that these treaties created except in the cases of Hong Kong and Macao.

Analysis of China's dispute behavior bears directly on the future of peace and stability in East Asia. Behavior in territorial disputes is a fundamental indicator of whether a state is pursuing status quo or revisionist foreign policies, an issue of increasing importance in light of China's rising power.⁴ In addition, China still has active territorial disputes, some of which have resulted in armed conflict in the past. Finally, although key to assessing its international relations since 1949, China's territorial disputes have not been studied systematically in the academic literature.⁵ The few comprehensive studies that do exist examine only China's settlements in the 1960s, not the 1990s, and were unable to benefit from the flowering of new Chinese-language source materials in the last ten years.⁶

3. The "unequal treaties" refer to agreements that the Qing dynasty signed with foreign powers in the nineteenth and early twentieth centuries. The agreements, often reached after a military defeat, contained one-sided terms requiring China to cede land, pay reparations, open treaty ports, or grant extraterritorial privileges to foreign citizens.

4. On China's status quo foreign policy in arenas other than territorial disputes, see Alastair Iain Johnston, "Is China a Status Quo Power?" *International Security*, Vol. 27, No. 4 (Spring 2003), pp. 5–56.

5. Scholars usually examine China's disputes through analysis of bilateral relations or international law. Examples include Luke T. Chang, *China's Boundary Treaties and Frontier Disputes* (New York: Oceana, 1982); Pao-min Chang, *The Sino-Vietnamese Territorial Dispute* (New York: Praeger, 1986); George Ginsburgs and Carl F. Pinkele, *The Sino-Soviet Territorial Dispute, 1949–64* (London: Routledge, 1978); Ting Tsz Kao, *The Chinese Frontiers* (Aurora, Ill.: Chinese Scholarly Publishing, 1980); Ying Cheng Kiang, *China's Boundaries* (Lincolnwood, Ill.: Institute of China Studies, 1985); Alastair Lamb, *The Sino-Indian Border in Ladakh* (Canberra: Australian National University Press, 1973); Neville Maxwell, *India's China War* (New York: Pantheon, 1970); J.R.V. Prescott, Harold John Collier, and Dorothy F. Prescott, *Frontiers of Asia and Southeast Asia* (Carlton: Melbourne University Press, 1977); Tsien-hua Tsui, *The Sino-Soviet Border Dispute in the 1970's* (New York: Mosaic Press, 1983); and Byron N. Tzou, *China and International Law: The Boundary Disputes* (New York: Praeger, 1990).

6. Greg Austin, *China's Ocean Frontier: International Law, Military Force, and National Development* (Canberra: Allen and Unwin, 1998); Harold C. Hinton, *Communist China in World Politics* (New York: Houghton Mifflin, 1966), pp. 273–336; Eric A. Hyer, "The Politics of China's Boundary Dis-

Within the broader study of territory and war, the past decade has witnessed increased study of the peaceful resolution of territorial disputes. Using mostly quantitative analysis, scholars have identified important empirical regularities in patterns of dispute settlement. Paul Huth's pathbreaking 1996 study, for example, found that democracy, alliances, disputes over land with economic value, the existence of multiple disputes, and prior defeat in armed conflict over contested land significantly increased the odds of a settlement being reached.⁷ Other scholars have stressed the pacific effects of leadership tenure, the salience of contested land, regime-type homogeneity, third-party intervention, and international norms.⁸ More recent efforts examine separately the decisions to open negotiations and offer concessions. In one of the most detailed studies to date, Huth and Todd Allee demonstrate that democratic leaders are less likely to threaten force and more likely to pursue negotiations in their territorial disputes with other states, as well as more likely to offer concessions to other democracies.⁹

Although major progress has been achieved in understanding the peaceful resolution of territorial disputes, several important puzzles remain. While democracies are more likely to pursue peaceful settlement efforts, they represent a minority of participants in twentieth-century territorial disputes. Instead, about two-thirds of the challengers in these conflicts are nondemocracies, such

putes and Settlements," Ph.D. dissertation, Columbia University, 1990; and Francis Watson, *The Frontiers of China* (New York: Praeger, 1966). For an important exception, see Allen R. Carlson, "Constructing the Dragon's Scales: China's Approach to Territorial Sovereignty and Border Relations in the 1980s and 1990s," *Journal of Contemporary China*, Vol. 12, No. 37 (November 2003), pp. 677–698.

7. Paul K. Huth, *Standing Your Ground: Territorial Disputes and International Conflict* (Ann Arbor: University of Michigan Press, 1996).

8. See, for example, Giacomo Chiozza and Ajin Choi, "Guess Who Did What: Political Leaders and the Management of Territorial Disputes, 1950–1990," *Journal of Conflict Resolution*, Vol. 47, No. 3 (June 2003), pp. 251–278; Gary Goertz and Paul F. Diehl, *Territorial Changes and International Conflict* (New York: Routledge, 1992); Paul R. Hensel, "Contentious Issues and World Politics: The Management of Territorial Claims in the Americas, 1816–1992," *International Studies Quarterly*, Vol. 45, No. 1 (March 2001), pp. 81–109; Paul K. Huth and Todd L. Allee, *The Democratic Peace and Territorial Conflict in the Twentieth Century* (Cambridge: Cambridge University Press, 2002), pp. 138–230; Arie M. Kacowicz, *Peaceful Territorial Change* (Columbia: University of South Carolina Press, 1994); and Beth A. Simmons, "Capacity, Commitment, and Compliance: International Institutions and Territorial Disputes," *Journal of Conflict Resolution*, Vol. 46, No. 6 (December 2002), pp. 829–856. On territorial norms, see Tanisha M. Fazal, "State Death in the International System," *International Organization*, Vol. 58, No. 2 (Spring 2004), pp. 311–344; and Mark W. Zacher, "The Territorial Integrity Norm: International Boundaries and the Use of Force," *International Organization*, Vol. 55, No. 2 (Spring 2001), pp. 215–250.

9. Huth and Allee, *The Democratic Peace and Territorial Conflict in the Twentieth Century*, pp. 138–230.

as China.¹⁰ Although they are on average less likely to seek negotiations than democracies, authoritarian states account for the majority of attempts to pursue peaceful settlement. Moreover, their behavior is difficult to predict. As they face fewer domestic constraints, authoritarian leaders can more easily choose between escalation and cooperation.¹¹ Precisely because such leaders are more likely to use force than their democratic counterparts, explaining and understanding their decisions to negotiate and compromise should yield important theoretical insights in the study of territorial disputes.

The ambiguity of authoritarian behavior also suggests that the causal mechanisms explaining why and when states might offer concessions require further attention. Although democratic challengers are more willing to pursue negotiations, they are just as likely as authoritarian ones to offer concessions once negotiations begin.¹² Other variables from previous studies that might capture incentives to compromise are often situational, referring to the broader context of the dispute, and do not necessarily reflect the initial motivation for leaders to offer concessions. In addition, many of these variables, such as alliances or the military balance, are relatively stable in any given case and offer less leverage in understanding attempts to compromise over time in the same dispute.

To explain why and when states might compromise in territorial disputes, this article presents a counterintuitive argument about the effects of domestic conflict on foreign policy. Diversionary war theory asserts that leaders facing domestic strife provoke conflicts with other states just to improve their position at home.¹³ By contrast, I argue that internal conflict often creates conditions for cooperation, producing a "diversionary peace" instead of war. Embattled leaders are willing to cooperate with other states in exchange for assistance in countering their domestic sources of insecurity. In territorial disputes, leaders are more likely to compromise when confronting internal

10. In the Huth and Allee data set, democracies are those states with a POLITY net democracy score equal to or greater than 6. Challengers are those states that seek to alter the status quo in a dispute. See <http://www.yale.edu/unsy/jcr/jcrdatadec02.htm>.

11. Huth and Allee, *The Democratic Peace and Territorial Conflict in the Twentieth Century*, pp. 77–78. In this study, variables measuring the political security of authoritarian challengers did not produce strong results. *Ibid.*, pp. 164, 211.

12. Democratic targets, however, are more likely to offer concessions than authoritarian ones. *Ibid.*, pp. 199–201.

13. Jack S. Levy, "The Diversionary Theory of War: A Critique," in Manus I. Midlarsky, ed., *Handbook of War Studies* (Boston: Unwin Hyman, 1989), pp. 259–288.

threats to regime security, including rebellions and legitimacy crises. Facing these types of internal threats, leaders are more likely to trade territorial concessions for assistance from neighboring states, such as suppressing rebels or increasing bilateral trade.

Regime insecurity best explains China's many attempts to compromise in its territorial disputes. Most of China's disputes are located on its long land border adjacent to frontier regions where the authority of the regime has been weak. Ethnic minorities dominate these frontiers, which make up more than half of the country and were never governed directly before 1949. During periods of regime insecurity, especially in the event of ethnic unrest near its international boundaries, China's leaders have been much more willing to offer concessions in exchange for cooperation that strengthens their control of these areas, such as denying external support to separatists or affirming recognition of Chinese sovereignty over these regions.

China's willingness to compromise in territorial disputes carries several implications for international relations theory. First, China's cooperative behavior as an authoritarian state underscores the importance of moving beyond democracy in the study of regime type and cooperation.¹⁴ Second, China's compromises question the role of reputation building in explaining the intractability of territorial disputes between states. China offered many concessions despite clear incentives that its simultaneous involvement in multiple conflicts created to signal toughness and resolve, not conciliation.¹⁵ Third, China's behavior challenges existing arguments about the foreign policies of revolutionary states.¹⁶ In China, radical politics at home failed to produce assertive or belligerent foreign policies in most of its territorial disputes.¹⁷ Fourth, China's cooperative response to regime insecurity helps to explain a key puzzle for the diversionary war hypothesis: why many periods of domestic strife fail to pro-

14. For one such effort, see Mark Peceny and Caroline C. Beer with Shannon Sanchez-Terry, "Dictatorial Peace?" *American Political Science Review*, Vol. 96, No. 1 (March 2002), pp. 15–26.

15. See, for example, Barbara F. Walter, "Explaining the Intractability of Territorial Conflict," *International Studies Review*, Vol. 5, No. 4 (December 2003), pp. 137–153, especially pp. 149–150.

16. See, for example, Stephen M. Walt, *Revolution and War* (Ithaca, N.Y.: Cornell University Press, 1996).

17. Perhaps the only clear-cut case of radical politics creating belligerence would be China's 1958 shelling of Taiwan. See Thomas J. Christensen, *Useful Adversaries: Grand Strategy, Domestic Mobilization, and Sino-American Conflict, 1947–1958* (Princeton, N.J.: Princeton University Press, 1996), pp. 194–241. On China's pattern of escalation in territorial disputes, see M. Taylor Fravel, "The Long March to Peace: Explaining China's Settlement of Territorial Disputes," Ph.D. dissertation, Stanford University, 2003.

duce crisis escalation and the use of force.¹⁸ Violence is less common than diversionary war theory predicts because, under certain conditions, leaders can have strong incentives to pursue cooperation instead of war to strengthen their domestic political security.

Although the initiation and escalation of China's many territorial disputes are important topics, this article focuses only on China's territorial compromises. I begin by discussing how regime insecurity creates incentives for cooperation in territorial disputes. I then describe China's ethnic geography and outline why regime insecurity is likely to create incentives for China to cooperate in disputes adjacent to its frontier regions populated by ethnic minorities. The next four sections trace China's efforts to compromise in periods when regime insecurity was most acute: namely, the revolt in Tibet, the failure of the Great Leap Forward, the upheaval following the Tiananmen crisis, and separatist violence in Xinjiang. To demonstrate the link between regime insecurity and cooperation, I tap a range of newly available Chinese-language source materials, including party history documents, oral histories, memoirs, government training manuals, and provincial gazetteers.

Regime Insecurity and Cooperation in Territorial Disputes

The notion of regime insecurity as a source of cooperation extends earlier scholarship on the relationship between domestic politics and foreign policy. To explain alignment in the developing world, Steven David argues that leaders "omnibalance," forming alliances to balance against the most pressing threat that they face, foreign or domestic.¹⁹ For many state leaders, especially in authoritarian regimes and new democracies, the most pressing threats to their political survival emanate from internal political challenges such as rebellions and coups.²⁰ To maximize their tenure in office, leaders form alliances,

18. For a complementary explanation based on selection effects, see Giacomo Chiozza and H.E. Goemans, "Avoiding Diversionary Targets," *Journal of Peace Research*, Vol. 41, No. 4 (July 2004), pp. 423–443.

19. Steven R. David, "Explaining Third World Alignment," *World Politics*, Vol. 43, No. 1 (January 1991), pp. 233–256.

20. On internal insecurity and foreign policy, see Mohammed Ayoob, *The Third World Security Predicament: State Making, Regional Conflict, and the International System* (Boulder, Colo.: Lynne Rienner, 1995); Edward Azar and Chung-In Moon, eds., *National Security in the Third World: The Management of Internal and External Threats* (Aldershot, U.K.: Edward Elgar, 1988); and Brian L. Job, ed., *The Insecurity Dilemma: National Security of Third World States* (Boulder, Colo.: Lynne Rienner, 1992).

even with external adversaries, to balance against more immediate internal foes.²¹

My argument about regime insecurity applies this insight of omnibalancing to international cooperation more broadly. When leaders face internal threats to their survival, they may use foreign policy in addition to domestic tools such as repression to enhance their political security. They may cooperate to achieve different types of support: (1) to gain direct assistance in countering internal threats, such as denying material support to opposition groups; (2) to marshal resources for domestic priorities, not defense; or (3) to bolster international recognition of their regime, leveraging the status quo bias of the international system to delegitimize domestic challengers.²² When leaders face internal threats, they may also cooperate to enhance their external security and preempt potential attempts by other states to profit from their domestic woes. These effects of regime insecurity are paradoxical: efforts to consolidate political power at home, often through repression, produce efforts to cooperate abroad. While such behavior is peaceful, its source is not necessarily benign.

Regime insecurity offers one causal logic to explain why and when leaders might pursue otherwise costly policies of compromise over disputed territory. In active territorial disputes, leaders choose among three general strategies.²³ A delaying strategy involves doing nothing except maintaining a state's claims through public declarations.²⁴ An escalation strategy involves the threat or use of force over disputed territory. A cooperation strategy excludes the threat or use of force and involves an offer to compromise by dividing control of contested land or dropping outstanding claims. In most cases, such compromise precedes the final settlement of a dispute in a bilateral treaty or agreement.

21. For a similar explanation of Egypt's alliances, see Michael N. Barnett and Jack S. Levy, "Domestic Sources of Alliances and Alignments: The Case of Egypt, 1962–1973," *International Organization*, Vol. 45, No. 3 (Summer 1991), pp. 369–395.

22. Recent historical and quantitative scholarship provides general support for this logic of regime insecurity in the origins of détente and a finding that politically insecure leaders are less likely to initiate an international crisis than politically secure ones. See Jeremi Suri, *Power and Protest: Global Revolution and the Rise of Détente* (Cambridge, Mass.: Harvard University Press, 2003); and Giacomo Chiozza and H.E. Goemans, "Peace through Insecurity: Tenure and International Conflict," *Journal of Conflict Resolution*, Vol. 47, No. 4 (August 2003), pp. 443–467.

23. Huth, *Standing Your Ground*. These strategies represent ideal types, but they are not mutually exclusive.

24. Leaders can delay by participating in negotiations without offering to compromise in the dispute.

The delaying strategy is usually the least costly alternative at any point in time. Escalation contains many risks, including the uncertainty associated with spirals of hostility or domestic political punishment for military defeat in addition to the costs of war. Cooperation is risky because concessions over territory can carry a high domestic political price, which may weaken a leader's position or even result in political death. From a leader's perspective, continuing the dispute through a delaying strategy is often better than offering concessions or failing on the battlefield.²⁵

Although delay is usually the least costly strategy for leaders to adopt, maintaining a claim to another's land still carries a price. By fostering uncertainty about the security of vital interests and mistrust of intentions more broadly, a territorial claim creates poor diplomatic relations with the opposing state. Poor bilateral ties, in turn, limit the potential for cooperation in security, diplomatic, or economic arenas. When a conflict has been militarized, the defense of contested land also diverts scarce resources from domestic priorities.

Regime insecurity generates incentives for compromise by increasing the cost of poor bilateral relations created by the presence of disputed territory. When regime survival is at stake, trading concessions for assistance to enhance regime security often outweighs the benefits of maintaining a claim through delay.²⁶ State leaders are more likely to offer concessions to manage two types of challenges to regime security: internal threats to territorial integrity and internal threats to political stability.

Internal threats to territorial integrity usually occur as unrest or rebellions that challenge the most basic indicator of a regime's authority—the control of the territory claimed by the state. When a rebellion erupts near international boundaries, territorial disputes adjacent to the area of unrest become much more costly to pursue because neighboring states can provide a range of support for rebels or even seek to intervene in the conflict. Leaders facing unrest or rebellion near their borders will be more likely to trade territorial concessions for direct assistance in crushing the uprising, such as (1) sealing borders; (2) attacking rebel bases; (3) denying refuge or material support to rebels; (4)

25. Delaying may also carry specific benefits because it allows a state to buy time to strengthen its tactical military position or maintain domestic support from key constituencies. Delaying is especially useful for states that occupy a majority of disputed land because the passage of time consolidates a favorable status quo.

26. Regime insecurity represents one pathway toward territorial compromise. External threats to a state's security may also increase the cost of delay and create incentives for compromise.

extraditing rebel leaders; (5) minimizing inadvertent escalation during hot pursuit; (6) providing assurances not to intervene; or (7) affirming their state's sovereignty over the region of unrest.²⁷

Internal threats to political stability appear in the form of social unrest such as large-scale protests that question the legitimacy of a regime. When such threats arise, even when far from borders, territorial disputes become much more costly to pursue because they distract leaders from adequately addressing domestic unrest. Leaders facing political instability will be more likely to trade territorial concessions for direct or indirect assistance in overcoming internal unrest, such as (1) increasing trade and investment to stimulate economic development; (2) minimizing external tensions to marshal resources domestically; (3) delegitimizing domestic rivals; or (4) preempting attempts to profit from a regime's internal weakness.

A key variable is the salience of the contested land. The greater the importance of the territory at stake, the larger the magnitude of the internal threat necessary to make compromise more attractive than delay. In most disputes, some bargaining space exists for trading territorial concessions for assistance in enhancing regime security. In some cases, however, one or more of the disputants sees no possible trade, which endows the dispute with perceived indivisibility and increases the cost of side payments necessary for a peaceful settlement.²⁸ Such bargaining space is greatly reduced, for example, in disputes over areas populated by coethnics for one or more of the states, such as Kashmir.²⁹

Regime insecurity offers a plausible pathway for compromise in territorial disputes for two reasons. First, the majority of twentieth-century territorial disputes involved authoritarian regimes where internal threats to regime security are likely to occur, including popular uprisings, coups, revolutions, and secessions.³⁰ Second, external cooperation may be particularly important for authoritarian leaders, especially in the developing world, whose political insti-

27. In the absence of a territorial dispute with a neighboring state, one observable implication of this argument is that leaders will offer concessions in other types of disputes for cooperation in addressing regime insecurity.

28. On indivisibility and territory, see Ron E. Hassner, "To Halve and to Hold: Conflicts over Sacred Space and the Problem of Indivisibility," *Security Studies*, Vol. 12, No. 4 (Summer 2003), pp. 1–33; Monica Duffy Toft, *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory* (Princeton, N.J.: Princeton University Press, 2003); and Walter, "The Intractability of Territorial Conflict."

29. Huth, *Standing Your Ground*, pp. 141–180.

30. David, "Explaining Third World Alignment."

tutions often suffer from limited capacity or “broadcasting power.”³¹ When the state is weak, cooperation with neighbors will be more valuable because these leaders have fewer resources with which to respond to their internal threats.

Explaining China's Compromises

China has offered substantial compromises in seventeen of its twenty-three territorial disputes active since 1949. As Table 1 demonstrates, it has often agreed to accept less than half of the territory being disputed. In fifteen of these disputes, its willingness to compromise created the conditions for a final territorial settlement through bilateral agreements. By contrast, China has never offered to compromise in six of its twenty-three disputes, consistently adopting delaying strategies instead.³² Regime insecurity best explains the variation in China's use of cooperation and delaying strategies.

ETHNIC GEOGRAPHY AND REGIME INSECURITY

In China, ethnic geography has shaped the regime's vulnerability to internal threats. It refers to the density and distribution of ethnic groups within a state. China's ethnic geography reflects an “empire state,” with an ethnic majority core surrounded by a large periphery of minorities.³³ Han Chinese constitute more than 90 percent of China's population. They live mostly on roughly 40 percent of the landmass along the coast, an area known as “China proper” or the “inner land” (*neidi*).³⁴ By contrast, ethnic minorities such as Tibetans and Mongols make up less than 10 percent of China's population. They reside on

31. Jeffrey Herbst, *States and Power in Africa: Comparative Lessons in Authority and Control* (Princeton, N.J.: Princeton University Press, 2000).

32. As discussed below, disputes over Hong Kong and Macao were settled without compromise when Britain and Portugal agreed to return these territories to China in 1984 and 1987, respectively.

33. Owen Lattimore, *Inner Asian Frontiers of China* (New York: American Geographical Society, 1940); Mao Zhenfa, ed., *Bianfang lun* [On frontier defense] (Beijing: Junshi kexue chubanshe, 1996); Gerald Segal, “China Changes Shape: Regionalism and Foreign Policy,” Adelphi Paper No. 287 (London: International Institute of Strategic Studies, 1994); Michael D. Swaine and Ashley J. Tellis, *Interpreting China's Grand Strategy: Past, Present, and Future* (Santa Monica, Calif.: RAND, 2000), pp. 21–96; Joseph Witney, “China: Area, Administration, and Nation Building,” Department of Geography Research Paper No. 123 (Chicago: University of Chicago, 1970); and Zheng Shan, ed., *Zhongguo bianfang shi* [China's frontier defense history] (Beijing: Shehui kexue wenxian chubanshe, 1995).

34. Bu He, ed., *Minzu lilun yu minzu zhengce* [Nationality theory and nationality policy] (Huhehaote: Neimenggu daxue chubanshe, 1995), p. 27.

Table 1. China's Territorial Disputes, 1949-2005

Disputed Area	Km ²	Talks		Agreements		Description of Settlement
		1949-84	1984-87	1984-87	1987-91	
Homeland Disputes						
Hong Kong	1,042	1982-84	1984-84	JD	JD	Territory was returned to China.
Macao	25	1986-87	1986-87	JD	JD	Territory was returned to China.
Taiwan	32,260	—	—	—	—	Dispute is still active.
Frontier Disputes						
Burma border	1,909	1956-57	1956-57	—	—	Talks were held after troops clashed on the border. China received 18% of the disputed land; Burma received a strategic highway, a salt mine, and the preferred watershed boundary in north.
Nepal border	2,476*	1960	1960	1960: BA 1960: BT 1961: BP	1960: BA 1961: BT 1963: BP	China received 6% of the disputed land and half of Mt. Everest; Nepal held most of the grazing areas and passes.
India border	125,000	1960	1960	—	—	China offered to hold 26% of the disputed land; dispute is still active.
North Korea border	1,165	1962	1962	1962: BT 1964: BP	1962: BT 1964: BP	China received 40% of the disputed crater lake on Changbai Mt.; North Korea held the remaining land.
Mongolia border	16,808	1962	1962	1962: BT 1964: BP	1962: BT 1964: BP	China received 29% of the disputed land.
Pakistan border	8,806**	1962	1962	1963: BA 1965: BP	1963: BA 1965: BP	China received 60% of the disputed land but transferred 1,942 square kilometers to Pakistan.
Afghanistan border	~7,381	1963	1963	1963: BT 1965: BP	1963: BT 1965: BP	China did not receive any of the disputed Wakhan corridor.
Russia border (eastern)***	~1,000	1964	1964	—	—	Consensus reached to divide islands evenly according to the <i>thalweg</i> principle.
		1969-78	1969-78	—	—	—
		1987-91	1987-91	1991: BA 1999: BP	1991: BA 1999: BP	China received 52% of the river islands; other areas were divided evenly.

Bhutan border	1,128	1984-present	1998: MTA	China reportedly offered to hold 24% of the disputed land; dispute is still active.
Laos border	18	1990-91	1991: BT 1993: BP	China reportedly received 50% of the disputed land.
Vietnam border		1977	—	—
	227	1992-99	1993: PriA 1999: BT	China received 50% of the disputed land.
Russia border (western)***	no data	1992-94	1994: BA 1999: BP	Agreement affirmed the line of actual control.
Kazakhstan border***	2,420	1992-98	1994: BA 1997: SA 1998: SA 2002: BP	China received approximately 22% of the disputed land.
Kyrgyzstan border***	3,656	1992-99	1996: BA 1998: SA 2004: BP	China received approximately 32% of the disputed land.
Tajikistan border***	28,430	1992-2002	1999: BA 2002: SA	China received 4% of the disputed Pamirs; other sectors were divided evenly.
Strategic islands on Russian border***	408	1964 1990-2004	— 2004: SA	— Control of Abagaitu and Hexiazi islands was divided evenly.

Offshore Island Disputes

White Dragon Tail Island	~5	1957	No formal agreement	Island was transferred to North Vietnam.
Paracel Islands	~10	—	—	Dispute is still active.
Spratly Islands	~5	—	—	Dispute is still active.
Senkaku Islands	~7	—	—	Dispute is still active.

NOTE: BA (boundary agreement), BP (boundary protocol), BT (boundary treaty), CBMs (confidence-building measures), MTA (maintenance of tranquility agreement), JD (joint declaration), PriA (principles agreement), and SA (supplemental agreement).

*Figure excludes the dispute over Mt. Everest; **figure excludes the dispute over K2; ***territory also disputed with the Soviet Union.

the other 60 percent of the landmass enveloping China proper, regions known as the “borderlands” (*bianjiang*) or “outer lands” (*waidi*).³⁵

This ethnic geography highlights the varied challenges to maintaining territorial integrity that China’s leaders confronted when founding the PRC in 1949. The political institutions of the new state were strong in the core but weak in the larger frontiers. China proper was relatively easy to govern because the Chinese Communist Party (CCP) inherited from the Ming dynasty a provincial system of direct rule, which vertically integrated townships and provinces with the central government.³⁶ As the civil war was waged mainly within China proper, the CCP also cultivated a large pool of Han cadres to staff the new government.

By contrast, the institutions of the new state were weak in the frontiers, the regions adjacent to China’s land borders. Many ethnic groups in the borderlands did not identify themselves as members of the PRC and previously sought at various times to seek independence. Unlike China proper, the CCP inherited no institutions of direct rule through which to govern these areas. As strategic buffer zones, the frontiers had historically been governed indirectly through arrangements that granted substantial autonomy to local leaders in exchange for loyalty to the emperor.³⁷ Indirect rule ensured the exclusion of foreign influence without having to establish a costly system of direct rule, but limited the rapid expansion of state authority in these regions when the PRC was established.³⁸ As the civil war was waged largely in China proper, the

35. On the frontier areas, see Ma Dazheng and Liu Ti, *Ershi shiji de Zhongguo bianjiang yanjiu* [China’s borderland research in the twentieth century] (Harbin: Heilongjiang jiaoyu chubanshe, 1998), pp. 1–60; and Niu Zhongxun, *Zhongguo bianjiang dili* [China’s frontier geography] (Beijing: Renmin jiaoyu chubanshe, 1991), pp. 1–7.

36. Charles O. Hucker, *A Dictionary of Official Titles in Imperial China* (Stanford, Calif.: Stanford University Press, 1985), pp. 71–96.

37. Nicola Di Cosmo, “Qing Colonial Administration in Inner Asia,” *International History Review*, Vol. 20, No. 2 (June 1998), pp. 24–40; and Ma Dazheng, ed., *Zhongguo bianjiang jinglue shi* [A history of China’s frontier administration] (Zhengzhou: Zhongzhou guji chubanshe, 2000), pp. 240–434.

38. Many forms of indirect rule were adopted, including the local chieftain system in southwestern tribal areas, military colonies in Xinjiang and parts of Mongolia, a protectorship in Tibet, and vassalage relations in Manchuria and other parts of Mongolia. See Di Cosmo, “Qing Colonial Administration in Inner Asia”; Joseph Fletcher, “The Heyday of the Ch’ing Order in Mongolia, Sinkiang, and Tibet,” in John K. Fairbank, ed., *The Cambridge History of China*, Vol. 10: *Late Ch’ing, 1800–1911* (Cambridge: Cambridge University Press, 1978), pp. 58–90; James Millward, *Beyond the Pass: Economy, Ethnicity, and Empire in Qing Central Asia, 1759–1864* (Stanford, Calif.: Stanford University Press, 1998); Luciano Petech, *China and Tibet in the Early Eighteenth Century: History of the Establishment of the Chinese Protectorate in Tibet* (Leiden, Netherlands: E.J. Brill, 1972); Morris Rossabi, *China and Inner Asia: From 1368 to the Present Day* (New York: PICA Press, 1975); and Warren W. Smith, *Tibetan Nation: A History of Tibetan Nationalism and Sino-Tibetan Relations* (Boulder, Colo.: Westview, 1996), pp. 115–150.

Table 2. China's Compromises by Type of Dispute

	Homeland	Frontier	Offshore	Total
Compromise attempted	0	16(14)	1(1)	17(15)
No compromise attempted	3(2)	0	3	6(2)
Total	3(2)	16(14)	4	23(17)

NOTE: The figure in each cell refers to the number of disputes in which compromise was attempted. The figure in parentheses refers to the number of disputes in which a final settlement was reached.

CCP also lacked ethnic minority cadres, which further limited its ability to extend its authority in the frontiers.

China's territorial disputes reflect these different challenges to maintaining the territorial integrity of an empire state.

HOMELAND DISPUTES. China has disputed three areas linked to the Han Chinese core. In these disputes, the main challenge to territorial integrity has been to regain those parts of China proper not under PRC control in 1949. They include China's contentious conflicts over Taiwan, Hong Kong, and Macao. China's leaders view these disputes as key to completing the mission of national unification that began with the fall of the Qing dynasty in 1911, a goal intertwined with the very legitimacy of the CCP.

Concessions for any reason are unlikely in homeland disputes. The overriding importance of completing national unification suggests that these conflicts are basically nonnegotiable. Few threats, internal or external, would be great enough to make any territorial compromise appear more attractive than delay and the achievement of unification.

As Table 2 indicates, China has never attempted to compromise in a homeland dispute. Disputes over Hong Kong and Macao were settled when Britain and Portugal concluded that the costs of continuing them, especially after the expiry of a ninety-nine-year lease for the New Territories in Hong Kong, far exceeded whatever benefit they still derived from their possession. In a matter of eighteen months of talks with China, Britain's negotiating position changed from demanding an extension of British sovereignty after 1997 to bargaining over the details of the territory's handover to China. Portugal followed Britain's lead and agreed in 1987 to return Macao to China on similar terms.³⁹ Al-

39. Robert Cottrell, *The End of Hong Kong: The Secret Diplomacy of Imperial Retreat* (London: John Murray, 1993).

though China did agree to maintain existing political institutions in both territories, this did not represent much of a compromise because, since the late 1970s, China had indicated its desire to maintain the political status quo to facilitate investment from these territories.

FRONTIER DISPUTES. China has disputed control over sixteen areas along its land border adjacent to its borderlands. These disputes stem from the challenge of consolidating control over vast frontiers through the implementation of direct rule. Frontier disputes arose from the ambiguity surrounding China's land borders when the PRC was established. While large portions of the border had been delimited in the unequal treaties, many of the descriptions contained in these agreements were vague, imprecise, and often not demarcated on the ground. Neighboring states also feared that China might pursue claims to the almost 2 million square kilometers of territory ceded in these agreements. Nevertheless, China has pursued mostly status quo goals in these disputes to secure the boundaries of the late Qing dynasty as defined by the unequal treaties.

Internal threats are likely to create incentives to compromise in frontier disputes. The weakness of the state increases the potential influence of neighbors within these regions, creating an opportunity to trade territorial concessions for support in governing the frontiers. Securing one of the longest land borders in the world poses a logistical challenge for China's armed forces even under optimal conditions. In addition, the presence of ethnic minorities, many of whom aspire to secede, intertwines territorial integrity with political stability.⁴⁰ Before the establishment of the PRC, many of these groups enjoyed much stronger economic and cultural ties with neighboring states than with China proper and did not see themselves as members of the new state.⁴¹ As Deng Xiaoping explained presciently in 1950 when describing China's southwestern frontier, "On a border this long . . . if the issue of ethnic minorities is not resolved, then the matter of national defense cannot be settled."⁴²

40. Of the 135 counties adjacent to China's international frontiers, 107 are ethnic autonomous regions. Bu He, *Minzu lilun yu minzu zhengce*, p. 27.

41. Parts of Xinjiang had been ruled or heavily influenced by the Soviet Union before 1949, while India and Nepal maintained special trading privileges in Tibet and hill tribes moved freely across China's southwestern borders in Yunnan and Guangxi. See Andrew D.W. Forbes, *Warlords and Muslims in Chinese Central Asia: A Political History of Republican Sinkiang, 1911–1949* (New York: Cambridge University Press, 1986); Alastair Lamb, *The McMahon Line: A Study in the Relations between India, China, and Tibet, 1904–1914* (London: Routledge and K. Paul, 1966); and Herold J. Wiens, *China's March toward the Tropics* (Hamden, Conn.: Shoe String Press, 1954).

42. *Deng Xiaoping wenxuan* [Deng Xiaoping's selected works] (Beijing: Renmin chubanshe, 1994), Vol. 2, p. 161.

When internal threats arise, neighboring states can provide assistance to help China's leaders maintain frontier stability. While separatist unrest near borders poses a direct challenge to the integrity of frontiers, political instability at the core also creates concerns about control of these regions, given the potential for separatist movements and the need to focus state resources elsewhere. With its relative military strength, however, China's territorial claims create suspicion and mistrust among neighbors, which limits the development of bilateral ties. Under these conditions, China's leaders can offer territorial concessions to improve relations with its neighbors and strengthen frontier control.

As Table 2 demonstrates, China has compromised at least once in each of its frontier disputes since 1949. Chinese sources, especially doctrinal studies produced by scholars from the People's Liberation Army (PLA), link the defense of frontiers with internal political stability.⁴³ These studies assert that insecure borders promote ethnic unrest by increasing external influence within China and reveal suspicions that neighbors may manipulate ethnic tensions to create instability.⁴⁴ Diplomatic relations with neighboring states are consistently highlighted as a key tool for maintaining internal control over frontiers in addition to border security, economic development, and political mobilization. Benefits of cooperation with neighbors include policing rebel activity, limiting the potential for miscalculation during pacification campaigns, and expanding cross-border trade.⁴⁵

OFFSHORE ISLAND DISPUTES. China has disputed four island groups. In these disputes, China's leaders have sought to secure a maritime frontier that previous governments lacked. Offshore island disputes arose from the ambiguity of sovereignty over small islands, rocks, and reefs that had never been administered by any of the claimants, including China. Today, the value of the islands stems mostly from maritime rights to resources in adjacent waters and as bases for the limited forward deployment of naval forces for surveillance and securing sea-lanes.

Regime insecurity is unlikely to create incentives for cooperation in offshore

43. See especially Mao, *Bianfang lun*; and Li Xing, ed., *Bianfang xue* [The study of frontier defense] (Beijing: Junshi kexue chubanshe, 2004). See also police training manuals for frontier defense, including Cai Xiru, ed., *Bianfang lilun* [Frontier defense theory] (Beijing: Jingguan jiaoyu chubanshe, 1996); and Ping Qingfu, ed., *Bianjing guanli xue* [The study of border management] (Beijing: Jingguan jiaoyu chubanshe, 1999).

44. Mao, *Bianfang lun*, pp. 232–234, 256–261.

45. *Ibid.*, pp. 241–255; and Wang Wenrong, ed., *Zhanlue xue* [The study of military strategy] (Beijing: Guofang daxue chubanshe, 1999), p. 270.

island disputes. Far from the mainland, these small, desolate, and unpopulated islands have little influence on regime security. As they do not target vital interests for similar reasons, China's claims to these islands have a limited impact on bilateral ties. In addition, during periods of regime insecurity, the only assistance that neighbors might offer in exchange for China's concessions would be diplomatic support. As a result, external, not internal, factors are most likely to produce efforts to compromise. Furthermore, as these islands are cheap for the claimants to dispute, requiring few troops to maintain a claim, states are most likely to adopt a delaying strategy to maximize the potential economic or strategic benefits.

China has compromised in only one offshore island dispute since 1949, White Dragon Tail, as discussed below. In its disputes over the Paracel (*xisha*), Spratly (*nansha*), and Senkaku (*diaoyu*) island groups, China has consistently adopted a delaying strategy and never offered to compromise.⁴⁶ It has held limited talks with individual states over the Spratlys, but these talks have never touched upon sovereignty, emphasizing instead escalation control.⁴⁷ In November 2002, China did sign a declaration with ASEAN states on a code of conduct concerning the South China Sea, but the agreement focused only on broad confidence-building measures, not sovereignty and dispute settlement.⁴⁸

REGIME INSECURITY AND TERRITORIAL COMPROMISE

Since 1949, China's pattern of compromise and delay has varied with the presence or absence of threats to regime insecurity: China has attempted compromise in response to these internal threats; otherwise, it has pursued delay.

Unrest among minorities near international boundaries has occurred during three distinct periods since 1949, each of which produced efforts by China's leaders to compromise in its territorial disputes. When a revolt peaked in Tibet in 1959, China moved to compromise in disputes with Burma, Nepal, and India. After ethnic unrest in Xinjiang in 1962, China pursued compromise with Mongolia, Pakistan, Afghanistan, and the Soviet Union. In the 1990s, sustained separatist violence in Xinjiang produced compromises in disputes with Kazakhstan, Kyrgyzstan, and Tajikistan.

46. Austin, *China's Ocean Frontier*.

47. *Zhongguo waijiao* [China's diplomacy] (Beijing: Shijie zhishi chubanshe, 2001), p. 63.

48. "Declaration on the Conduct of Parties in the South China Sea," November 2002, <http://www.aseansec.org/13163.htm>.

Political instability has also occurred during three different periods, two of which produced attempts to compromise. In the aftermath of the Great Leap Forward in the early 1960s, concerns about bolstering political stability throughout the country created additional incentives to compromise in disputes with Mongolia, Afghanistan, Pakistan, and the Soviet Union, as well as North Korea. Following the 1989 Tiananmen crisis and worries about the viability of its socialist system, China attempted compromise in disputes with its socialist neighbors the Soviet Union, Laos, and Vietnam; in addition, it reached confidence-building agreements with India and Bhutan. A third period of political instability during the early phase of the Cultural Revolution (1966–69) failed to produce efforts to compromise. China's senior leaders, however, created this chaos deliberately, which did not initially reflect a threat to the regime's security from society. Moreover, during the most acute phase, much of the central government ceased to function, including the foreign ministry.

In the absence of internal threats to regime security, China's leaders have adopted delaying strategies instead. China attempted to compromise only twice in disputes before 1960. It did not offer new compromises in any territorial dispute from the mid-1960s until the late 1980s.⁴⁹

Nevertheless, regime insecurity leaves unexplained three attempts by China to compromise. First, China and Burma held negotiations in 1956 after border patrols clashed. Even though China offered to exchange some disputed areas, the impetus for the talks was to reduce the potential for escalation, as Nationalist troops from Taiwan had taken refuge in Burma and periodically raided the border of Yunnan Province.⁵⁰ Second, in 1957, China transferred White Dragon Tail Island (*bailongwei dao*) in the Tonkin Gulf to North Vietnam. Newly available sources indicate that Chairman Mao Zedong ordered this compromise to aid Hanoi in its conflict with the United States.⁵¹ Third, in October 2004, China and Russia agreed to divide control of two disputed river islands: Abagaitu in the Argun River and Heixiazi at the confluence of the Amur and Ussuri rivers. As no internal threats to regime security preceded this

49. During this period, China held negotiations with the Soviet Union (1969–78), Vietnam (1977), and India (1981–present).

50. Han Huaizhi and Tan Jingqiao, eds., *Dangdai zhongguo jundui de junshi gongzuo (shang)* [The military work of contemporary China's armed forces] (Beijing: Zhongguo shehui kexue chubanshe, 1989), Vol. 1, pp. 367–379.

51. Li Dechao, "Bailong weidao zhengming [Rectification of White Dragon Island's name]," *Zhongguo bianjiang shidi yanjiu baogao*, Vols. 1–2, No. 3 (1988), pp. 21–23; and Mao, *Bianfang lun*, p. 137.

agreement, China's compromise most likely stemmed from external factors, especially the need to deepen ties with Russia.⁵²

In addition to regime insecurity, China's relative military power has influenced its willingness to compromise in frontier disputes. With a large standing army, China has more leverage in disputes on its land border where it can most easily project its military power. While the authority of the state remains weak in the frontiers, China's relative strength decreases the risk that other states will perceive concessions in frontier disputes as a sign of weakness. By contrast, China has faced real limits to projecting power in offshore island disputes and faces greater risks that any concessions will be perceived as signaling weakness. Nevertheless, while military power increases the likelihood that China might compromise in a given frontier dispute, it cannot explain the variation in the timing of China's efforts to compromise or the motivation for compromise. On land, its relative power has been largely constant since 1949, while attempts to compromise have varied widely over time.

The following sections examine the relationship between regime insecurity and compromise in China's territorial disputes. The analysis demonstrates a clear correlation between the onset of internal threats and efforts to compromise. Where possible given limited access to archival material, the analysis also demonstrates that China's leaders acted in accordance with the hypothesized mechanisms of regime insecurity.

The Tibetan Revolt

In 1959, a revolt in Tibet against Chinese rule sparked the largest internal threat to the PRC's territorial integrity. The outbreak of the revolt dramatically increased the cost of maintaining territorial disputes with Burma, Nepal, and India. China offered concessions in these disputes in exchange for cooperation to crush the rebellion.

UNREST IN THE SOUTHWESTERN FRONTIER

After occupying Tibet in 1951, the PRC confronted severe obstacles to governing the region. Tibet maintained strong social, cultural, and economic ties with

52. "Russian Foreign Minister Gives TV Interview Focusing on Ties with Neighbors," Moscow NTV (in Russian), November 14, 2004, Foreign Broadcast Information Service (FBIS) CEP-2004-1114-000063.

its Himalayan neighbors, while the central government lacked any legacy institutions and ethnic cadres with which to integrate Tibet into the new state. As a result, China's leaders chose to govern Tibet indirectly through the Dalai Lama and existing Tibetan institutions. From the mid-1950s, however, a series of armed uprisings against Chinese rule in the ethnic Tibetan Kham region of Sichuan Province grew into a widespread revolt within Tibet proper.⁵³ At its height in late 1958 and early 1959, the revolt involved perhaps more than 50,000 rebels, who controlled most of Tibet apart from Lhasa, Shigatse, and the main highways.⁵⁴ The revolt was especially threatening because of the central government's already weak position in the region and because Tibet's basically undefended international border left it vulnerable to external influence.⁵⁵

Before the Tibetan revolt erupted, China had rebuffed efforts by neighboring states to open negotiations over disputed territory. In 1954, Chinese Premier Zhou Enlai and Indian Prime Minister Jawaharlal Nehru touched on the border in their talks, but agreed not to pursue it.⁵⁶ In 1956, Nepal sought to open talks with China over their border, but China demurred.⁵⁷ The same year, China and Burma held negotiations after a clash between border patrols, but they stalled in early 1957 when China would not accede to Burma's demands concerning four disputed areas.⁵⁸

53. Gompo Tashi Andrugstang, *Four Rivers, Six Ranges: Reminiscences of the Resistance Movement in Tibet* (Dharamsala: Information and Publicity Office of H.H. The Dalai Lama, 1973); and Tsering Shakya, *The Dragon in the Land of the Snows: A History of Modern Tibet since 1947* (New York: Penguin Compass, 1999), pp. 170–218.

54. Although A. Tom Grunfeld estimates that the number of rebels ranges from 50,000 to 200,000, Chinese sources indicate that the number ranges from 23,000 to 87,000. See Grunfeld, *The Making of Modern Tibet*, rev. ed. (Armonk, N.Y.: M.E. Sharpe, 1996), pp. 134–135; Li Peisheng and Li Guozhen, eds., *Pingxi xizang panluan* [Suppressing the Tibetan rebellion] (Lasa: Xizang renmin chubanshe, 1995), p. 17; Jamyang Norbu, "The Tibetan Resistance Movement and the Role of the CIA," in Robert Barnett, ed., *Resistance and Reform in Tibet* (London: Hurst and Company, 1994), p. 189; and Yang Qiliang, *Wang Shangrong jiangjun* [General Wang Shangrong] (Beijing: Dangdai zhongguo chubanshe, 2000), p. 403.

55. The PLA began to garrison the entire border only after the 1959 revolt. See Jiang Siyi and Li Hui, eds., *ZhongYin bianjing ziwai fanji zuozhan shi* [History of China's war of self-defensive counterattack against India] (Beijing: Junshi kexue chubanshe, 1994), p. 86; and Wu Lengxi, *Shinian lunzhan: 1956–1966 ZhongSu guanxi huiyilu* [Ten years of polemics: A recollection of Sino-Soviet relations from 1956 to 1966] (Beijing: Zhongyang wenxian chubanshe, 1999), p. 212. Author interview, Beijing, February 2003.

56. Jin Chongji, ed., *Zhou Enlai zhuan* [Biography of Zhou Enlai] (Beijing: Zhongyang wenxian chubanshe, 1998), Vol. 3, p. 1492.

57. S.D. Muni, *Foreign Policy of Nepal* (New Delhi: National Publishing House, 1973), p. 104.

58. Jin, *Zhou Enlai zhuan*, pp. 1229–1324; Dorothy Woodman, *The Making of Burma* (London: Cresset Press, 1962), pp. 535–536; and Yao Zhongming, "Zhou Enlai zongli jieju ZhongMian bianjie wenti de guanghui yeji" [Premier Zhou Enlai's glorious achievement in settling the Sino-

The Tibetan revolt, however, increased the cost to China of maintaining disputes with India, Nepal, and Burma. The main focus of the regime's response to the rebellion was a brutal pacification campaign combined with political reform designed to penetrate Tibetan society and implement direct rule.⁵⁹ But achievement of these domestic goals required peaceful borders and stable ties with neighboring states. The presence of territorial disputes threatened to complicate consolidation of central authority in Tibet because neighbors might provide support for rebels or even seek to intervene in the conflict. As PLA units pursued rebels fleeing to the south, they encountered Indian troops and clashed violently on August 25, 1959, near Longju. At the time, Chinese leaders believed that India was harboring rebel groups and allowing them to operate from bases in India.⁶⁰

CHINA COMPROMISES WITH BURMA, NEPAL, AND INDIA

In response to growing tensions on the border, China moved to compromise in its territorial disputes with Tibet's neighbors. Although India was the most important state, compromise was also pursued with Nepal and Burma to stabilize portions of Tibet's border and help to pressure India to reach a compromise agreement with China. Immediately following the Longju clash, Zhou ordered an investigation of the incident. On September 8, Mao convened a Politburo meeting, whose participants agreed that China should prepare to seek a negotiated settlement to all aspects of its border dispute with India.⁶¹ After the meeting, China first approached Nepal and Burma to open negotiations. On September 24, Zhou wrote to Burmese Prime Minister Ne Win, stating his interest in reopening talks based on earlier correspondence between the two sides.⁶² On October 9, Zhou met with a visiting Nepalese minister and expressed China's willingness to enter into border talks.⁶³ After a second armed

Burmese border problem], in Pei Jianzhang, ed., *Yanjiu Zhou Enlai: Waijiao sixiang yu shijian* [Studying Zhou Enlai: Diplomatic thought and practice] (Beijing: World Affairs Press, 1989), pp. 94–111.

59. Li and Li, *Pingxi xizang panluan*; and Yang Chengwu, "Xizang panluan" [The Tibetan rebellion], in *Zongcan moubu: huiyi shiliao* [General Staff Department: Recollections] (Beijing: Jiefangjun chubanshe, 1997), pp. 532–548.

60. Maxwell, *India's China War*.

61. Wu, *Shinian lunzhan*, p. 212.

62. On Ne's June 4, 1959, letter and Zhou's last offer of June 1958, see Woodman, *The Making of Burma*, p. 536.

63. Li Ping and Ma Zhisun, eds., *Zhou Enlai nianpu (zhong)* [A chronicle of Zhou Enlai's life] (Beijing: Zhongyang wenxian chubanshe, 1997), Vol. 2, p. 260; and Leo E. Rose, *Nepal: Strategy for Survival* (Berkeley: University of California Press, 1971), p. 223.

clash occurred between Chinese and Indian forces in late October 1959, Zhou offered to hold talks with Nehru. In December, Nehru still refused to meet with Zhou.

In January 1960, the Politburo Standing Committee (PBSC) discussed the Chinese-Indian dispute during a ten-day working meeting in Hangzhou. PBSC members agreed that the border with India should be settled swiftly through negotiations based on the principle of "give and take" (*huliang hurang*). According to one participant's recollection, the PBSC agreed that "China should make some concessions [and] India should make some concessions."⁶⁴ The PBSC also agreed to adopt a similar approach on China's other disputed borders and decided that China should seek a quick settlement with Burma and Nepal.

Official documents link these high-level decisions to the internal situation in Tibet. Newly available General Staff Department regulations from May 1960 outlined the leadership's general policy direction for stabilizing Tibet: "To stabilize our southwestern border region quickly, we must not only bring stability to the interior (*neibu*) but also to the exterior (*waibu*)."⁶⁵ An internal PLA bulletin stressed the same theme: "The PLA must work hard to make our southwest and northwest regional borders peaceful and secure. This is the best way to settle the problem of our [frontier] regions."⁶⁶

Although India remained its most important negotiating partner, China sought first to reach agreements with Burma and Nepal. In early January 1960, Zhou invited Prime Minister Ne Win to visit Beijing. On January 28, 1960, the two sides signed a preliminary agreement, where China acceded to many of Burma's demands from early 1957.⁶⁷ In early February, perhaps persuaded by China's agreement with Burma, Nehru agreed to meet with Zhou. In early March, before Zhou traveled to New Delhi, Nepalese Prime Minister Bishweshwar Koirala visited Beijing for negotiations in which China and Nepal reached a preliminary agreement to settle their dispute based on the customary line of control.⁶⁸ During and after these talks, China also gained

64. Wu, *Shinian lunzhan*, p. 248.

65. Jiang and Li, *ZhongYin bianjing ziwei fanji zuozhan shi*, p. 458.

66. Quoted in J. Chester Cheng, ed., *Politics of the Chinese Red Army* (Stanford, Calif.: Hoover Institution, 1966), p. 191.

67. Jin, *Zhou Enlai zhuan*, pp. 1929–1324; and Woodman, *The Making of Burma*, pp. 535–536.

68. Arthur Lall, *How Communist China Negotiates* (New York: Columbia University Press, 1968), pp. 199–200; and Yang Gongsu, *Zhongguo dui waiguo qinlue ganshe xizang difang douzheng shi* [History of China's struggle against foreign aggression and intervention in Tibet] (Beijing: Zangxue chubanshe, 1992), pp. 322–325.

other assistance in securing its southwestern frontier, including agreements with Nepal to regulate the movement of people across the border and to increase border trade, as well as permission to conduct limited military operations on Nepalese territory against Tibetan rebels and in Burma against remnant Nationalist troops.⁶⁹

Before departing for New Delhi, Zhou personally drafted a plan for his talks with Nehru. In his most optimistic scenario, Zhou hoped to reach an agreement that was “the same as with Burma and Nepal,” namely one based on compromise.⁷⁰ In those agreements, the acceptance of the McMahon Line with Burma and affirmation of the implied direction of the line with Nepal clearly suggested that China would also accept the line as the boundary in the eastern sector with India, which would have addressed India’s largest concern.⁷¹ Newly available sources indicate that, during the talks, Zhou proposed a territorial swap with Nehru. In their sixth meeting, Zhou offered to recognize India’s position in the eastern sector if India accepted China’s sovereignty over the Aksai Chin area in the west.⁷²

The April 1960 talks between Zhou and Nehru failed spectacularly. Nevertheless, China’s attempt to compromise in its dispute with India supports the logic of regime insecurity. The outbreak of rebellion in Tibet dramatically increased the cost of disputing territory with Burma, Nepal, and India. By seeking to focus on the consolidation of authority in Tibet and securing its borders, China dropped any large or inflexible claims it might have pressed against its Himalayan neighbors. Domestic priorities for political stability in Tibet trumped competition over disputed territory. Without the Tibetan rebellion in 1959, China would probably not have even entered into negotiations with these states, much less offered significant territorial concessions.

External balancing offers the main alternative explanation for China’s willingness to compromise following the revolt in Tibet. Based on balance of threat

69. John W. Garver, *Protracted Contest: Sino-Indian Rivalry in the Twentieth Century* (Seattle: University of Washington Press, 2001), p. 148; Han and Tan, *Dangdai zhongguo jundui de junshi gongzuo (shang)*, pp. 369–379; and Yang Gongsu, *Zhongguo dui waiguo qinlue ganshe xizang difang douzheng shi*, pp. 324–325.

70. Zhou Enlai quoted in Li and Ma, *Zhou Enlai nianpu (zhong)*, p. 302; and Shi Zhongquan, *Zhou Enlai de zhuoyue gongxian* [Zhou Enlai’s great contributions] (Beijing: Zhongyang wenxian chubanshe, 1993), p. 393.

71. The McMahon Line serves as roughly the line of control between China and India in the eastern sector. Britain drew the line during the Simla Conference on Tibet in 1913–14 attended by Chinese and Tibetan representatives, but it was not ratified by the Chinese government. See Maxwell, *India’s China War*, pp. 46–56.

72. Li and Ma, *Zhou Enlai nianpu (zhong)*, pp. 304–314.

theory, this explanation asserts that China settled not to consolidate its domestic position within Tibet, but to balance against external threats from neighboring states to its influence in the region.⁷³ Despite the emerging Chinese-Soviet split, China's compromise attempts in the early 1960s were not designed to improve China's position vis-à-vis the Soviet Union. China's initial offers to compromise with Burma, Nepal, and India occurred in late 1959 and early 1960 when relations between China and the Soviet Union remained cordial despite disagreement over policy toward the United States and other issues that surfaced in 1958. Most important, China's efforts to compromise from January to April 1960 occurred before the July 1960 decision to withdraw the remaining Soviet experts working in China and the Soviet decision to provide India with limited military assistance in the middle of 1961.⁷⁴

Likewise, China's attempts to compromise with Burma and Nepal were not part of an effort to balance against India. Although India was more threatening to China after 1959, this was not due to increased Indian diplomatic activity in the region or a rise in Indian power. Instead, China's leaders viewed India as threatening because of their own insecurity in Tibet and their beliefs about India's ambitions in the area. More generally, if China was seeking to counter India's growing influence in the region, territorial compromise with India probably would not have furthered this objective. If China and India were really only competing for influence over buffer states such as Nepal and to a lesser extent Burma, China probably would have adopted a much more assertive policy in the territorial dispute. Instead, China's leaders were countering internal threats to territorial integrity, which they believed that India could exacerbate or exploit.

The Great Leap Forward

In the spring of 1962, China faced renewed unrest in the frontiers during a period of intense regime instability following the failure of the Great Leap Forward (1958–60).⁷⁵ This combination of internal threats to territorial integrity and political stability increased the cost of maintaining frontier disputes with

73. See, for example, Hyer, "The Politics of China's Boundary Disputes and Settlements"; and Watson, *The Frontiers of China*.

74. Allen S. Whiting, *The Chinese Calculus of Deterrence: India and Indochina* (Ann Arbor: University of Michigan Press, 1975), p. 73.

75. The Great Leap Forward was a campaign that used mass mobilization techniques, including the collectivization of agriculture, to achieve rapid economic growth.

its neighbors. China pursued compromise in disputes with North Korea, Mongolia, India, Pakistan, Afghanistan, and the Soviet Union to recover from the economic crisis by easing external tensions.

POLITICAL INSTABILITY IN CHINA PROPER

In early 1962, the dramatic failure of the Great Leap Forward and the devastating famine that it unleashed posed a serious threat to China's political stability. A combination of misguided agricultural policies, natural disasters, and bureaucratic incompetence led to the starvation of somewhere between 20 and 30 million people from 1959 to 1962. Grain production in 1960 was only 75.5 percent of that in 1958. Industrial output also decreased significantly.⁷⁶ China's gross domestic product declined by 27 percent in 1961 alone.⁷⁷ In response, China's leaders focused on rebuilding the economy and consolidating the regime.⁷⁸ Retrenchment became official policy at the Ninth Plenum in January 1961 with the slogan "readjustment, consolidation, filling out, and raising standards." Mao Zedong withdrew from day-to-day policymaking as Liu Shaoqi and Deng Xiaoping moved to oversee reconsolidation efforts. To relieve pressure in urban areas, and presumably to quell potential unrest, the Central Committee decided to reduce China's urban population by 25 million through rustication.⁷⁹

At the same time, China also faced renewed internal threats to its territorial integrity. From April to May 1960, more than 60,000 Kazakhs fled from Xinjiang to the Soviet Union, and violent protests erupted in Yining, a town near the border. At the time, China faced gaping vulnerabilities because Xinjiang's border was basically undefended—the PLA maintained only eight frontier defense stations and a total of 110 soldiers on this boundary.⁸⁰ Moreover, recent archival research demonstrates that Chinese leaders believed that Moscow was provoking ethnic unrest in the area by issuing false Soviet iden-

76. Kenneth Lieberthal, "The Great Leap Forward and the Split in the Yanan Leadership," in John King Fairbank and Roderick MacFarquhar, eds., *The Cambridge History of China*, Vol. 14: *The Emergence of Revolutionary China, 1949–1965* (Cambridge: Cambridge University Press, 1987), pt. 1, p. 318.

77. World Development Indicators, various years, <http://devdata.worldbank.org/dataonline>.

78. Roderick MacFarquhar, *The Origins of the Cultural Revolution*, Vol. 3: *The Coming of the Cataclysm, 1961–1966* (New York: Columbia University Press, 1997), pp. 1–121.

79. Personal communication with Lei Guang, May 2002.

80. *Gongan budui: zongshu, dashiji, biaoce* [Public security troops: Summary, chronology, statistics] (Beijing: Jiefangjun chubanshe, 1997), p. 78.

tity papers to Uighurs and Kazakhs, broadcasting propaganda encouraging them to flee, and facilitating the flow of people across the border.⁸¹

As the domestic situation deteriorated, external challenges to China's territorial integrity exacerbated this regime insecurity. In early 1962, India executed a "forward policy" of establishing posts in disputed areas to challenge China's territorial claims. Although the outposts themselves were not especially threatening, they were viewed as a renewed challenge to China's control of Tibet because they occurred just as China's leaders had completed the pacification campaign and consolidated control of the region. As Zhou explained in 1960, the Indian leadership "want[s] Tibet . . . as a 'buffer state' under India's influence, becoming its protectorate."⁸² In March and April 1962, Taiwan's leader, Chiang Kai-shek, mobilized his forces for an assault on the mainland. China's leaders correctly concluded that Chiang sought to profit from instability on the mainland, which increased concerns about territorial integrity and frontier stability more broadly.⁸³

Taken together, these events greatly increased China's costs of maintaining territorial disputes along the frontier. To support regime consolidation and economic rebuilding, China pursued a moderate foreign policy and sought to avoid territorial conflicts and tensions on three fronts. In addition, government documents connect this foreign policy orientation with the domestic situation. In January 1962, for example, three foreign policy advisers in a letter to Zhou Enlai and other senior leaders linked the recovery from the economic crisis with a relaxation of tensions abroad.⁸⁴

CHINA COMPROMISES WITH NORTH KOREA, MONGOLIA, PAKISTAN,
AFGHANISTAN, AND THE SOVIET UNION

As regime insecurity grew in April and May, China moved to open talks with North Korea, Mongolia, and Pakistan. Previously, in 1959 and 1961, it had declined Pakistani attempts to discuss their disputed border. But on February 28,

81. Li Danhui, "Dui 1962 nian Xinjiang Yita shijian qiyan de lishi kaocha (xu) [Historical investigation of the origins of the 1962 Yita incident in Xinjiang (cont.)]," *Dangshi yanjiu ziliao*, No. 5 (1999), pp. 1–8.

82. *Zhou Enlai waijiao wenxuan* [Zhou Enlai's selected works on diplomacy] (Beijing: Zhongyang wenxian chubanshe, 1990), p. 270.

83. See, for example, Wang Shangrong, "Xin zhongguo jiansheng hou jici zhongda zhanzheng" [Several significant wars after the emergence of new China], in Zhu Yuanshi, ed., *Gongheguo yaoshi koushushi* [Oral histories of important events of the republic] (Changsha: Hunan renmin chubanshe, 1999), pp. 277–278.

84. Xu Zehao, ed., *Wang Jiaxiang nianpu* [A chronicle of Wang Jiaxiang's life] (Beijing: Zhongyang wenxian chubanshe, 2001), pp. 488–489.

1962, Beijing finally agreed to consider talks, which were announced in May.⁸⁵ Likewise, in 1957 and 1960, China had rebuffed Mongolian efforts to open talks. On April 13, 1962, however, it moved to hold negotiations with Mongolia and proposed measures to maintain peace on the border in light of recent tensions.⁸⁶ When Mongolia did not reply, Zhou Enlai met with the Mongolian ambassador in early May to press for talks.⁸⁷ During this same period, interviews indicate that China also pushed for talks with North Korea, which Beijing had refused to hold in the early 1950s and in 1960.⁸⁸ In June 1962, Zhou met with the North Korean ambassador to discuss their border dispute.⁸⁹

Over the summer of 1962, tensions on the Chinese-Indian border continued to escalate. As India expanded the scope of the forward policy, China resumed patrolling in disputed areas. In late July, after a clash in the Chip Chap valley, China used multiple diplomatic channels to open talks with India, repeating these efforts in September and early October. When India would not agree to unconditional talks, China launched a military campaign to bring it to the negotiating table.⁹⁰ During this same period, China began to hold talks with North Korea, Mongolia, and Pakistan, reaching draft or final agreements by the end of the year. In all cases, China offered significant concessions, giving North Korea most of the disputed Changbai (Paekdu) mountain area, conceding roughly 12,000 square kilometers to Mongolia, and transferring 1,942 square kilometers to Pakistan.⁹¹

In the short term, the timing of these agreements was linked to China's failed efforts to compel India to negotiate and the need to secure restive fron-

85. Mujtaba Razvi, *The Frontiers of Pakistan: A Study of Frontier Problems in Pakistan's Foreign Policy* (Karachi-Dacca: National Publishing House, 1971), pp. 166–193; and Wang Taiping, ed., *Zhonghua renmin gongheguo waijiao shi, 1957–1969* [Diplomatic history of the People's Republic of China, 1957–1969] (Beijing: Shijie zhishi chubanshe, 1998), pp. 102–104.

86. Wang, *Zhonghua renmin gongheguo waijiao shi, 1957–1969*, p. 101.

87. Li and Ma, *Zhou Enlai nianpu (zhong)*, p. 477. See also Cai, *Bianfang lilun*, p. 11; and Han Nianlong, ed., *Dangdai Zhongguo waijiao* [The diplomacy of contemporary China] (Beijing: Zhongguo shehui kexue chubanshe, 1988), p. 154.

88. Author interview, Beijing, June 2002.

89. Li and Ma, *Zhou Enlai nianpu (zhong)*, p. 486.

90. China wanted to hold unconditional talks on all disputed areas, whereas India only agreed to hold talks on the western sector if China withdrew from all areas claimed by India. See Whiting, *The Chinese Calculus of Deterrence*.

91. On Pakistan, see Razvi, *The Frontiers of Pakistan*, pp. 175–177. On North Korea, see Cai, *Bianfang lilun*, p. 9. On Mongolia, see Henry S. Bradsher, "Sino-Soviet Rift Catches Mongolia," *Christian Science Monitor*, January 6, 1964; and Wang, *Zhonghua renmin gongheguo waijiao shi, 1957–1969*, p. 102.

tiers during a period of regime instability. In addition, however, the successful conclusion of these agreements helped China to prepare for talks with the Soviets, which China's leaders had decided to pursue in the spring of 1962.⁹² Similar to China's agreements with Burma and Nepal in 1960, treaties with North Korea and Mongolia, as "fraternal" socialist neighbors, were concluded in part to persuade Moscow to reach a similar agreement with Beijing. In December 1962, for example, the *People's Daily* opined that the Chinese-Mongolian agreement "set a good example for socialist countries in handling their mutual relations."⁹³

In April 1963, after announcing negotiations with Afghanistan, China approached the Soviet Union to hold talks. After several months of correspondence, the two sides met in February 1964. Newly released sources indicate that both sides achieved far more progress in settling their differences than was previously known.⁹⁴ In particular, the two sides reached a consensus concerning the eastern sector of the border along the Amur and Ussuri Rivers that was almost identical to the agreement signed in 1991, which gave China control of 52 percent of the contested islands totaling roughly 1,000 square kilometers.⁹⁵ The two sides also began discussions on the western sector in Soviet Central Asia. Negotiators adjourned in August 1964 and planned to resume talks in Moscow on October 15, 1964, the day after Nikita Khrushchev was sacked. Soviet leadership change and Mao's famous comments to members of the Japanese Communist Party in July 1964 that China had "yet to settle [its] account" for territories ceded to czarist Russia prevented the resumption of talks. Although Mao quickly claimed that he was only "firing empty canons," Soviet leaders concluded that China harbored irredentist claims despite the draft agreement and earlier statements in the 1964 talks that it would accept the boundaries of these unequal treaties.⁹⁶

Similar to compromise attempts after the Tibetan revolt, external balancing

92. China did not pursue talks immediately because it sought to undertake additional archival research on the historical changes of the border and strengthen the PLA's frontier defense stations before negotiations commenced. Author interview, Beijing, July 2001.

93. *Renmin ribao* [People's daily], December 27, 1962.

94. Li Lianqing, *Lengnuan sui yue: yi bolan de ZhongSu guanxi* [Hot and cold times: The twists and turns of Chinese-Soviet relations] (Beijing: Shijie zhishi chubanshe, 1999), pp. 323-324; Tang Jiaxuan, ed., *Zhongguo waijiao cidian* [Dictionary of China's diplomacy] (Beijing: Shijie zhishi chubanshe, 2000), p. 725; and Wang, *Zhonghua renmin gongheguo waijiao shi, 1959-1969*, pp. 254-256. Author interviews, Beijing, July 2001-June 2002.

95. Tang, *Zhongguo waijiao cidian*, p. 725.

96. Yang Kuisong, "The Sino-Soviet Border Clash of 1969: From Zhenbao Island to Sino-American Rapprochement," *Cold War History*, Vol. 1, No. 1 (August 2000), p. 23.

offers a persuasive alternative explanation for China's behavior in 1962. This explanation asserts that China's many compromises in 1962 reflected concerns about the increasing threat posed by deteriorating relations with Moscow. Nevertheless, China did not seek to settle with these same states during earlier periods of tension with the Soviet Union, especially after the withdrawal of advisers in 1960. Chinese-Soviet ties were relatively stable in 1961, as economic cooperation continued and the sale of MiG fighter aircraft was broached. Moreover, China also held substantive negotiations with the Soviet Union, which is inconsistent with balancing behavior. If China were balancing against the Soviet Union, its best strategy might have been to strengthen its border security, while downplaying the presence of a dispute to bide time to improve its military position. This is precisely what China did after the open split in 1964 with the rapid militarization of the Chinese-Soviet border. Finally, many of these states offered little leverage to China in its dispute with the Soviet Union. Pakistan was a member of the Southeast Asian Treaty Organization, an anti-communist alliance, while Afghanistan was not part of the Soviet orbit at the time. Moreover, Chinese-Mongolian ties had deteriorated in the mid-1950s, well before China had moved to settle the dispute, while North Korea was firmly on China's side in its ideological competition with the Soviet Union.

Tiananmen Upheaval

In 1989, the upheaval in Tiananmen Square posed an internal threat to the stability of China's communist regime. This legitimacy crisis, exacerbated by the weakening of other socialist governments worldwide, increased the cost of maintaining territorial disputes with the Soviet Union, Laos, and Vietnam. China offered concessions in these disputes in exchange for cooperation to counter diplomatic isolation and ensure the continuation of Deng Xiaoping's domestic reform agenda.

CRISIS AT THE CORE

The end of the Cold War had a paradoxical effect on how China's leaders perceived their security. The decline and then disintegration of the Soviet Union greatly increased China's territorial security, removing any serious threat to its northern border and eliminating potential encirclement by Soviet forces based in Vietnam. Moreover, this appeared to present an ideal moment for China to

regain territory it had long disputed with the Soviet Union and its republics, which were in no position to resist Chinese pressure.

Nevertheless, the end of the Cold War ushered in a period of unprecedented regime insecurity for China's leaders. The demonstrations in Tiananmen highlighted popular dissatisfaction with the CCP's leadership and the state of reform, while the violent crackdown and subsequent suppression further alienated China's leaders from its people.⁹⁷ External trends aggravated these internal sources of insecurity.⁹⁸ Sanctions and the rapid deterioration of relations with the United States and other Western states threatened to undermine China's opening and reform policy, which required access to foreign capital, technology, and markets. The collapse of socialism in Eastern Europe and then the disintegration of the Soviet Union underscored China's vulnerability to internal change and left China's leaders with few diplomatic sources of support for their increasingly embattled regime.

In response, paramount leader Deng Xiaoping led an effort to consolidate the position of the CCP through maintaining domestic stability and continuing economic reform. To support these goals, Deng crafted a moderate foreign policy based in part on improving ties with regional states to counter China's diplomatic isolation and maintain access to markets and sources of investment.⁹⁹ Deng instructed in September 1989 that China's leaders "should be calm . . . quietly immerse ourselves in practical work and to accomplish something, something for China."¹⁰⁰ In the months that followed, this became known as the 20-character policy: "Observe things coolly, deal with things calmly, keep a firm footing, hide our capacities and bide our time, get some things done" [*lengjing guancha, chenzhuo yingfu, wenzhu zhenjiao, taoguang yanghui, yousuo zuowei*].¹⁰¹ Improving ties with neighboring states through ter-

97. Chinese sources refer explicitly to the concept of political security in contrast to territorial security and economic security. See, for instance, Yan Xuetong, "Shiyan Zhongguo de anquan huanjing [Preliminary analysis of China's security environment]," *Dangdai guoji wenti yanjiu*, No. 4 (1994), pp. 35–41.

98. Richard Baum, *Burying Mao: Chinese Politics in the Age of Deng Xiaoping* (Princeton, N.J.: Princeton University Press, 1994), pp. 247–310.

99. See, for example, Zhang Baijia, "Jiushi niandai de zhongguo neizheng yu waijiao" [China's domestic politics and diplomacy in the 1990s], *Zhonggong dangshi yanjiu*, No. 6 (2001), pp. 29–34.

100. *Deng Xiaoping wenxuan* [Deng Xiaoping's selected works] (Beijing: Renmin chubanshe, 1993), Vol. 3, p. 321.

101. For an authoritative statement of these phrases, see Qian Qichen's entry in Wang Taiping, ed., *Deng Xiaoping waijiao sixiang yanjiu lunwen ji* [Collected papers on the study of Deng Xiaoping's diplomatic thought] (Beijing: Shijie zhishi chubanshe, 1996).

ritorial compromise, especially other socialist states, was part of this broader strategy. As Foreign Minister Qian Qichen remarked in 1990, "Diplomacy is the extension of internal affairs."¹⁰²

CHINA COMPROMISES WITH THE SOVIET UNION, LAOS, AND VIETNAM

China's most important compromise during this period was with the Soviet Union. The impetus for the 1991 agreement, however, came not from Beijing but from Moscow. In 1986, President Mikhail Gorbachev stated that the Soviet Union would accept the *thalweg* principle along the eastern sector of the border, thereby acceding to one of China's long-standing demands and signaling a return to the terms of the 1964 draft agreement.¹⁰³ Before the Tiananmen crisis, China insisted upon achieving a package deal that would settle disputes along the eastern and western sectors of the Chinese-Soviet border. After 1989, however, China agreed to pursue separate agreements because improved relations with the Soviet Union had become much more important. In 1990, the pace of talks quickened, with two meetings of the joint working group as well as meetings of a border survey group and an agreement drafting group.¹⁰⁴ In the 1991 eastern sector agreement, China received approximately 52 percent of the disputed areas.¹⁰⁵ China offered further concessions when disputes arose during the demarcation process.¹⁰⁶

China also compromised in frontier disputes with its other socialist neighbors, Laos and Vietnam. By compromising with these states, China sought not only to normalize relations and thus strengthen neighboring socialist regimes, but also to facilitate the economic development of the frontiers. China and Laos opened talks in early 1990 and signed an agreement to divide evenly their disputed areas in 1991. China and Vietnam opened negotiations in 1991 and reached a preliminary agreement in 1993 on principles for settling their dispute. The animosity created by the 1979 China-Vietnam war and a decade

102. Quoted in *Beijing Review*, October 15–21, 1990, p. 12.

103. The Soviet decision is consistent with the logic of regime insecurity because the growing Soviet economic crisis prompting glasnost and perestroika increased the cost of strained relations with China.

104. *Zhongguo waijiao gailan* [Survey of China's diplomacy] (Beijing: Shijie zhishi chubanshe, 1991), p. 235.

105. Tang, *Zhongguo waijiao cidian*, p. 725.

106. Akihiro Iwashita, *A 4,000 Kilometer Journey along the Sino-Russian Border* (Sapporo: Slavic Research Center, 2004). Prior to the collapse of the Soviet Union, China had considered dropping its claim to the Pamir mountains in exchange for the return of Black Bear Island. Author interviews, Beijing, June–July 2001.

of border clashes delayed reaching a final agreement evenly dividing the disputed areas until 1999. Obstacles to a final settlement included the need for massive mine-sweeping operations along the border as well as internal resistance on both sides from those who opposed making concessions to a former adversary.¹⁰⁷ During this period, China signed agreements to increase development of frontier regions through cross-border trade.

At the same time, China also moved to improve ties with India. In 1991, a joint working group established in 1989 began substantive discussions, which led to the signing of the confidence-building agreements in 1993 and 1996.¹⁰⁸ Both agreements helped to stabilize the border, improving relations and reducing potential for inadvertent escalation between the two sides. In 1996, China reportedly again offered a compromise based on exchanging positions in the eastern and western sectors.¹⁰⁹ Similar to Vietnam, the legacy of war and internal opposition posed serious obstacles to a final settlement. The 1993 and 1996 confidence-building measures, however, have established an effective boundary along the Line of Actual Control, giving China most of what it sought—improved ties with an important neighbor to create breathing space for internal consolidation and facilitate regional trade.

External balancing provides an unconvincing alternative explanation for China's behavior during this period: that China sought to balance against increasing U.S. power after the Cold War, not bolster the security of its socialist regime. Two factors weigh against this interpretation. During this period, China's security policy, including defense spending, military ties, and military training, did not keep pace with the dramatic shift in the global balance of power after 1991.¹¹⁰ If China's compromise attempts were the result of balancing against the United States, they were exceptional from the perspective of Chinese security policy at the time. Second, China's compromises with the Soviet Union preceded its collapse by several years. Although the initiative for

107. For example, the PLA reportedly withdrew its members from the border joint working group in protest for the concessions that were made. Author interview, Beijing, July 2002.

108. Waheguru Pal Singh Sidhu and Jing-dong Yuan, "Resolving the Sino-Indian Border Dispute: Building Confidence through Cooperative Monitoring," *Asian Survey*, Vol. 41, No. 2 (March/April 2001), pp. 351–376.

109. "Opportunity to Resolve Border Dispute with China Missed," *Telegraph* (Calcutta), May 15, 1997, in FBIS FTS-1997-0515-001285.

110. Alastair Iain Johnston, "Realism(s) and Chinese Security Policy in the Post-Cold War World," in Ethan B. Kapstein and Michael Mastanduno, eds., *Unipolar Politics: Realism and State Strategies after the Cold War* (New York: Columbia University Press, 1999), pp. 261–318.

these talks stemmed from Moscow, the pace quickened dramatically after the Tiananmen crackdown but before the Soviet collapse. The 1991 Chinese-Soviet draft agreement was concluded in 1990, before the full extent of Soviet weakness was known, and signed in April 1991, before the August revolution that set in motion the collapse of the Soviet Union. Likewise, China compromised in its dispute with Laos and opened talks with Vietnam before the disintegration of the Soviet Union.

Xinjiang Unrest

Soon after the Tiananmen upheaval, ethnic unrest in Xinjiang posed a new internal threat to the PRC's territorial integrity. The demonstrations and armed uprisings increased the cost of maintaining territorial disputes with Xinjiang's neighbors, Kazakhstan, Kyrgyzstan, and Tajikistan. China moved to compromise in these disputes in exchange for cooperation to reduce external support for separatists.

INSTABILITY IN THE NORTHWESTERN FRONTIER

In the early 1990s, ethnic unrest erupted in Xinjiang, compounding the legitimacy crisis after Tiananmen. Mounting grievances among Uighur and other ethnic minorities against the central government resulted in an unprecedented number of demonstrations, bombings, assassinations, and armed clashes with security forces throughout the 1990s.¹¹¹ Although the scale of unrest was small compared with the Tibetan revolt, it was nevertheless a source of considerable concern for China's leaders, especially after Tiananmen. The ethnic unrest in Xinjiang underscored the breadth of dissatisfaction with the regime in frontier areas at a time when it faced instability and discontent among its core constituents.

China's response to this unrest included a comprehensive effort to engage the newly independent states in Central Asia. For China's leaders, territorial disputes had become much more costly to maintain because they blocked as-

111. CCP Organization Department, ed., *Zhongguo diaochaogao: xin xingshi xia renmin neibu maodun yanjiu* [China investigative report: Research on internal contradictions of the people under new circumstances] (Beijing: Zhongyang bianyi chubanshe, 2001), pp. 243–281; Man Zongzhou, "Xinjiang minzu fenlie zhuyi de houdong ji fan minzu fenlie douzheng yingdang zhuyi de wenti [Issues to note in Xinjiang's ethnic separatist activities and the antiseparatist struggle]," *Xinjiang sheke luntan*, No. 3 (1996), pp. 43–45; and James Millward, "Violent Separatism in Xinjiang: A Critical Assessment," Policy Studies No. 6 (Washington, D.C.: East-West Center Washington, 2004).

sistance that China sought to maintain stability within Xinjiang. In particular, China needed cooperation with its neighbors to prevent the spread of pan-Islamic and pan-Turkic forces to the region, limit external support for separatists within Xinjiang, and increase cross-border trade as part of a broader strategy to reduce tensions among ethnic groups through development. Attempts to settle outstanding territorial disputes were an important component of China's engagement of the region. Facing renewed internal unrest, China's leaders valued stable and secure borders more than ever before. For example, a leaked document from the CCP's Central Committee on China's Xinjiang policy instructed the government to secure the border and use diplomacy to "urge these countries to limit and weaken the activities of separatist forces inside their border."¹¹²

CHINA COMPROMISES WITH KAZAKHSTAN, KYRGYZSTAN, AND TAJIKISTAN

The collapse of the Soviet Union might have presented China with an ideal opportunity to regain the more than 34,000 square kilometers of territory it claimed in Central Asia. In the context of ethnic unrest, however, China chose to improve ties with the newly independent states to deny external support to separatist groups in Xinjiang. China first pursued compromise in its dispute with Kazakhstan, Xinjiang's largest neighbor, starting talks less than a year after the collapse of the Soviet Union. In the 1994, 1996, and 1997 agreements, China made significant concessions, holding roughly 22 percent of disputed areas.¹¹³ In 1996 and 1999, China made similar concessions in agreements with Kyrgyzstan, where it received about 32 percent of the disputed land.¹¹⁴ The civil war that erupted after the collapse of the Soviet Union, however, stalled negotiations with Tajikistan. Talks resumed in 1997, leading to a preliminary agreement in 1999 that divided one disputed sector evenly and a supplemental agreement in 2002. China made a large concession in this agreement, dropping most of its 28,000-square-kilometer claim to the Pamir mountains, a claim it had pressed since the late nineteenth century.¹¹⁵

112. Central Committee, "Document No. 7 (1996)," in Human Rights Watch, *China: State Control of Religion*, Update No. 1 (New York: Human Rights Watch, 1998).

113. Aleksey Gulyayev, "China-Kazakhstan: The Border Dispute Continues," *Izvestiya*, April 30, 1999, in FBIS FTS-1999-0504-001086; and Zhang Zhouxiang, *Xinjiang bianfang gaiyao* [Summary of Xinjiang's frontier defense] (Urmuchi: Xinjiang renmin chubanshe, 1999), p. 135.

114. Viktoriya Panfilova, "Is Kyrgyzstan Being Reduced in Size?" *Nezavisimaya gazeta*, May 16, 2001, in FBIS CEP-2001-0516-000097; and Zhang, *Xinjiang bianfang gaiyao*, p. 135.

115. In the 2002 supplemental agreement, China received only 1,000 of the 28,000 square kilome-

In return, China received assistance in eliminating external support for separatists within Xinjiang. In numerous public statements, Central Asian leaders pledged their help to China. During the signing of the 1997 border agreement, for example, President Askar Akayev stated that Kyrgyzstan “would not allow any force to make use of its soil to conduct activities against China.”¹¹⁶ Likewise, Kazakhstan repatriated suspected separatists to China, dissolved political parties, closed newspapers, and arrested suspected militants.¹¹⁷ China also used the border negotiations as well as demilitarization talks to lead the establishment of a new regional grouping, the Shanghai Cooperation Organization (SCO), which focuses partly on enhancing political stability of member states through trade, counterterrorism cooperation, and confidence-building measures.¹¹⁸

The timing of these compromises do not support arguments based on external threats, particularly efforts to balance against U.S. influence in the region. China signed many agreements well before the upsurge of U.S. interest in Caspian oil in the mid-1990s. While China is wary of a U.S. military presence in the region, before the terrorist attacks in New York and Washington, D.C., on September 11, 2001, this presence was limited to training exercises for peace-keeping operations held in 1997 and 2000. China did become much more concerned about the U.S. presence after the start of Operation Enduring Freedom in Afghanistan in October 2001, but by this time Beijing had signed all but one agreement resolving its territorial disputes with Xinjiang’s neighbors. Likewise, the joint statements issued at various summit meetings and through the SCO did not include many direct references to the United States and stressed instead problems associated with the “three forces” of terrorism, separatism, and extremism.¹¹⁹

ters of the Pamir mountain area. See “Tajikistan Transfers 1,000 km Territory to China,” May 21, 2002, Interfax, in FBIS CEP-2002-0521-000161.

116. “Chinese, Kyrgyzstan Presidents Sign Supplementary Border Agreement,” Xinhua News Agency, August 26, 1999, from Lexis-Nexis.

117. See, for example, Jean-Christophe Peuch, “Central Asia: Uighurs Say States Yield to Chinese,” Radio Free Europe, <http://www.rferl.org/nca/features/2001/03/29032001104726.asp>; “Uighur Separatist Exiles See End to Help from Central Asia,” Agence France-Presse, June 18, 1997, from Lexis-Nexis; and Human Rights Watch, *China: Human Rights Concerns in Xinjiang* (New York: Human Rights Watch, 2001).

118. Xu Tao and Li Zhiye, eds., *Shanghai Hezuo Zuzhi: Xin anquan guan yu xin jizhi* [Shanghai Cooperation Organization: New security concept and new mechanism] (Beijing: Shishi chubanshe, 2002).

119. *Zhongguo waijiao* [China’s diplomacy] (Beijing: Shijie zhishi chubanshe, various years).

Conclusion

Regime insecurity best explains China's pattern of cooperation and delay in its territorial disputes. China's leaders have compromised when faced with internal threats to regime security—the revolt in Tibet, the instability following the Great Leap Forward, the legitimacy crisis after the Tiananmen upheaval, and separatist violence in Xinjiang. The timing of compromise efforts, official documents, and statements by China's leaders demonstrate that internal threats, not external ones, account for why and when China pursued cooperation.

While further research is necessary, support for the logic of regime insecurity in China's many disputes serves as a plausibility probe for a more general application of the argument. In territorial disputes, similar domestic sources of compromise might arise under two sets of conditions. First, unrest near international borders should create incentives for leaders to cooperate with adjacent states, such as Iraq's 1975 decision to compromise in a dispute with Iran to crush a Kurdish rebellion.¹²⁰ Second, state leaders, especially authoritarian ones, should be more likely to compromise in their disputes when facing political instability or legitimacy crises, such as Argentina's decision to compromise in a dispute with Uruguay or Peru's attempts to settle its dispute with Ecuador before the outbreak of the border war in February 1995.¹²¹

China's cooperation in territorial disputes reveals a pattern of behavior far more complex than a singular view of China as a territorially ambitious state. China's growing military and economic power has not yet translated into increased territorial revisionism. In the past two decades, China has not issued new territorial demands nor increased the scope of land claimed despite its rising power. Since 1949, China has pursued revisionist claims in its homeland disputes and in offshore island disputes over the Spratlys and Senkakus, but these claims have remained constant despite increases in China's power. With the exception of the Senkakus, the PRC claimed all homeland areas and other offshore islands between 1949 and 1951. These claims reflect a vision of what ought to constitute a modern Chinese state, not ambitions conditioned by

120. Jasim M. Abdulghani, *Iraq and Iran: The Years of Crisis* (London: Croom Helm, 1984), p. 155.

121. Beth A. Simmons, "Territorial Disputes and Their Resolution: The Case of Ecuador and Peru," *Peaceworks* No. 27 (Washington, D.C.: United States Institute of Peace, 1999), pp. 11–12; and Joseph S. Tulchin, *Argentina and the United States: A Conflicted Relationship* (Boston: Twayne, 1990), p. 137.

China's position in the international system. China, weak or strong, has sought to unify Taiwan with the mainland.

In its other territorial disputes, China has pursued mostly status quo goals. The handling of the nineteenth-century unequal treaties that often ceded vast tracts of land reflects this complex approach to territory.¹²² In negotiations with neighbors since 1949, the PRC has accepted the general boundaries delimited by these very agreements. Not coincidentally, most of these boundaries lie adjacent to ethnic minority regions. The only past agreements that China has sought to overturn are those linked to the territories of Hong Kong and Macao, also not coincidentally the only Han Chinese areas that have been ceded. In addition, by drafting new boundary agreements with neighbors when settling territorial disputes, China has defined the precise location of most of its land borders. These texts, some of them hundreds of pages in length, remove any ambiguity about the extent of the PRC's sovereignty and raise the costs for pursuing future claims that would breach these agreements.

The prognosis for the settlement of China's six remaining disputes is uncertain. Despite the recent mainland visits of Taiwan's opposition leaders, the PRC is unlikely to abandon its long-standing goal of unification. In addition, the increasing strength of the state in the frontier regions suggests that regime insecurity may be less likely to create incentives for compromise in China's last two frontier disputes with India and Bhutan. As the April 2005 Chinese-Indian agreement on principles for settling their dispute demonstrates, external factors are likely to play a stronger role in this conflict's final settlement, even though the parameters of compromise were established decades ago. Offshore, China and the other claimants have little incentive to compromise over sovereignty because these island groups cost little to dispute and may yield economic or strategic advantages. In addition, the role of nationalism in Chinese foreign policy remains a wild card that might constrain the future ability of China's leaders to pursue compromise.¹²³

Nevertheless, the territorial settlements made possible by China's compromises have had important strategic effects in East Asia. China's settlements are linked to the absence of war with opposing states. China has not participated

122. On China's broader approach to sovereignty, see Allen R. Carlson, *Unifying China, Integrating with the World: Securing Chinese Sovereignty in the Reform Era* (Stanford, Calif.: Stanford University Press, 2005).

123. Nationalism, however, did not constrain China's ability to reach compromises in the 1990s and beyond, including the important 2002 supplemental agreement with Tajikistan.

in military conflict over contested areas with neighboring states with which it has settled territorial disputes.¹²⁴ China's compromises have also enabled the active engagement of the region since the late 1990s that is the hallmark of China's "new diplomacy."¹²⁵ By settling disputes and eliminating ambiguity about the location of its borders, China has reassured its neighbors about its intentions and potential ambitions, lessening the security dilemma. Regional diplomatic initiatives would have been much more difficult to pursue under the shadow of hot territorial conflicts, especially in light of ongoing tensions over Taiwan. China's continued need to engage East Asia also suggests that these settlements will endure in the coming years.

124. The one exception concerns clashes on the border with Burma during the Cultural Revolution involving Burmese communist insurgents. See "Burma Lists Toll at China's Border," *New York Times*, November 8, 1969.

125. Evan S. Medeiros and M. Taylor Fravel, "China's New Diplomacy," *Foreign Affairs*, Vol. 82, No. 6 (November/December 2003), pp. 22–35; and David Shambaugh, "China Engages Asia: Reshaping the International Order," *International Security*, Vol. 29, No. 3 (Winter 2004/05), p. 67.



FOREIGN POLICY RESEARCH INSTITUTE

E-NOTES

September 2013

A SALUTATION TO ARMS: Asia's Military Buildup, Its Reasons, and Its Implications

By Felix K. Chang



Felix K. Chang is an FPRI Senior Fellow, as well as the co-founder of Avenir Bold, a venture consultancy. He was previously a consultant in Booz Allen Hamilton's Strategy and Organization practice; among his clients were the U.S. Department of Energy, U.S. Department of Homeland Security, U.S. Department of the Treasury, and other agencies. Earlier, he served as a senior planner and an intelligence officer in the U.S. Department of Defense and a business advisor at Mobil Oil Corporation, where he dealt with strategic planning for upstream and midstream investments throughout Asia and Africa. His publications include articles in American Interest, National Interest, Orbis, and Parameters.

This year, Asian defense spending will surpass that of Europe for the first time in over half a millennium. A steep drop in European military expenditures after the Cold War, a concurrent and steady rise in Chinese expenditures, and a recent sharp increase in defense spending across the rest of Asia rapidly closed the gap between the two continents. Reasons vary for Asia's military buildup and many, whether right or wrong, have begun to darkly speculate about its implications for the region. But one thing seems certain: Asia's military buildup is no flash in the pan; it is likely to endure.

For the most part, China played down its increased military expenditures throughout the 1990s. But its defense spending was never as low as it claimed, nor probably as wisely spent. At the start of the decade, much of China's military budget was devoted to preparations for national mobilization and maintaining large standing conventional forces. But within a few years, China began to reallocate that budget, shifting resources from ground forces and pouring them into its navy and air force. Chinese ground forces were subjected to a series of deep cuts that demobilized over a million troops; the army shrank from over 120 division equivalents to fewer than 60 more heavily mechanized division equivalents by the end of the decade.¹ Just as significantly, the process also freed up resources that fueled the research, development, and acquisition of new weapon systems.

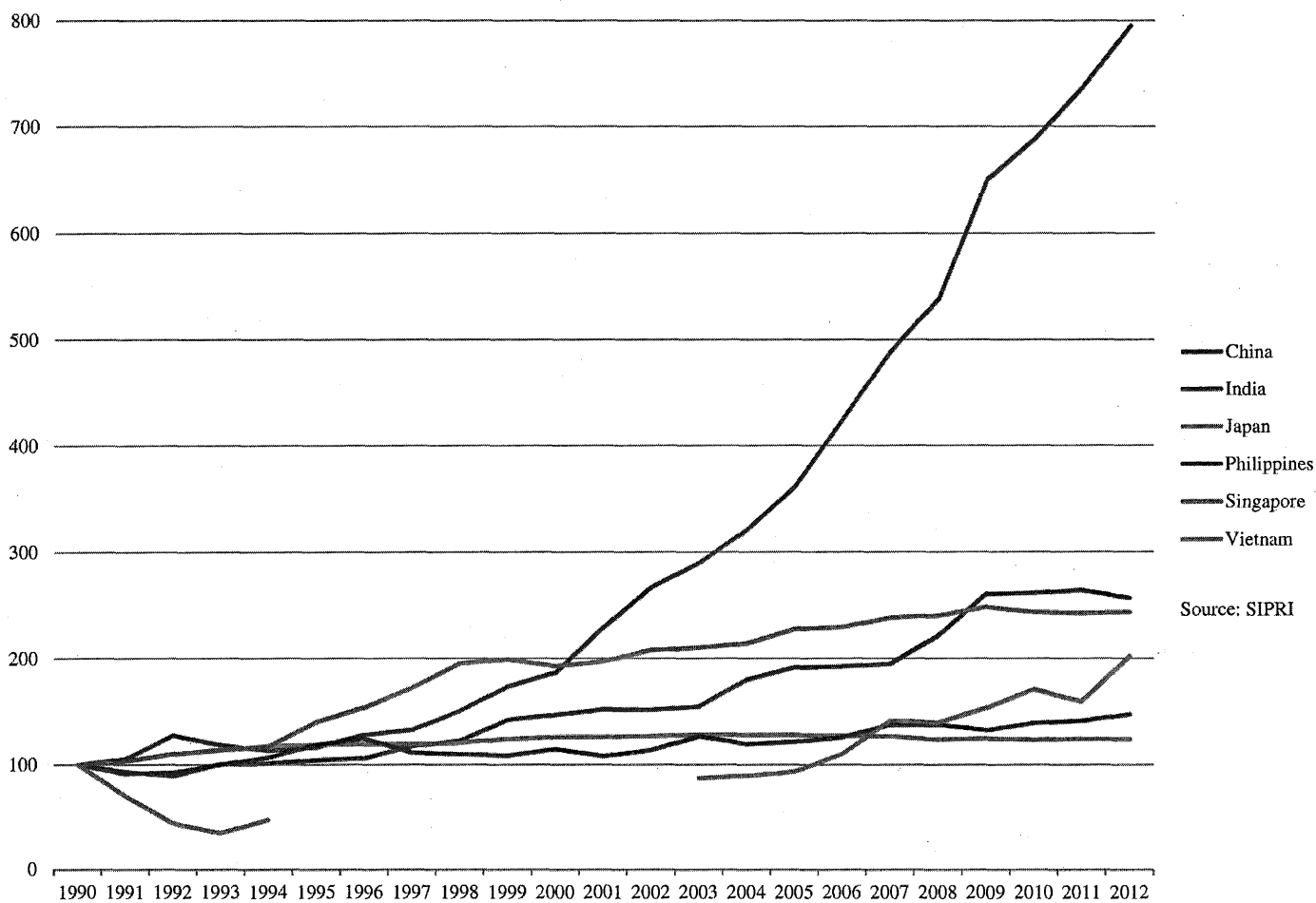
Chinese shipyards produced small batches of progressively more capable warships at first, and then far more rapidly a decade later. Although the navy did purchase destroyers and submarines from Russia, their numbers, in retrospect, were small, especially after China began serial production of its own modern surface combatant and submarine classes and ultimately refurbished and re-commissioned a former Soviet aircraft carrier, the *Liaoning*, in 2012. It may yet build a further two of its own in the coming years.² The Chinese navy even constructed a major new naval base at the southern tip of Hainan Island that includes an underground tunnel for nuclear attack and ballistic missile submarines. Over that same period, China's air force began to transform itself, steadily replacing its older fighters with more modern Russian Su-27SK and Su-30MKK fighters and indigenously-built J-10 and J-11 fighters. To create its new fleet, China heavily invested in not only reverse engineering Russian designs, but also

¹ Division equivalents include only airborne, armored, and infantry formations; three independent brigades or regiments are counted as a division equivalent. Fourteen infantry divisions were merged into the People's Armed Police.

² A second aircraft carrier might be under construction at Jiangnan Group's Changxing Island shipyard near Shanghai. James Hardy, "Images suggest China's first carrier under construction," *Jane's Defence Weekly*, Aug. 2, 2013.

laying the groundwork for a domestic aerospace industry that could develop its own next-generation fighters. Meanwhile, the air force also acquired the kinds of aircraft that it would need on an “informationized” battlefield, such as A-50 and Y-8W airborne early warning and control (AEW&C) aircraft and H-6U aerial refueling tankers. China excelled in missile and rocket technology too. Its conventional forces received new air-to-air missiles and their first undersea-launched anti-ship cruise missiles; its strategic forces rolled out not only newer ballistic missiles, but also those potentially accurate enough to target a ship at sea (if paired with an oceanic surveillance system with sufficient fidelity). And fearing American dominance of space, China put into orbit its own military satellites as well as designed and tested anti-satellite missiles, first in 2007 and then possibly again in early 2013.³

Asia's Military Expenditures Index, 1990–2012



However, as formidable as China’s defense industry has become, it does have its shortcomings. Even as Chinese shipyards launch new warship classes, many of them are powered with Ukrainian gas turbine engines and protected by Russian air search radars. And though China’s Chengdu Aircraft Design Institute and Shenyang Aircraft Corporation surprised many observers with the unveiling of stealthy next-generation fighters, the J-20 and J-31, most of the country’s upgraded fighter fleet is still propelled by Russian and Ukrainian-designed and manufactured turbofan engines. Indeed, with all the advances its defense industry has made, some observers were surprised to learn that China was in negotiations with Russia to buy as many as 48 of its new Su-35 fighters.⁴

³ Andrea Shalal-Esa, “U.S. sees China launch as test of anti-satellite muscle: source,” Reuters, May 15, 2013; Richard D. Fisher, Jr., “China’s Aviation Sector: Building Toward World Class Capabilities,” testimony for the U.S.-China Economic and Security Review Commission Hearing on *China’s Emergent Military Aerospace and Commercial Aviation Capabilities*, May 20, 2010; Duncan Lennox, “More details emerge of Chinese ASAT technology,” *Jane’s Defence Weekly*, Jan. 31, 2007; John Wilson Lewis and Xue Litai, *Imagined Enemies: China Prepares for Uncertain War* (Stanford, CA: Stanford University Press, 2006), pp. 230-241.

⁴ “Su35: Moscow to Shed Tears for the Second Time,” *Kanwa Defense Review*, Jul. 30, 2013; Reuben F. Johnson, “Airshow China:

Certainly China has not been alone in modernizing its armed forces. Many other Asian countries also did so, beginning in the early 1990s. But almost all soon fell afoul of some economic woes. As Japan's economy struggled through the first of its two "lost decades," the Japanese self-defense forces managed to maintain its force structure—benefitting only from the largess Tokyo bestowed on its aerospace and shipbuilding firms, which turned out a small but steady stream of F-2 fighters and new warships. India's military fared far worse. For much of the 1990s, it saw its strength sapped conducting counterinsurgency operations in Kashmir and its budget shrivel from high inflation and a weak currency that ate away at its foreign purchasing power (a situation repeating today). As a result, Indian troops would fight the Kargil War in 2002 with largely outdated equipment. Finally, when the Asian Financial Crisis struck South Korea and much of Southeast Asia in 1997–1998, it severely dented their military modernization ambitions. Thailand, for example, had aspired to become an Asian naval power. Early in the 1990s, Bangkok even funded the new construction of East Asia's first aircraft carrier, the *Chakri Naruebet*, commissioned in 1996. But soon after the crisis, the ship idled at port, sailing for only one day a month as funds to maintain it and its complement of Harrier jets declined (as did the number of operational jets). A similar fate befell Malaysia's modern F/A-18 and MiG-29 fighters. With high maintenance costs, they rarely flew and their combat readiness suffered.⁵

When the economic clouds over Asia finally lifted in the first decade of the new century, many countries in the region were slow to resume their military upgrade programs. The first to do so was India. But its biggest challenge turned out to be its own bureaucracy; even though funds were allocated to modernize its military's equipment and organization, a significant portion was never spent, while other monies were wasted. While many point to India's fifty-year old Arjun tank program as the paragon of such inefficiency, more practically worrying was the 27-year wait the army had to endure to receive any new artillery. Still, India has had success in upgrading its bases along the disputed border with China and in the Bay of Bengal and, after several cost overruns, putting into service a refurbished Soviet aircraft carrier, the *Vikramaditya*, as well as a new Akula-class nuclear attack submarine. Also recovering from its debt crisis by mid-decade, South Korea revitalized its military modernization plans. Since then, it has procured new tanks, armored fighting vehicles, Aegis-equipped destroyers, and six Type 214 submarines.⁶

By and large, most other Asian countries accelerated their military modernization programs only within the last few years. Vietnam turned to its former Russian patrons to acquire new sophisticated air defense systems, Su-30MK2 fighters, and, most impressively, an order for six Kilo-class submarines. It also requested Russian assistance to restore its (and former American) naval base at Da Nang. Indonesia also began large-scale modernization in 2012 with multiple orders of combat vehicles, three South Korean-built Type 209 submarines, a small number of Su-30MK2 fighters, and a much bigger number of transport and training aircraft. But possibly the most dramatic turnabout occurred in the Philippines, which had allowed the parts of its armed forces designed for external defense to decline to near non-existence. That changed in 2011, when Manila procured two retired American high-endurance cutters and began discussions with Japan for ten small patrol boats. Since then, the Philippine government has scoured the world for military hardware, recently negotiating for a dozen South Korean fighters and even considering two Italian guided-missile frigates.⁷

Russian engine makers face quandary over potential Chinese sale," *Jane's Defence Weekly*, Nov. 15, 2012; Reuben F. Johnson, "China continues to test J-20 engine and airframe," *Jane's Defence Weekly*, May 24, 2012; Reuben F. Johnson, "Russia confirms Chinese request for Su-35s, S-400 systems," *Jane's Defence Weekly*, Mar. 1, 2012; "China's J-20: future rival for air dominance?" International Institute for Strategic Studies *Strategic Comment* 17:4, Feb. 2011; Reuben F. Johnson, "Russian industry skeptical about China's 'fifth-gen' fighter," *Jane's Defence Weekly*, Jan. 7, 2011; Jeremy Page, "A Chinese Stealth Challenge," *Wall Street Journal*, Jan. 4, 2011.

⁵ "With Agent 220 Million, No Agent 110 Million," *Kuala Lumpur Security Review*, Jun. 8, 2010; Wassana Nanuam, "Thai Navy Chief Reveals Lack of Cash Hindering Arms Maintenance, Sub Deal," FBIS-EAS-2001-0513, *Bangkok Post*, May, 13, 2001; David Saw, "Funding Crisis Threatens To Submerge Navy," *Defense Daily International*, Sep. 15, 2000.

⁶ Rahul Bedi, "Defending Delhi, running the budgetary obstacle course," *Jane's Defence Weekly*, Mar. 9, 2012; Mrityunjay Mazumdar and James Hardy, "India rolls out reactivated airstrip in Arunachal Pradesh," *Jane's Defence Weekly*, Nov. 29, 2011; Rahul Bedi, "Indian army laments continuing weapon delays," *Jane's Defence Weekly*, Sep. 21, 2010; Rahul Bedi, "Indian Navy to build two new bases and upgrade defences of island territories," *Jane's Defence Weekly*, Apr. 16, 2010; Manu Pubby, "Probe stalls upgrade, Army to get new guns," *Indian Express*, Jun. 24, 2009, <http://www.indianexpress.com/story-print/480429>.

⁷ Alexis Romero, "DND approves terms for acquisition of 12 fighter jets," *Philippine Star*, Jul. 11, 2013, <http://www.philstar.com/headlines/2013/07/11/964112/dnd-approves-terms-acquisition-12-fighter-jets>; "Philippines to buy 2 frigates from Italy," Agence France-Presse, Jul. 3, 2013.

Asia's Navies, 1995–2015

	China				India				Japan			
	Ships	Tonnage	Subs	Tonnage	Ships	Tonnage	Subs	Tonnage	Ships	Tonnage	Subs	Tonnage
1995	52	127,355	56	130,636	40	123,584	19	49,320	62	214,150	16	44,006
2000	60	158,270	64	153,765	40	114,159	18	47,645	54	242,400	16	48,125
2005	67	194,531	61	149,270	47	144,162	16	43,570	54	275,272	16	51,443
2010	78	257,361	56	169,621	46	148,344	16	43,570	52	298,813	16	56,220
2015*	91	389,920	69	226,521	56	255,524	17	46,623	54	370,093	17	63,716

	Philippines				Singapore				Vietnam			
	Ships	Tonnage	Subs	Tonnage	Ships	Tonnage	Subs	Tonnage	Ships	Tonnage	Subs	Tonnage
1995	1	1,776	0	0	6	3,630	0	0	9	12,535	0	0
2000	4	4,101	0	0	6	3,630	1	1,229	10	10,623	2	224
2005	4	4,101	0	0	6	3,630	4	4,916	12	11,673	2	224
2010	4	4,101	0	0	12	23,136	4	4,916	14	10,244	2	224
2015*	6	10,807	0	0	12	23,136	6	8,168	16	14,512	5	9,599

Notes:

* These include the completion of currently ordered serial production by 2015 and continuation of normal retirement patterns.

Sources:

The Military Balance 1995–2013 eds.; *Jane's Fighting Ships* 1995–2013 eds.; and author's estimates.

Among the latest countries to accelerate its military buildup is Japan. While Japan has continued its measured shipbuilding program that averages one new attack submarine and one new surface combatant each year, that pace may increase in the coming years. Already, it is replacing its older combat ships with far more powerful ones. Its two 1970s-era Shirane-class destroyers, carrying three helicopters apiece, will be replaced by two new 22DDH-class "helicopter destroyers"—each of which will nominally embark about a dozen helicopters—even though their size and displacement more closely resemble those of American Wasp-class amphibious assault ships, which are capable of operating V/STOL combat aircraft and up to 40 helicopters. The first 22DDH-class destroyer, the *Izumo*, was launched in August 2013. And given the victory of Japanese Prime Minister Shinzō Abe's coalition in Japan's upper house elections one month earlier, it is possible that he will push through new measures to speed the procurement of warships and coast guard vessels, although the recent depreciation of the Japanese yen may force him to extend the purchase of American combat aircraft.⁸

Apart from China, only Singapore has steadily devoted resources to upgrading all three branches of its armed forces since the 1990s. As a result, the island nation has been able to transform its once provincial defensive forces into a modern military with substantial power projection capabilities, including not only attack submarines, but ones with advanced air-independent propulsion and not only F-15SG and F-16C/D fighters, but ones backed by several networked AEW&C and aerial refueling aircraft. Today, Singapore is already preparing itself to receive delivery of the second of its follow-on Archer-class submarines and will likely be the second Asian country, after Japan, to acquire the American F-35 Joint Strike Fighter.

⁸ Yuka Hayashi, "As Tensions Rise, Pacifist Japan Marches Into a Military Revival," *Wall Street Journal*, Jul. 18, 2013; Paul Kallender-Umezū, "Japan Might Delay F-35 Purchases," *Defense News*, Jun. 10, 2013.

Asia's Air Forces, 1995–2015

	China**				India**				Japan			
	≤3G Fighters	≥4G Fighters	AEW&C	Tankers	≤3G Fighters	≥4G Fighters	AEW&C	Tankers	≤3G Fighters	≥4G Fighters	AEW&C	Tankers
1995	5,500	26	0	0	743	102	0	0	160	165	10	0
2000	3,138	125	0	6	649	117	0	6	110	160	14	0
2005	1,513	334	2	13	608	172	0	6	110	200	14	0
2010	1,098	507	9	13	463	182	1	6	65	245	14	4
2015*	864	913	9	13	455	322	5	6	60	240	17	4

	Philippines				Singapore**				Vietnam			
	≤3G Fighters	≥4G Fighters	AEW&C	Tankers	≤3G Fighters	≥4G Fighters	AEW&C	Tankers	≤3G Fighters	≥4G Fighters	AEW&C	Tankers
1995	11	0	0	0	106	7	4	0	190	0	0	0
2000	11	0	0	0	87	49	4	2	177	12	0	0
2005	11	0	0	0	53	62	4	9	193	16	0	0
2010	0	0	0	0	51	62	4	9	193	16	0	0
2015*	0	12	0	0	43	84	4	9	63	34	0	0

Key:

≤3G Fighters = third (or earlier) generation fighters; ≥4G Fighters = fourth (or later) generation fighters; AEW&C = airborne early warning and control aircraft; Tankers = aerial refueling aircraft

Notes:

* These include the completion of currently ordered serial production by 2015 and continuation of normal retirement patterns.

** China includes naval air force, but excludes four Y-8J AEW and three Y-8T command and control aircraft; India includes naval air force; Singapore includes four dual-use aerial refueling and transport KC-130 aircraft.

Sources:

The Military Balance 1995–2013 eds.; *Jane's All the World's Aircraft* 1995–2013 eds.; *Jane's World Air Forces* 2005–2013 eds.; and author's estimates.

REASONS FOR ASIA'S MILITARY BUILDUP

Naturally, the reasons behind Asia's military buildup are varied and often intertwined. A number have less to do with strategic considerations than domestic ones. One reason concerns domestic political calculations. In countries where militaries have intervened in politics, civilian politicians sometimes use larger defense budgets to buy military quiescence. Past studies of Asian civil-military relations have revealed that this may have been the case in countries like Indonesia and Thailand.¹ Another reason deals with military expenditures that are directed to support favored domestic companies or industries or provide local employment. Of course, that is likely to be true to some degree in any country with a sizable defense industry, like India, Japan, and Singapore. Japan's regular orders for warships and submarines may have reflected its hope to maintain the country's shipbuilding base as much as its desire to improve the country's security, especially after commercial ship orders largely migrated to lower-cost China and South Korea. A third (and somewhat counterintuitive) reason is a growing appreciation among national leaders of how military power can contribute to humanitarian relief efforts. When the 2004 Indian Ocean tsunami devastated the province of Aceh in Indonesia, the Indonesian military could do little to help but watch as American and Australian troops came ashore from offshore ships to deliver aid and search for victims. Humbled by the event,

Indonesia has since set aside more of its military budget for transport ships and aircraft.⁹

Even so, strategic considerations related to changes in the geopolitical environment have played the biggest role in Asia's military buildup in the years after the Cold War and particularly over the last decade. As early as 1991, many Chinese leaders—after witnessing the undeniable success of American arms and organization during the Persian Gulf Conflict—realized that they needed to modernize their armed forces. But institutional interests made progress slow; it was not until after Beijing's failure to deter American intervention in its attempt to intimidate Taiwan with ballistic missile tests in 1995 and 1996 did China's military transformation resume in earnest. China's leadership was finally convinced that its traditionally mainland-bound forces were inadequate to counter American naval and air strength in the western Pacific Ocean and that only a comprehensive military modernization could hope to keep American forces at arms' length as well as prevent other countries from either impinging on its "core interests"—including its territorial claims in the East and South China Seas and on the Himalayan frontier—or geopolitically encircling China.¹⁰

Unfortunately, China's accelerated military modernization created a security dilemma for its Asian neighbors. As China became more militarily powerful, other Asian countries saw their relative security decline. Unsurprisingly, India was quick to act, given its historic suspicion of China—its enemy during the 1962 Sino-Indian border conflict and an ally of its long-time adversary, Pakistan. And over the course of the 2000s, New Delhi also grew concerned over China's rapid development of dual-use civil-military infrastructure—airports, railways, and roads—in Tibet and its growing commercial interests in the Indian Ocean. No stranger to encirclement schemes, Indian security experts saw China developing a "string of pearls" across the Indian Ocean that could one day encircle India. And so, even as China often benchmarks itself against the United States, India came to benchmark its military capabilities against those of China and has found itself wanting. Indeed, Indian fears of China's growing ability to rush massive reinforcements to their disputed border in Arunachal Pradesh has led India to station six mountain divisions (including two newly-raised ones) there to face only three Chinese mountain brigades on the other side. In 2013, India even decided "in principal" to fund a new "strike" corps, to give its frontline units a better offensive capacity.¹¹

In recent years, many other Asian countries have begun to similarly react to China's military buildup, though they had earlier welcomed China's rise, because it had brought them economic benefits. Between the late 1990s and the first half decade of the new century, Beijing's diplomatic "charm offensive" raised the hopes of many Southeast Asians, who were pleased with China's seeming willingness to accept their preference to prioritize economic development over political conflict and consider the region's multilateral norms. But as China's confidence grew with its economic influence and military strength, Beijing began to assert the primacy of its interests in its disputes with Southeast Asia. Rather than embrace multilateral dialogue, China seemed to increasingly sideline Southeast Asian concerns and pursue its aims either alone or through only bilateral negotiations. That has been true of China's recent approaches to conflicts over both its use of the Mekong River and (more famously) its maritime borders in the South China Sea.¹²

In late 2007, Beijing raised the status of the administrative authority governing the Paracel and Spratly Islands to that of a "county-level city" in Hainan province. Then, it listed its South China Sea claims among its "core interests"—those over which it is willing to fight. Sensing the start of a slippery slope, several Southeast Asian countries publicly confronted China about its assertiveness at the 17th ASEAN Regional Forum in 2010. China was incensed by the rebuff. Thereafter Chinese patrol vessels have occasionally harassed oil exploration ships from the Philippines and Vietnam—the militarily weakest disputants in the South China Sea—by cutting the cables towing their ships' seismic equipment. In 2012, China further fired tensions when it built structures on Philippine-claimed Amy Douglas Reef and triggered a months-long maritime standoff. As a result, despite the willingness of many

⁹ Aurel Croissant, David Kuehn, and Philip Lorenz, *Breaking With the Past?: Civil-Military Relations in the Emerging Democracies of East Asia* (Honolulu: East-West Center, 2012), pp. 14-15, 17-18, 34-41; Paul W. Chambers, "Thailand on the Brink: Resurgent Military, Eroded Democracy," *Asian Survey* 50:5 (2010), pp. 835-58; "Seeking a modern role," *Economist*, Mar. 20, 2012; "Old soldiers, old habits," *Economist*, Sep. 21, 2006; Terence Lee, seminar, "The Armed Forces and Regime Maintenance: Explaining the Role of the Military in 1973 and 1992 Thailand, and 1998 Indonesia," Institute of Southeast Asian Studies, Feb. 10, 2006.

¹⁰ Gerald Segal, "East Asia and the 'Constraint' of China," *International Security* 20:4 (1996), pp. 107-135.

¹¹ Rahul Bedi, "Indian Finance Ministry approves 90,000-strong mountain strike corps," *Jane's Defence Weekly*, May 29, 2013.

¹² Felix K. Chang, "The Lower Mekong Initiative and U.S. Foreign Policy in Southeast Asia: Energy, Environment, and Power" *Orbis* 57:2 (2013), pp. 282-299.

Southeast Asian countries to give China the benefit of the doubt that its military buildup was part of a “peaceful rise” or narrowly directed against its wayward province of Taiwan, they now view China with far more circumspection and their own military buildups with greater urgency.

As similar series of events occurred over the Japanese-controlled Senkaku (Diaoyu) Islands in the East China Sea. Starting in 2010, Chinese and Japanese patrol boats have confronted one another in the waters surrounding the islands. Then, after a risky move that Tokyo thought would calm tensions badly misfired, Beijing stepped up its maritime patrols in the area and allowed its citizens to vent their anger against Japanese commercial interests in China. On the other hand, more Japanese have conceded the need to boost their defense preparations, despite their generally pacifist sentiments. As a result, Tokyo has deployed an additional squadron of F-15J fighters to Okinawa and maintained around-the-clock coast guard patrols near the disputed islands. But such sustained demands placed on Japan’s self-defense forces and coast guard have begun to strain their equipment, prompting the need for newer and more capable aircraft and ships.¹³

Some have now suggested that Asia’s military buildup might point to the existence of one or more arms races—situations where conflicting interests or mutual fears cause competitive increases in arms between two states or coalition of states. So far that is not the case, at least not in the literal sense, if for no other reason no Asian state or coalition of states can afford to directly compete with China’s pace and scale of military modernization, barring a downturn in the Chinese economy. Rather than an arms race, much of Asia’s military buildup can be characterized as an arms catch-up. As Asian countries abandoned their purely bandwagon policies toward China, they have scrambled to strengthen their relative military power—partly through their own military modernizations and partly through closer ties with external powers—to hedge against China’s rise. No doubt that is also why Asian countries, like India, Japan, the Philippines, and Vietnam, whose interests seemed so disparate in the past have become so keen on economic and security cooperation today.

IMPLICATIONS OF ASIA’S MILITARY BUILDUP

The primary concern of many of those who follow Asia’s military buildup is that the increased level of armaments will likely lead to conflict, whether through miscalculation or design. In addition, they could argue that the possibility for miscalculation is made all the more likely because of the overlapping military catch-up efforts—China’s attempt to catch-up to the United States, India’s to China, Japan’s to China, and the rest of Asia’s to China—which creates a complex situation where the actual, functioning balance of power is difficult to ascertain.

Even so, sometimes overlooked are countervailing factors that could lessen the possibility of conflict. First, though Asian countries are rearming in response to China’s increased defense spending and more assertive behavior, they would prefer not to see China as an adversary and hope that it could continue to be a source of economic vitality for the region. Indeed, all Asian countries, even China, have underlined the benign nature of their intentions. Second, all the countries of Asia share common interests that bind them as states, such as promoting economic growth, deterring terrorism, and foiling transnational crime. And third, as history has demonstrated, military expansion can also result in agreements to limit arms, especially as they become more costly to accrue. It was just such a concern that led the world’s five leading naval powers to agree to a naval arms limitation treaty at the Washington Conference in 1921–1922.¹⁴

But if economic benefits and common interests are insufficient to allay qualms over the intentions of possible rivals—as often is the case—and Asia’s military buildup continues, then those countries playing catch up with China would be well advised to do so through the acquisition of new military technologies. Rather than try to match Chinese forces in terms of absolute numbers of aircraft and ships, they could attempt an asymmetric approach with new technologies against which China has fewer defenses. Much like China’s attempt to thwart American carrier battle groups with conventionally-armed ballistic missiles, Asian countries could emphasize some combination of new technology and tactics that can compensate for smaller quantities. Such systems could include supersonic land-based cruise missiles (and radar systems that support them), stealthy attack submarines, armed

¹³ Kosuke Takahashi and James Hardy, “UAVs, marines, BMD top Japanese MoD’s policy recommendations,” *Jane’s Defence Weekly*, Jul. 30, 2013.

¹⁴ Paul Kennedy, *The Rise and Fall of British Naval Mastery* (London: Macmillan, 1983), p. 274; Roger Dingman, *Power in the Pacific: The Origins of Naval Arms Limitation, 1914-1922* (Chicago: University of Chicago Press, 1976).

unmanned aerial vehicles, or even long-range standoff weapons launched from unconventional platforms, like the P-8A maritime patrol aircraft. Doing so would enable Asian countries to more quickly approach parity with China, regardless of the current gap in conventional military power, and bring greater security and stability to the region.

For the United States, Asia's military buildup can be seen as a source of both concern and comfort. Naturally, China's continued military expansion and benchmarking against American capabilities are troubling, though not yet alarming, unless the United States curtails its own military modernization efforts. But China's growing power has created new tensions in the dynamics of America's bilateral security arrangements across the region. On the one hand, if American security guarantees are too firm, then its security partners could embroil the United States in an unwanted conflict. On the other hand, if American security guarantees are too weak, then its security partners could decide that their interests might be better served by currying favor with China. So far, that is not the case. China's new arms and recent maritime assertiveness have led many Asian countries to seriously invest in their own defenses for the first time since the Cold War. And to the extent that these countries are friendly to the United States, Washington can take some cheer from the fact that for now others are willing to share more of the balancing burden in Asia.

In any case, it is far from certain that Asia's military buildup will inexorably lead to crisis or war. What matters in the end is not the region's quantity of armaments, but rather the region's perceptions of power and intentions. At the moment, those of China concern many Asian countries. Yet if they, along with the United States, collectively gather enough power to persuade China to temper its provocative behavior, then their military buildup will have contributed to the region's security and stability. Conversely, if China's military power continues to grow relative to that of its neighbors, then one can expect more confrontations to come—no matter the quantities of arms amassed.

FPRI, 1528 Walnut Street, Suite 610, Philadelphia, PA 19102-3684

For more information, contact Eli Gilman at 215-732-3774, ext. 255, email fpri@fpri.org, or visit us at www.fpri.org.

What We've Learned from China's Air Defense Zone (so Far)

December 9, 2013 · in Commentary and Analysis

warontherocks.com

Last week, U.S. Vice President Joe Biden visited Japan, China, and South Korea. The agenda of the previously planned trip—trade promotion—was largely scrapped. Instead, Biden spent most of his time in the region doing damage control, after China's unilateral declaration of controlled air space over the East China Sea. Instead of discussing trade and financial liberalization, the vice president fielded questions about America's resolve from anxious officials in Tokyo and Seoul. While Biden was in Beijing, China's officials undoubtedly made a similar appraisal of U.S. fortitude, albeit from the opposite perspective.

China's declaration on November 23rd of an air defense identification zone (ADIZ) covering much of the East China Sea, including the disputed Senkaku/Diaoyu Islands, has revealed much about China's leaders and their priorities. What we have learned from this episode should cause analysts to recalibrate their assumptions about China's intentions and behavior.

1. Regarding East Asia, China's leaders don't have much use for diplomacy. China's declaration of an ADIZ was sudden and unilateral, done without consulting its neighbors. This contrasts with Japan's ADIZ declaration, which, although protested by Taiwan, was at least done with prior consultation with Taiwanese authorities. South Korea's adjustment of its ADIZ, done in response to China's declaration, was also done after consultation with China and Japan, a difference noted by the U.S. State Department.

One might presume that the objective of Beijing's declaration is to improve China's awareness of activity in the air space to its east. The best way to accomplish that goal, especially with ostensibly non-hostile neighbors, would be to negotiate information-sharing agreements, air space management procedures, "hot line" arrangements, and other such cooperative measures. If China's main concern was air space awareness, multilateral cooperation would be the most effective means of achieving that awareness.

China's rejection of this approach in favor of a unilateral air control declaration reveals several disturbing conclusions about China's decision-makers. First, rather than an attempt to improve air space awareness, the declaration is foremost another act of "salami-slicing" in the region, a gradual program to establish "facts on the ground" that Beijing hopes will eventually add up to recognition of Chinese sovereignty over disputed territory in the region (Matthew Hipple recently explained at *War on the Rocks* why China's declaration more closely resembles a sovereignty grab instead of a true ADIZ). China's hoped-for gains come at the expense of the sovereignty and security of China's neighbors. If true, China would increasingly resemble a revisionist adversary, albeit a careful, slow-moving one.

China's dismissal of diplomacy indicates that its leaders have high confidence in China's strengthening position in the region; why bother with negotiations when the trends are so favorable? China's stiff-arm of ASEAN over negotiations for a code of conduct for dispute resolution in the South China Sea is another example of Beijing's opinion of diplomacy with its neighbors.

2. China's leaders care about more than economic growth. Many analysts point to deep economic interdependence as a reason to dismiss the possibility of a major power conflict in East Asia. China's actions make that a questionable conclusion. If maximizing economic growth and reducing China's dependence on vulnerable crude oil imports from the Middle East were the paramount objectives of China's leaders, why are they stepping up China's territorial assertions and in doing so, inciting security

competitions in the East and South China Seas? According to the U.S. Energy Information Administration, there may be enough crude oil under the South China Sea for sixty years of Chinese consumption. Oil under the East China Sea could potentially supply China's needs for another fifteen years. If China's dominant concerns were its economic growth rate and energy security, it would seem to have a strong interest in settling its maritime disputes with its neighbors (as it has done with its land borders to its west and north). It could then cooperate with these neighbors to exploit the enormous hydrocarbon resources under the two seas, reasonably assuring its future economic and security interests.

But, instead of pursuing that seemingly rational "profit maximizing" strategy, China is engaging in persistent territorial "salami-slicing," accepting the attendant risks of brinkmanship and possible confrontation. China's actions reveal that other goals, such as achieving control over claimed territory, are at least as important to China's leaders as economic growth.

3. ***China's leaders don't seem to care about an anti-China backlash in the region.*** Before 2008, Chinese "smile diplomacy" appeared to predominate across the region. Since then, "snarl diplomacy" has been more in use. The result has been a backlash against China and an accelerating security competition in the region. An arms race in the region appears underway, with submarines as a major focus of the competition. Non-Chinese defense spending in the region is expected to jump 55% in 2013-2018, compared to the prior five-year period.

These trends have been visible to China's leaders for several years. Indeed, it seems clear that China's assertiveness was a major factor prompting the return of nationalist Prime Minister Shinzo Abe and his party in Japan, confirmed in two strong election results in 2012 and 2013. Abe's defense policy has focused on the Chinese threat, with budget increases forthcoming and military forces now repositioning to Japan's southwest. The U.S. is renewing its military links with the Philippines and slowly expanding them in Vietnam. And India is expanding its security ties with ASEAN and in East Asia. A growing list of security linkages among China's neighbors offer yet more indicators that countries in the region are hedging against China's assertions.

This clear security backlash did not deter China from declaring its East China Sea ADIZ. Indeed, the declaration very likely damaged China's relations with South Korea, one of the few examples of recent Chinese diplomatic success. In response to the declaration, South Korea, Japan, and the United States have increased their military air patrols and intelligence-gathering in the East China Sea. China's actions are worsening China's security.

Based on the backlash that had been building for years before the ADIZ announcement, further reinforcement of U.S. and allied military activity in East Asia should hardly have been a surprise to China's leaders. Yet, they proceeded with the declaration anyway. Why? These leaders may be possessed by a quasi-Marxist belief in historical inevitability, a confidence in the "correlation of forces" trending in China's favor. Or, as the next section will explore, there may be other forces compelling China's leaders to make decisions that seem so unfavorable for China's security.

4. ***Domestic nationalism may be driving China's external policy.*** None of the explanations for this Chinese "own goal" are reassuring. If the ADIZ declaration was a product of mid-level bureaucratic machinery at work, it would mean that China's top leaders don't have control over the policy decisions of their government, in particular those taken by the People's Liberation Army (PLA). If top-level leaders made this decision without an appreciation for the backlash it would catalyze in the U.S. government and

across the region, it would indicate that these leaders don't understand their counterparts. In that case, U.S. and allied leaders should not assume that China's leaders will act competently during a crisis.

Perhaps the most plausible explanation is that China's leaders made the ADIZ decision in order to placate nationalist factions in the PLA, the government, or society at large. This explanation is no more comforting than those listed above. It would imply that Chinese "hawks" can at least occasionally drive China's external policy, with all the risks that accompany such moves. That would seem to be a recipe for action-reaction cycles in a region where nationalism is now on the rise. It may also mean that China's leaders fear that the other bases of their political legitimacy—such as the Communist Party's moral position in Chinese society or the government's ability to deliver material benefits—have faded away. If so, these leaders may have concluded that the manipulation of nationalist sentiment may be one of their only remaining levers of social control.

East Asia has avoided a major power conflict for seven decades because the presence of U.S. military power has maintained the region's stability. But, tensions in the region are now rising. The proximate issue in dispute is the sovereignty of a few uninhabited islets in the East and South China Seas. The fundamental issue is the rapid expansion of China's power and how the United States and China's neighbors will cope with this sudden upending of East Asia's security balance. China's rise is coinciding with a perception of U.S. decline and an appearance that Washington has become disillusioned with the global security responsibilities that it has long taken on. Such perceptions have likely influenced decision-making in Beijing, with increased risk in the region now the result.

China's sudden declaration of an ADIZ in the East China Sea has revealed much about the leadership in Beijing. The ADIZ declaration was not a reasonable policy to enhance China's security; if China truly wanted better awareness over the air space to its east, it could have chosen to cooperate with its neighbors to get that information. Instead, the ADIZ declaration is another act of territorial "salami-slicing," which will only accelerate the security competition between China and its neighbors. We can further conclude that economic development and energy security are not the only goals of China's leaders; if they were, China's leaders could have taken other paths that would not have placed these goals at risk. U.S. and allied policymakers should thus not assume that economic interdependence will always deter conflict in the region. It should also be no comfort to post-modern observers in the West that nationalism could be a significant driver of China's external policy, and possibly a prop supporting the internal legitimacy of China's government.

Military tensions among the three largest economies in the world are now rising. The underlying security trends in East Asia suggest that the crisis over China's ADIZ will not be a one-time event, especially if China continues to act like a revisionist power. Regrettably, the era of interstate conflict may not be over just yet.

Robert Haddick is an independent contractor at U.S. Special Operations Command. He writes here in a personal capacity. In 2014, Naval Institute Press will publish Haddick's book on the rise of China's military power and U.S. strategy in East Asia.

Session 4 – The West, Women and War

INTEGRATING WOMEN INTO THE INFANTRY

Captain Adam N. Wojack, U.S. Army

WOMEN HAVE NOT been given a chance to succeed on an equal footing with men in the military. Even after decades of reform, initiated by the All-Volunteer Force (AVF) in 1973, women are excluded from six principal job groups: infantry, armor, short-range air defense, cannon artillery, combat engineers, and special forces. Opponents of total gender integration point out that women cannot perform the physically demanding work these job groups require. Opponents also say that putting women into units where only men have traditionally served will jeopardize the Army's combat readiness by ruining unit cohesion.

This article proposes that the U.S. Army integrate women into the infantry branch. It will dispel practical notions that a woman is too "weak" to do an infantryman's job and that her presence will destroy team spirit and ground maneuver units' fighting effectiveness. This article does not dispute those who believe it is wrong for the United States to send women to fight close combat battles, nor is it an advocate for those who wish to destroy gender barriers simply because they exist. It acknowledges the personal nature of those points of view and avoids them altogether. Instead, this article assumes a sociopolitical climate in which only practical debate is waged about whether to integrate women into the infantry. The issue, then, is not about right and wrong but about suitability and feasibility. Can women do the infantryman's job, and how can the Army help them do it? The key assumption, here, is that American women would volunteer to become infantry soldiers if given the chance.

Why Women in the Infantry?

Ground combat units contain the only jobs closed to women in land-based military forces today. Before the AVF, which recruited women to replace some of the Army's postdraft manpower losses, women made up 3 percent of all soldiers in the Army. Today, women account for 14 percent of all

At the end of Operation Desert Storm in 1991, there were 13 women killed in action. Of those, four were termed "hostile deaths" out of a total of 148 U.S. combat deaths.

soldiers and 20 percent of all recruits.¹ They fly attack helicopters, command military police companies, drive infantry soldiers into combat on trucks, and "man" logistics bases far forward, or in the midst of ground maneuver forces. In the past 15 years, women have been killed in combat. At the end of Operation Desert Storm in 1991, there were 13 women killed in action. Of those, four were termed "hostile deaths" out of a total of 148 U.S. combat deaths.² Interestingly, two of those deaths occurred when an Iraqi Scud missile hit a temporary barracks housing combat service support units far behind the forward edge of the battle area.³

These deaths seem to back the notion that today's battlefield is no longer as well-defined as it once was. For example, U.S. offensive doctrine calls for attacking the enemy's lines of communication, in addition to his main defenses, to disrupt their combat forces' resupply.⁴ The theory is that, if successful, the enemy's maneuver forces will run out of rations, ammunition, and the will to fight, in that order. It is no secret that the United States' conventional threat uses the same doctrine. Our field trains, brigade support areas, and division support areas are the key objectives of conventional enemy attacks. It is also no secret that most Army women work in these areas.

This doctrine transforms all soldiers—men and women—in field command and control and/or logistics areas into front-line combatants, at least in the enemy's eyes. Why attack through infantry and

The views expressed in this article are those of the author and do not necessarily reflect the position of the Department of the Army, the Department of Defense, or any other government office or agency.—Editor

armor when the division rear can be penetrated? Of course, this says nothing about why women belong in the infantry.

Proponents of giving women the right to serve in ground combat units usually use a combination of arguments: an equal opportunity to serve is every

Sports scientists generally agree that a woman can run 90 percent as fast as a man over all distances and is about two-thirds as strong in the upper body. However, upper body strength assessments may soon prove to be inaccurate. Women's weightlifting world records hover at around 70 percent of the men's record in most comparable weight classes. In some classes, the women's record is as high as 78 percent.

American's right; current technologies are gender-neutral; and other nations allow women in the infantry.⁵ This list omits perhaps the most compelling reason to integrate women into the infantry and other ground combat fields: given the contemporary operating environment, women are in close proximity to combat regardless of where they are on the battlefield, so they might as well be allowed to fight offensively.

At this point, opponents of gender integration point out why women do not belong in the infantry:

- Women lack the upper body strength to perform an infantryman's job.
- Women have certain hygiene needs that would demand special treatment in the field.
- Women are too valuable as reproducers of human life to be wasted in ground combat.
- Women are nondeployable while pregnant.
- Women would destroy the cohesion of previously all-male ground combat units.⁶

The oldest argument against allowing women in the infantry is that women are too weak. The genesis of this argument is as old as society and civilization—women are the weaker sex; a woman's duty is to bear and raise children; men are supposed to protect women. Whether these assumptions are myth or reality, they have governed social thought for centuries. Women did not serve in the military, and women did not play sports. Women who chose to work were restricted to teaching, clerking, and nursing. Even when women were allowed in the military during World War II, they did not receive the same military training as the men. Instead, women received pointers on how to maintain trim figures and an attractive appearance.

Women were partly responsible for this. In World War II, the Navy's Women Accepted for Voluntary Emergency Service (WAVES) initially attracted more women than the Army's Women's Army Corps (WAC). The WAVES' navy blue uniform was considered more stylish than the olive drab the WACs wore.⁷

The women's movement of the late 1960s and early 1970s changed all of that. American women demanded equal treatment and equal opportunity in all aspects of society and very nearly got that. In 1972, Congress enacted a law known as Title IX that made it illegal for any school that received Federal funds to spend more money on men's athletics than it did on women's athletics. More than anything before or since, Title IX made it acceptable and attractive for women to pursue athletic dreams and hone their athletic prowess at the high school and collegiate levels.⁸ In the 25-plus years since Title IX was introduced, women's athletics in America have grown exponentially. Today, women in high school and college compete in many of the same sports in which men compete, and women now play professional soccer and basketball in national, televised leagues.

Title IX's key contribution was giving women the government's stamp of approval to be athletic without risking losing their "womanness." Meanwhile, women gained much athletic ground on men. Consider the world record progression in the marathon. In the past 30 years, the men's record has gone from 2 hours, 9 minutes (2:09) to 2:05, a 3-percent improvement. Over the same period, the women's record has improved from 3:01 to 2:18, an almost 24-percent improvement. The women's record went from being 71 percent of the men's record to 90 percent.⁹

Note also the pole vault event, which was closed to women because track and field authorities considered women either to be too weak or the event too dangerous, or both. Since the International Association of Athletics Federations opened the event to women in 1992, the women's record has gone from an initial 4.05 meters (m) to 4.81 m, a 16-percent improvement. In the same period, the men's record improved only .04 percent, from 6.12 m to 6.14 m.¹⁰

This is not to say that women will continue to improve at the pace of the past 30 years and bypass men's athletic accomplishments. What this shows is a true picture of women's athletic potential. Simply put, before Title IX, women were not encouraged to play and, on the whole, played at a misleadingly low level compared to men. Since Title IX, women have been able to realize their physiological potential in athletics.



Army nurses prepare to disembark from a troop transport at Liverpool's Princess Dock in 1944.

In World War II, the Navy's Women Accepted for Voluntary Emergency Service (WAVES) initially attracted more women than the Army's Women's Army Corps (WAC). The WAVES' navy blue uniform was considered more stylish than the olive drab the WACs wore.

Sports scientists generally agree that a woman can run 90 percent as fast as a man over all distances and is about two-thirds as strong in the upper body. However, upper body strength assessments might soon prove to be inaccurate. Women's weightlifting world records hover at around 70 percent of the men's record in most comparable weight classes. In some classes, the women's record is as high as 78 percent.¹¹ Bear in mind that weightlifting, like pole vaulting, was introduced to the Olympic Games as a medal sport for women only during 2000.

Scientists point out that a woman's athletic potential is limited by specific physiology. American women are, on average, 4 inches shorter than the average man and 40 pounds lighter. Women also have from 6.5 to 13 pounds more body fat than men and from 40 to 48 pounds less lean body mass, or muscle weight. Women also possess about one-tenth of the amount of testosterone as men, a hormone that is key to strength development. Because of this, scientists say that even the strongest, fastest woman can never expect to surpass the strongest, fastest man.¹²

However, even if physiology allows a woman to be two-thirds as strong as an average man, most women are actually much less strong than that. Sports physiologists believe this condition is culturally induced. In our society, strength is viewed as a masculine trait, and small, frail bodies are considered to be feminine. Sex stereotypes such as these do much to program behavior and prevent individuals from fulfilling their full potential.¹³ In the past, this has meant discouraging women from engaging in

weight training and the more strenuous sports (football, basketball, soccer) that men have traditionally played. While there has been progress since Title IX, change is slow.

Change is even slower in the Army. The AVF and Title IX occurred at around the same time. Both acknowledged a need for greater and more varied roles for women in society and in the Armed Forces. While Title IX spawned a generation of professional women athletes, the AVF seemed content to protect the status quo. Consider the Army Physical Fitness Test (APFT). The current APFT, introduced in 1999, is only a revision of the 1984 paradigm that introduced 2 minutes each of pushups and situps and the timed 2-mile run.

The Army developed the 1984 standards by testing a large group of soldiers without familiarizing or training them on the new events and then recording the scores. The Army APFT minimum standard became the number of pushups and situps and 2-mile run time recorded by soldiers at the bottom of the 90 percent that passed. In sum, the Army's landmark 1984 APFT makeover, intended to bring soldiers to a higher level of health and physical readiness, was based on the achievements of the 11th percentile.¹⁴

Men's standards were actually high enough to be both challenging and realistic. The youngest men, 18 to 21 years old, had to perform a minimum of 42 pushups and 52 situps, and run 2 miles in 15:54 to pass the APFT. However, women's standards were much lower. The youngest women had to do 18

pushups and 50 situps and run 2 miles in 18:54.¹⁵

Comparatively, women did 43 percent of the upper body work the men did, performed roughly the same amount of abdominal work, and ran 84 percent of the men's minimum. These scores do not

The obvious comparison to total gender integration is the Armed Forces' racial integration that President Truman ordered in 1948. . . . After integration, blacks and whites agreed that black soldiers performed better in racially mixed units because competition with white soldiers improved not only their soldier skills but their self-confidence as well.

correspond with a woman's physiological potential. Assuming that most soldiers train to meet rather than exceed APFT standards, Army women since 1984 have done only about 64 percent of the upper body work of which they are capable and about 90 percent of their ability on the 2-mile run.

By the mid-1990s, the Army, realizing its standards were too low in some places and too high in others, rewrote the APFT's minimum standards. As of 1999, women and men were required to do the same number of situps across all age groups. The Army also increased the minimum number of pushups for the youngest women—by one. The 2-mile run time minimums remained the same. In determining the new standards, the Army used the 1984 testing strategy, eliminating the bottom 10 percent as failures and adopting the next lowest score as the minimum standard. Participants were not put through a special physical training program to raise standards. Soldiers had performed against the 1984 standards throughout their periods of service and performed predictably.¹⁶

Given this study, one could say the Army has no one to blame but itself for any shortcomings in the perceived or tested physical abilities of female soldiers. In short, the Army has not given women a chance to succeed physically on a par with men.

Physique, which is only one argument against introducing women into the infantry, is also the one the critics are most ready to concede in light of the great strides women have made in athletics in the past 30 years. Other issues—nondeployability, special hygiene and privacy needs, and their status as procreators—also deserve attention.

Opponents of women in the infantry cite nondeployability because of pregnancy as a reason not to have female grunts, but consider the Gulf war. Overall deployability rates throughout the Gulf war were 91 percent for women and 98 percent for

men. Reportedly, half of all soldiers who did not deploy had medical problems. Because of the 7-percent disparity, the consensus among men—without proof—was that many women were getting pregnant so they would not have to deploy. It is likely that some women did just that. Even so, the pregnancy rate for women in the military during the Gulf war remained the same as the peacetime rate. It has even been suggested that, as a temporary disability, men missed more work time due to sports injuries than women missed while pregnant.¹⁷

Privacy and hygiene needs are the next issue. In short, men and women require separate latrines, showers, and living quarters, especially in the field. This was also tested during the Gulf war. The Army discovered that by using common sense and having respect for each others' needs, men and women soldiers could share limited field latrines and showers without incident. Billeting is a concern only when it is limited to tents or actual field conditions. Commanders in the Gulf tried several different strategies: separate tents when available; women's sections of tents separated by hanging towels or blankets and integrated tents where privacy was minimal.

In the end, women found they preferred sharing tents with those they worked with and handling any privacy issues in the same common-sense manner as was used with latrines and showers. Rumors of sexual liaisons in tent cities were common during the Gulf war, but two reasons probably kept such behavior to a minimum: the lack of privacy and familiarity within any unit. In the words of one woman assigned to a unit deployed to the Gulf, "We know their wives and girlfriends so we don't expect trouble."¹⁸

The third issue is women's place as procreators and nurturers. Opponents say women are too valuable to society to risk in direct combat because they bear children. But this issue pulls at emotions rather than at intellect. In the early 1990s, Air Force Chief of Staff General Merrill A. McPeak spoke for many when he said, "I just can't get over this feeling of old men ordering young women into combat. . . . I have a gut-based hangup there. And it doesn't make a lot of sense in every way. I apologize for it."¹⁹ In its original manifestation, the Equal Rights Amendment narrowly missed being approved, perhaps because of this one issue.²⁰ But the issue is very much alive.

The actual percentage of women in direct combat would probably be quite low. No one can imagine the total number of women in ground combat units surpassing 25 percent of all personnel anytime in the distant future after any type of integration. With 10 divisions in the Active Army totaling 500,000, 25 percent equates to 20,000 women in di-



Male and female members of the 1st Armored Division headquarters staff wait patiently for their turn to shower, 22 February 1991.

Commanders in the Gulf tried several different strategies: separate tents when available; women's sections of tents separated by hanging towels or blankets and integrated tents where privacy was minimal. In the end, women found they preferred sharing tents with those they worked with and handling any privacy issues in the same common-sense manner as was used with latrines and showers.

rect combat roles. Twenty thousand women represents about .03 percent of the approximately 60 million American women who are currently in their reproductive prime.²¹ Of course, a world war would more than likely pull more women into ground combat units, but in that case, global and human survival would be at stake. In comparison to peacetime, the number of women who would voluntarily serve in ground combat units is low. For example, Canada, a nation with an armed force of about 65,000, currently has six women infantry soldiers.

On the flip side of this issue, women are the most valuable human military asset because some of the gravest threats today come not from conventional armies but from asymmetric forces such as global terrorists. This type of threat uses surprise to achieve its goals. The current threat also operates from Third-World, religiously fundamental countries or societies where women's rights do not exist. Such male-dominated, paternalistic, and sexist threat groups probably do not expect to meet resistance from women, especially those who are not dressed in traditional military garb. Well-trained women could become America's greatest source of asymmetric combat power.

The last issue that opponents of gender integration bring up is cohesion. They wonder if allowing women into ground combat roles improves or undermines combat readiness. The hard truth is that right now, integrating women into the infantry and other currently all-men combat arms units would more than likely hurt morale initially. Infantrymen and

leaders would fumble their way to finding out just how to deal with women. The news media would pay close attention and generate excessive publicity, both positive and negative. It would be a bumpy ride for a while, but more than likely, the Army and the infantry would adjust. The U.S. Army would be better for it, not worse.

The obvious comparison to total gender integration is the Armed Forces' racial integration that President Harry S. Truman ordered in 1948. Based on unit cohesion alone this was a risky move that most whites and many blacks opposed. Whites argued that black soldiers were unreliable and careless, and blacks maintained they would not get fair treatment in racially integrated units. After integration, blacks and whites agreed that black soldiers performed better in racially mixed units because competition with white soldiers improved not only their soldier skills but their self-confidence as well.

Researchers at the time also found that desegregation did not hurt combat effectiveness. Residual racism still existed, but it was offset by the realization that blacks could be as competent at soldiering as whites and that formal integration was improving black soldiers' skills. Interestingly, black soldiers' complaints of racism or unfair treatment actually decreased in integrated units.²²

Following this example, gender integration could be as simple and successful as racial integration. But women are not separated from men by skin color alone. In gender integration, women actually have less in common with men than white and black

According to a 1997 RAND study that the Department of Defense sponsored, "Gender differences alone did not appear to erode cohesion. Cohesion was reported high in units where people believed the command emphasized unity and the importance and necessity of all members and divisions in accomplishing the mission."

men did with each other in 1948. White and black men were already infantry soldiers. Women, if they are to be accepted by men in previously all-male fields, must not only prove themselves equal to men, they must also demolish generations-old perceptions of being the protected rather than the protector. This is the recipe for cohesion.

Social scientists today prefer to divide cohesion, or the feelings that bind individuals to the immediate group, into two types: task and social.²³ Opponents argue that introducing women into ground combat units would immediately erode those units' social cohesion, which they argue is more important than task cohesion. Opponents also say that any well-trained group of men and women can develop task cohesion to accomplish virtually any work problem, regardless of how members feel about each other. They go on to say that while task cohesion may be enough to get the job done in the civilian work force, it is not enough in the military. In the military, the intimacy and isolation of combat demand high social cohesion.²⁴ This leaves only one question: Can women bond socially with men?

According to a 1997 RAND study that the Department of Defense sponsored, "Gender differences alone did not appear to erode cohesion. Cohesion was reported high in units where people believed the command emphasized unity and the importance and necessity of all members and divisions in accomplishing the mission."²⁵ Even more important, the study went on to say, "High social cohesion, or bonding on a social level, can have deleterious effects on performance outcomes and task cohesion, because people start to prioritize friendship and social activities over performing their jobs."²⁶

The bottom line regarding either social or task cohesion is fair and equal treatment. Women and men can bond to form effective units in any job field or situation as long as the women feel they will be treated equally and the men perceive that the women will not receive special treatment.

There is one additional lesson about gender integration and unit cohesion that our service academies and military colleges have taught us: Once women comprise more than 20 percent of a unit or class, they are judged as individuals and not as representatives of their gender. Successful women cadets in

the group become fellow cadets, not female cadets. Overall unit acceptance soon follows.²⁷

On the matter of cohesion, caution goes with promise. The lessons that service academies and NATO allies have learned, particularly in Canada, as they went through gender-integration trials, tells us that it takes roughly a year to break down pre-existing, negative, sexist attitudes. It takes quite a while before mixed-gender units function more efficiently and at higher levels of capability than all-male units. Therefore, says one social scientist, "Until American women are given the opportunity to dispel the prejudicial opinions ensconced within the U.S. military, those opposed to extended integration assist in the perpetuation of these preconceived notions."²⁸

Integrating Women Into the Infantry

Contrary to what many believe, only Canada has succeeded at desegregating its infantry. When the topic of military fighting women comes up, many point to Israel as an example of a nation with a gender-integrated ground combat force. But this is not true. Israeli women have not served in combat roles since Israel's War of Independence in 1948, and even then, most say they served because of desperate need. Today, unmarried Israeli women are drafted and serve short tours in the Israel Defense Forces, but they are restricted to clerical and non-combat medical fields. They are excluded from any duty involving imminent danger. In reality, American women have far greater military opportunities than do Israeli women.²⁹

Canada, the only modern nation with women in its infantry ranks, began its gender-assimilation program in 1989 and met mostly with failure. It eventually succeeded on a small scale, learning hard lessons along the way. These lessons are the key to successful gender integration in other armies.

First, Canada's volunteer women went through regular segregated basic training, performing no more than minimum women's standards before integrated infantry training. Women's physical fitness standards in Canada are lower than men's, so the women arrived at infantry school already behind the men in overall fitness.

Once in the integrated infantry school, the women were piecemealed among the training platoons. The average composition of a platoon contained only two or three women to 30 to 40 men. This led the women to feel little or no peer support.

Finally, the Canadian forces selected too many women who could not meet the physical standards needed to perform an infantryman's job. This was probably because of a lack of volunteers. Some Canadian women dropped out of the program early, feeling that videos of infantry training and recruiters' descriptions misled them.³⁰

Also, sexual harassment in Canada's infantry was a problem that was not addressed. The first woman to become an officer in Canada's infantry was a young woman named Sandra Perron. For years, coworkers subjected Perron to what she termed "constant emotional and psychological harassment."³¹ Perron did not complain; instead, she quit the army in 1995. Several years later she spoke out about the abuse. She recounted one incident of being tied to a tree, beaten, and left in the snow without boots for 4 hours. An investigation revealed that peers who were competing with her for promotions resented Perron.³²

Canada reacted to the Perron incident by instituting policies and training designed to eliminate sexual harassment. Current women infantry soldiers in Canada credit Perron for breaking down barriers and forcing the army to rethink its position on sexual harassment. Still others feel Perron handled the situation incorrectly. Another woman infantry officer, Maureen Wellwood, told a reporter, "The key is to talk about it. Sandra Perron should not have kept quiet. . . . She was very strong, but she accepted it at the beginning, and it kept going."³³ Wellwood said that there are still many men who oppose women in the infantry. "And there still will be years from now. But now the people who harass get into trouble, and not the other way around."³⁴

Of interest is the still minuscule total number of women serving in Canada's infantry: six. In an armed force of 65,000, that number achieves the critical mass of 20 percent in only one echelon: a single platoon. It appears that Canada's infantry women will be isolated for some time. Even so, the Canadian Government has taken notice and is learning. In a 1999 article on recruiting women into its armed forces, the Canadian Department of National Defence was quoted as "hoping 25 percent of all new enlistees will be women."³⁵

The U.S. Army can benefit from Canada's experience. In fact, integrating women into its infantry need not be difficult or painful as long as it is approached with common sense and a common purpose. That common purpose should be success. It should not integrate women into traditionally all-male units unless it is serious about creating an environment for their success. It can accomplish this by synthesizing the arguments and lessons learned.

The Plan

Adopt higher APFT standards for women. Use the current 1999 standards for men as a starting point, and set the women's minimum standard for the run at 90 percent of the men's—17:40 rather than 18:54. Make the pushup standard for women 70 percent of the men's standard—29 rather than 19. Situp standards should remain equal. These standards are overall improvements that the Army needs to increase women's physical condition. This will

give the Army a more physically fit force and dispel the notion that women soldiers have a lower, easier standard than men.

Infantry one-station unit training (OSUT) for enlisted women should be preceded by a women-only physical fitness trainup of from 4 to 8 weeks. This would close the physical readiness gap between men and women before integrated training starts. Women would be indoctrinated into the infantry physical workload by training against men's APFT standards

Our service academies and military colleges have taught us [that once] women comprise more than 20 percent of a unit or class, they are judged as individuals and not as representatives of their gender. Successful women cadets in the group become fellow cadets, not female cadets.

with weight training and rucksack marching. Only those women who meet the minimum standards of the men's APFT would graduate. From there, graduates would be integrated into infantry OSUT at Fort Benning, Georgia, along with the men trainees. Many women could start OSUT in better physical shape than the male trainees. This should improve initial unit cohesion by inspiring competition and respect among the untrained men and the physically ready women.

No fewer than eight women trainees, or a sufficient number to reach 20 percent of the whole, would be assigned to any training platoon—approximately 40 soldiers—at OSUT. This is consistent with social scientists' critical mass observation on the number of minorities in a majority group that are necessary to ensure adequate peer support for the minority and acceptance by the majority.

Women in OSUT would receive the same haircuts as the men, would not be allowed to wear makeup, and would compete against the same physical standards as the men. This uses lessons learned from successful gender integration at the Virginia Military Institute (VMI) in the mid-1990s. VMI benefited from the hard lessons The Citadel learned after its much-publicized forced integration of Shannon Faulkner in 1995. VMI treated the women like the men but monitored harassment. VMI leaders discovered that the men immediately accepted women who succeeded under those conditions.³⁶

Cadre at OSUT would include women drill sergeants and officers who had successfully completed infantry OSUT or the officer equivalent. Each training company with women basic trainees would have at least one woman drill sergeant and one officer. The initial low numbers of women in OSUT could

result in concentrating all women trainees into one training company. That would not be counterproductive. In fact, it might foster peer support and the majority's peer approval.

Women in OSUT would train to achieve the men's standards throughout training but would meet the women's standards in their age groups to pass the final APFT. Women would meet all other standards required of men, including the 5-mile run in 45 minutes and all road marches carrying the same equipment as the men.³⁷

Upon graduation, women infantry soldiers would be grouped into cohorts and assigned to the same field unit. The 20-percent guideline would be strictly adhered to. If a battalion received a cohort of six women, all would be assigned to one company and one platoon. This 20 percent guideline would be an integration tool rather than a permanent procedure. Once women were successfully integrated into the infantry, they would be individually assigned and re-assigned just as other soldiers are.

Women infantry officer trainees would precede women enlisted infantry trainees. Women at the U.S. Military Academy; in the Reserve Officer Training Course; and in Officer Candidate School would be allowed to enter the infantry branch. Upon commissioning, women officers would attend the Infantry Officer Basic Course (IOBC) at Fort Benning. They would continue to be assigned using the 20-percent guideline. Women IOBC graduates would then go

to Ranger School, using the principles described. For all female infantry officer training, women cadre members would be essential, especially at Ranger School. Successfully integrating women officers into infantry units would establish a path for young women to follow and ensure commissioned officer support.

Finally, each unit containing women infantry soldiers would designate a field grade officer within that unit to issue integration instruction and to oversee gender integration. The field grade officer would also coach, teach, and mentor women infantry soldiers and their leaders throughout the process. Company commanders would be directly responsible for their women infantry soldiers' training and welfare.

If integrating women into the infantry proceeded as outlined, not only would women succeed, but sexual harassment in the Army would decline. The greater respect women earn for themselves as true equals with military men will foster this.

The infantry will most likely struggle at first, but once women reach critical mass in units and in the overall infantry, unit cohesion and combat readiness should improve in ways we now probably cannot imagine. Change is part and parcel of the U.S. Army. Just as racial integration in the Armed Forces was considered dangerous 50 years ago, integrating females into ground combat units seems crazy today. For sure, the debate will continue, and opponents will continue to fight it. However, they are running out of solid arguments as well as time. **MR**

NOTES

1. Gerry J. Gilmore, "Key Factor in Army's Gender Issue: How Leaders Adjust to Change," *Army LINK News*, 18 September 1997, at <www.dtic.mil/armylink/news/Sep1997/a19970918footstg2.html>.
2. D'Ann Campbell, "Combating the Gender Gulf," *MINERVA: Quarterly Report* (Fall/Winter 1992): 26.
3. Linda D. Kozaryn, "Families, Friends Honor Desert Storm Fallen," *Fort Detrick Standard*, 27 February 2001, at <www.dcmilitary.com/armystandard/6_05national_news/5436-1.html>.
4. U.S. Army Field Manual (FM) 100-5, *Operations*, "Purposes of the Offense" (Washington, DC: U.S. Government Printing Office (GPO), June 1993), 7-0.
5. Anna Simons, "Women in Combat Units: It's Still a Bad Idea," *Parameters* (Summer 2001): 89.
6. Campbell, 13-31.
7. Brian Mitchell, *Women in the Military: Flirting With Disaster* (Washington, DC: Regency Publishing, 1998), 5.
8. Flora Davis, *Moving the Mountain: The Women's Movement in America Since 1960* (New York: Simon and Schuster, 1991), 215.
9. "World All-Time Best Performances—Marathon," The Athletics Site, 6 October 2001, at <www.athletix.net/Athletics/Statistics/wrMarathon>.
10. "World All-Time Best Performances—Pole Vault," The Athletics Site, 6 October 2001, at <www.athletix.net/Athletics/Statistics/wrPV.html>.
11. "World Weightlifting Records," International Weightlifting Federation, 6 October 2001, at <www.iwf.net/wrec/wrec.html>.
12. William P. Ebben and Randall L. Jensen, "Strength Training for Women: Debunking the Myths That Block Opportunity," *The Physician and Sportsmedicine* (May 1998): 2, at <www.physportsmed.com/issues/1998/05may/ebben.html>.
13. *Ibid.*
14. Louis F. Tomasi, Rey Regueros, Gene Fober, and Matthew Christenson, "Age, Gender, and Performance on the U.S. Army Physical Fitness Test (APFT)," *1995 Army Physical Fitness Test Update Survey* (Fort Benning, GA: U.S. Army Physical Fitness School, 1995), 1-8, at <www.benning.army.mil/usapfts/Research/1995APFTUpdateSurvey.htm>.
15. FM 21-20, *Physical Fitness Training* (Washington, DC: GPO, 30 September 1992), chap. 14.
16. Tomasi et al., 2.
17. Campbell, 19.

18. *Ibid.*, 21.
19. Stephanie Gutmann, *The Kinder, Gentler Military: Can America's Gender-Neutral Fighting Force Still Win Wars?* (New York: Scribner, 2000), 272-73.
20. Davis, 396-98.
21. U.S. Census Bureau, "Table DP-1: Profiles of General Demographic Characteristics for the United States, 2000," 8 October 2001, at <www.census.gov/dmd/www/database.html>.
22. Sherie Mershon and Steven Schlossman, *Foxholes and Color Lines: Desegregating the U.S. Armed Forces* (Baltimore, MD: Johns Hopkins University Press, 1996), 234-40.
23. Kim Field and John Nagl, "Combat Roles for Women: A Modest Proposal," *Parameters* (Summer 2001): 83.
24. Simons, 95.
25. Field and Nagl, 83.
26. *Ibid.*
27. D'Ann Campbell with Francine D'Amico, "Lessons on Gender Integration from the Military Academies," *Gender Camouflage: Women and the U.S. Military* (New York: New York University Press, 1999), 67.
28. Lieutenant Colonel David K. Leit, Canadian Forces, "Equal Rights and U.S. Combat Exclusion Policy for Women: A Congressional Lobby Confrontation" (master's thesis, Maxwell Air Force Base, Air War College, May 1988), 14.
29. Gutmann, 204-65.
30. William T. Stooksbury, "The Marine Corps' Infantry Selection and Assignment Process: Is It Ready for Gender Neutrality?" (master's thesis, Naval Postgraduate School, March 1994), 61-69.
31. Julie Demers, "Looking for Able Women: The Canadian Army's Female Recruitment Campaign," 27 September 2001, at <www.geocities.com/Wellesley/3321/win10c.htm>.
32. *Ibid.*
33. *Ibid.*, 1.
34. *Ibid.*, 2.
35. *Ibid.*, 1.
36. Campbell with D'Amico, 77.
37. Infantry OSUT Graduation Requirements (Fort Benning, GA: Infantry Training Brigade, 8 October 2001), at <www.benning.army.mil/itb/GRADREQ.htm>.

Captain Adam N. Wojack, U.S. Army, is the assistant operations officer, S3, 2d Brigade, 1st Infantry Division, Schweinfurt, Germany. He received a B.A. from Baruch College of the City University of New York. He has also served as S3 Air, 1st Battalion, 21st Infantry Regiment; rifle platoon leader, 1st Battalion, 21st Regiment; assistant S4, 2d Brigade; and anti-tank platoon leader, 1st Brigade, 14th Infantry Regiment, 25th Infantry Division, Schofield Barracks, Hawaii. He served with the 101st Airborne Division during the Persian Gulf war.

Copyright © 2003 EBSCO Publishing

Marine Corps Gazette, July 2012

Get Over It! We Are Not All Created Equal



Photo by Capt Katie Petronio

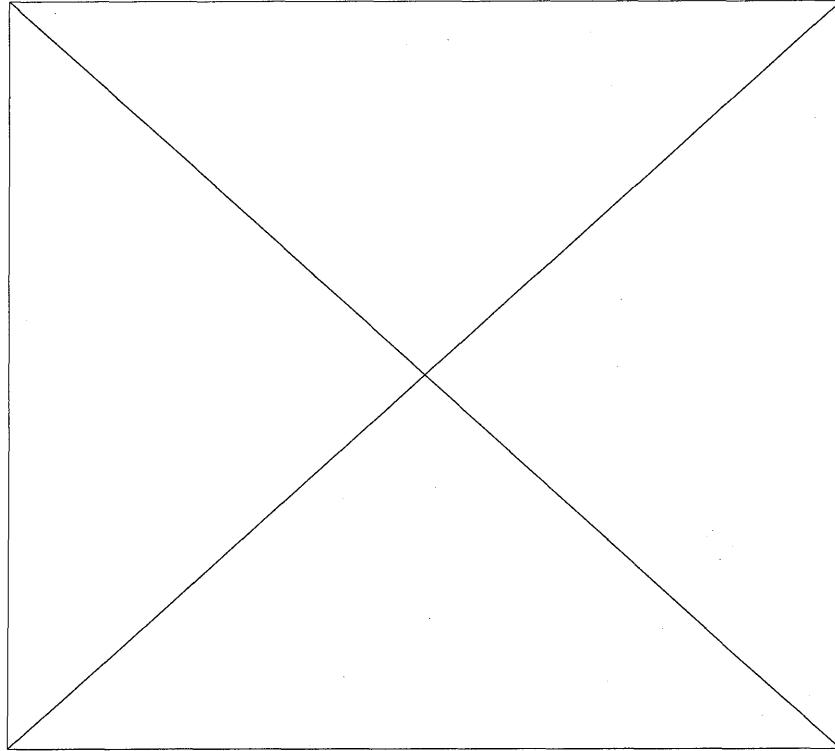
Description:

Women can conduct and lead combat operations; that is not the issue.

Author:

Capt Katie Petronio

The Marine Corps Times recently published a handful of articles in regard to opening Infantry Officer Course (IOC) to females and the possibility of integrating women into the infantry community. In mid-April the Commandant directed the “integration” of the first wave of female officers into IOC this summer following completion of The Basic School (TBS). This action may or may not pave the way for female Marines to serve in the infantry as the results remain to be seen. However, before the Marine Corps moves forward with this concept, should we not ask the hard questions and gain opinions of combat-experienced Marines (male and female alike) as to the purpose, the impact, and the gains from such a move? As a combat-experienced Marine officer, and a female, I am here to tell you that we are not all created equal, and attempting to place females in the infantry will not improve the Marine Corps as the Nation’s force-in-readiness or improve our national security.



As a company grade 1302 combat engineer officer with 5 years of active service and two combat deployments, one to Iraq and the other to Afghanistan, I was able to participate in and lead numerous combat operations. In Iraq as the II MEF Director, Lioness Program, I served as a subject matter expert for II MEF, assisting regimental and battalion commanders on ways to integrate female Marines into combat operations. I primarily focused on expanding the mission of the Lioness Program from searching females to engaging local nationals and information gathering, broadening the ways females were being used in a wide variety of combat operations from census patrols to raids. In Afghanistan I deployed as a 1302 and led a combat engineer platoon in direct support of Regimental Combat Team 8, specifically operating out of the Upper Sangin Valley. My platoon operated for months at a time, constructing patrol bases (PBs) in support of 3d Battalion, 5th Marines; 1st Battalion, 5th Marines; 2d Reconnaissance Battalion; and 3d Battalion, 4th Marines. This combat experience, in particular, compelled me to raise concern over the direction and overall reasoning behind opening the 03XX field.

Who is driving this agenda? I am not personally hearing female Marines, enlisted or officer, pounding on the doors of Congress claiming that their inability to serve in the infantry violates their right to equality. Shockingly, this isn't even a congressional agenda. This issue is being pushed by several groups, one of which is a small committee of civilians appointed by the Secretary of Defense called the Defense Advisory Committee on Women in the Service (DACOWITS). Their mission is to advise the Department of Defense (DoD) on recommendations, as well as matters of policy, pertaining to the well-being of women in the Armed Services from recruiting to employment. Members are selected based on their prior military experience or experience with women's workforce issues. I certainly applaud and appreciate DACOWITS'

mission; however, as it pertains to the issue of women in the infantry, it's very surprising to see that none of the committee members are on active duty or have any recent combat or relevant operational experience relating to the issue they are attempting to change. I say this because, at the end of the day, it's the active duty servicemember who will ultimately deal with the results of their initiatives, not those on the outside looking in. As of now, the Marine Corps hasn't been directed to integrate, but perhaps the Corps is anticipating the inevitable—DoD pressuring the Corps to comply with DACOWITS' agenda as the Army has already "rogered up" to full integration. Regardless of what the Army decides to do, it's critical to emphasize that we are not the Army; our operational speed and tempo, along with our overall mission as the Nation's amphibious force-in-readiness, are fundamentally different than that of our sister Service. By no means is this distinction intended as disrespectful to our incredible Army. My main point is simply to state that the Marine Corps and the Army are different; even if the Army ultimately does fully integrate all military occupational fields, that doesn't mean the Corps should follow suit.

I understand that there are female servicemembers who have proven themselves to be physically, mentally, and morally capable of leading and executing combat-type operations; as a result, some of these Marines may feel qualified for the chance of taking on the role of 0302. In the end, my main concern is not whether women are capable of conducting combat operations, as we have already proven that we can hold our own in some very difficult combat situations; instead, my main concern is a question of longevity. Can women endure the physical and physiological rigors of sustained combat operations, and are we willing to accept the attrition and medical issues that go along with integration?

As a young lieutenant, I fit the mold of a female who would have had a shot at completing IOC, and I am sure there was a time in my life where I would have volunteered to be an infantryman. I was a star ice hockey player at Bowdoin College, a small elite college in Maine, with a major in government and law. At 5 feet 3 inches I was squatting 200 pounds and benching 145 pounds when I graduated in 2007. I completed Officer Candidates School (OCS) ranked 4 of 52 candidates, graduated 48 of 261 from TBS, and finished second at MOS school. I also repeatedly scored far above average in all female-based physical fitness tests (for example, earning a 292 out of 300 on the Marine physical fitness test). Five years later, I am physically not the woman I once was and my views have greatly changed on the possibility of women having successful long careers while serving in the infantry. I can say from firsthand experience in Iraq and Afghanistan, and not just emotion, that we haven't even begun to analyze and comprehend the gender-specific medical issues and overall physical toll continuous combat operations will have on females.

I was a motivated, resilient second lieutenant when I deployed to Iraq for 10 months, traveling across the Marine area of operations (AO) and participating in numerous combat operations. Yet, due to the excessive amount of time I spent in full combat load, I was diagnosed with a severe case of restless leg syndrome. My spine had compressed on nerves in my lower back causing neuropathy which compounded the symptoms of

restless leg syndrome. While this injury has certainly not been enjoyable, Iraq was a pleasant experience compared to the experiences I endured during my deployment to Afghanistan. At the beginning of my tour in Helmand Province, I was physically capable of conducting combat operations for weeks at a time, remaining in my gear for days if necessary and averaging 16-hour days of engineering operations in the heart of Sangin, one of the most kinetic and challenging AOs in the country. There were numerous occasions where I was sent to a grid coordinate and told to build a PB from the ground up, serving not only as the mission commander but also the base commander until the occupants (infantry units) arrived 5 days later. In most of these situations, I had a sergeant as my assistant commander, and the remainder of my platoon consisted of young, motivated NCOs. I was the senior Marine making the final decisions on construction concerns, along with 24-hour base defense and leading 30 Marines at any given time. The physical strain of enduring combat operations and the stress of being responsible for the lives and well-being of such a young group in an extremely kinetic environment were compounded by lack of sleep, which ultimately took a physical toll on my body that I couldn't have foreseen.

By the fifth month into the deployment, I had muscle atrophy in my thighs that was causing me to constantly trip and my legs to buckle with the slightest grade change. My agility during firefights and mobility on and off vehicles and perimeter walls was seriously hindering my response time and overall capability. It was evident that stress and muscular deterioration was affecting everyone regardless of gender; however, the rate of my deterioration was noticeably faster than that of male Marines and further compounded by gender-specific medical conditions. At the end of the 7-month deployment, and the construction of 18 PBs later, I had lost 17 pounds and was diagnosed with polycystic ovarian syndrome (which personally resulted in infertility, but is not a genetic trend in my family), which was brought on by the chemical and physical changes endured during deployment. Regardless of my deteriorating physical stature, I was extremely successful during both of my combat tours, serving beside my infantry brethren and gaining the respect of every unit I supported. Regardless, I can say with 100 percent assurance that despite my accomplishments, there is no way I could endure the physical demands of the infantrymen whom I worked beside as their combat load and constant deployment cycle would leave me facing medical separation long before the option of retirement. I understand that everyone is affected differently; however, I am confident that should the Marine Corps attempt to fully integrate women into the infantry, we as an institution are going to experience a colossal increase in crippling and career-ending medical conditions for females.

There is a drastic shortage of historical data on female attrition or medical ailments of women who have executed sustained combat operations. This said, we need only to review the statistics from our entry-level schools to realize that there is a significant difference in the physical longevity between male and female Marines. At OCS the attrition rate for female candidates in 2011 was historically low at 40 percent, while the male candidates attrite at a much lower rate of 16 percent. Of candidates who were dropped from training because they were injured or not physically qualified, females were breaking at a much higher rate than males, 14 percent versus 4 percent. The same

trends were seen at TBS in 2011; the attrition rate for females was 13 percent versus 5 percent for males, and 5 percent of females were found not physically qualified compared with 1 percent of males. Further, both of these training venues have physical fitness standards that are easier for females; at IOC there is one standard regardless of gender. The attrition rate for males attending IOC in 2011 was 17 percent. Should female Marines ultimately attend IOC, we can expect significantly higher attrition rates and long-term injuries for women.

There have been many working groups and formal discussions recently addressing what changes would be necessary to the current IOC period of instruction in order to accommodate both genders without producing an underdeveloped or incapable infantry officer. Not once was the word “lower” used, but let’s be honest, “modifying” a standard so that less physically or mentally capable individuals (male or female) can complete a task is called “lowering the standard”! The bottom line is that the enemy doesn’t discriminate, rounds will not slow down, and combat loads don’t get any lighter, regardless of gender or capability. Even more so, the burden of command does not diminish for a male or female; a leader must gain the respect and trust of his/her Marines in combat. Not being able to physically execute to the standards already established at IOC, which have been battle tested and proven, will produce a slower operational speed and tempo resulting in increased time of exposure to enemy forces and a higher risk of combat injury or death. For this reason alone, I would ask everyone to step back and ask themselves, does this integration solely benefit the individual or the Marine Corps as a whole, as every leader’s focus should be on the needs of the institution and the Nation, not the individual?

Which leads one to really wonder, what is the benefit of this potential change? The Marine Corps is not in a shortage of willing and capable young male second lieutenants who would gladly take on the role of infantry officers. In fact we have men fighting to be assigned to the coveted position of 0302. In 2011, 30 percent of graduating TBS lieutenants listed infantry in their top three requested MOSs. Of those 30 percent, only 47 percent were given the MOS. On the other hand, perhaps this integration is an effort to remove the glass ceiling that some observers feel exists for women when it comes to promotions to general officer ranks. Opening combat arms MOSs, particularly the infantry, such observers argue, allows women to gain the necessary exposure of leading Marines in combat, which will then arguably increase the chances for female Marines serving in strategic leadership assignments. As stated above, I have full faith that female Marines can successfully serve in just about every MOS aside from the infantry. Even if a female can meet the short-term physical, mental, and moral leadership requirements of an infantry officer, by the time that she is eligible to serve in a strategic leadership position, at the 20-year mark or beyond, there is a miniscule probability that she’ll be physically capable of serving at all. Again, it becomes a question of longevity.

Despite my personal opinion regarding the incorporation of females into the infantry community, I am not blind to the fact that females play a key role in countering the gender and cultural barriers we are facing at war, and we do have a place in combat operations. As such, a potential change that I do recommend considering strongly for

female Marine officers is to designate a new secondary MOS (0305) for a Marine serving as female engagement team (FET) officer in charge (OIC). 0305s would be employed in the same way we employ drill instructors, as we do not need an enduring FET entity but an existing capability able to stand up based on operational requirements. Legitimizing a program that is already operational in the Corps would greatly benefit both the units utilizing FETs and the women who serve as FET OICs. Unfortunately, FET OICs today are not properly screened and trained for this mission. I propose that those being considered for FET OIC be prescreened and trained through a modified IOC with an appropriately adjusted physical expectation. FET OICs need to better understand the infantry culture and mindset and work with their 0302 brethren to incorporate FET assistance during specific phases of operations to properly prepare them to serve as the subject matter experts to a regimental- or battalion-level infantry commander. Through joint OIC training, both 0302s and FET OICs can start to learn how to integrate capabilities and accomplish their mission individually and collectively. This, in my mind, is a much more viable, cost-effective solution, with high reward for the Marine Corps and the Nation, and it will also directly improve the capabilities of FET OICs.

Finally, what are the Marine Corps standards, particularly physical fitness standards, based on—performance and capability or equality? We abide by numerous discriminators, such as height and weight standards. As multiple Marine Corps Gazette articles have highlighted, Marines who can run first-class physical fitness tests and who have superior MOS proficiency are separated from the Service if they do not meet the Marine Corps' height and weight standards. Further, tall Marines are restricted from flying specific platforms, and color blind Marines are faced with similar restrictions. We recognize differences in mental capabilities of Marines when we administer the Armed Services Vocational Aptitude Battery and use the results to eliminate/open specific fields. These standards are designed to ensure safety, quality, and the opportunity to be placed in a field in which one can sustain and succeed.

Which once again leads me, as a ground combat-experienced female Marine Corps officer, to ask, what are we trying to accomplish by attempting to fully integrate women into the infantry? For those who dictate policy, changing the current restrictions associated with women in the infantry may not seem significant to the way the Marine Corps operates. I vehemently disagree; this potential change will rock the foundation of our Corps for the worse and will weaken what has been since 1775 the world's most lethal fighting force. In the end, for DACOWITS and any other individual or organization looking to increase opportunities for female Marines, I applaud your efforts and say thank you. However, for the long-term health of our female Marines, the Marine Corps, and U.S. national security, steer clear of the Marine infantry community when calling for more opportunities for females. Let's embrace our differences to further hone in on the Corps' success instead of dismantling who we are to achieve a political agenda. Regardless of the outcome, we will be "Semper Fidelis" and remain focused on our mission to protect and defend the United States of America.

Foreign Affairs

Let Women Fight

Ending the U.S. Military's Female Combat Ban

By Megan H. MacKenzie

From our November/December 2012 Issue



UPDATE: January 23, 2013

On Wednesday, the U.S. Department of Defense announced that it would lift the ban on women in combat. This landmark decision reverses the 1994 "direct ground combat rule," which held that "women shall be excluded from assignment to units below the brigade level whose primary mission is to engage in direct combat on the ground."

The policy change is long overdue. The last few decades had made the ban largely irrelevant; increasing counterinsurgency warfare virtually erased the concept of combat front lines and female soldiers' contributions to the wars in Afghanistan and Iraq were undeniable. The policy had nevertheless continued to officially exclude women from 7.3 percent of army positions, largely in Infantry, Armor, and Special Forces. More importantly, it had limited women's career paths and promotion opportunities and contributed to gendered stereotypes about war as ultimately "the business" of men.

There are physically fit, tough women who are suitable for combat, and weak, feeble men who are not.

The decision to remove the exclusion now is a sound one based on careful consideration of several factors. Specifically, there was increasing support from within the military leadership -- including from Leon Panetta, Secretary of Defense, and Martin Dempsey, Chairman of the Joint Chiefs, who had both acknowledged that now is the time to remove gender-based barriers to service. There have also been distinct changes in public attitudes about women's capabilities and roles in war. According to a Washington Post-ABC News poll, the vast majority of Americans support allowing women into combat roles. Meanwhile, studies by the U.S. Army Research Institute of Environmental Medicine and the U.S. Government Accountability Office, along with various military and academic experts, have dispelled myths about women's impact on unit cohesion and their physical abilities. It could not have hurt, of course, that the Department of Defense is facing a lawsuit from several female service members (backed by the American Civil Liberties Union) who rightly claim that the exclusion was discriminatory and unconstitutional.

By changing the policy now, the Department of Defense is recognizing women's contributions to the wars in Iraq and Afghanistan and acknowledging that times have changed, both in terms of the ways wars are fought and in terms of attitudes about appropriate roles for women in the forces. The next battles for female soldiers will be ensuring that this policy is implemented effectively, stamping out any remaining sexist attitudes, and fighting to ensure that the military addresses its outstanding sexual violence problem.

ORIGINAL ARTICLE: November/December 2012

Today, 214,098 women serve in the U.S. military, representing 14.6 percent of total service members. Around 280,000 women have worn American uniforms in Afghanistan and Iraq, where 144 have died and over 600 have been injured. Hundreds of female soldiers have received a Combat Action Badge, awarded for actively engaging with a hostile enemy. Two women, Sergeant Leigh Ann Hester and Specialist Monica Lin Brown, have been awarded Silver Stars -- one of the highest military decorations awarded for valor in combat -- for their service in Afghanistan and Iraq.

Yet the U.S. military, at least officially, still bans women from serving in direct combat positions. As irregular warfare has become increasingly common in the last few decades, the difference on the ground between the frontline and support roles is no longer clear. Numerous policy changes have also eroded the division between combat and noncombat positions. More and more military officials recognize the contributions made by female soldiers, and politicians, veterans, and military experts have all begun actively lobbying Washington to drop the ban. But Congress has not budged.

Proponents of the policy, who include Duncan Hunter (R-Calif.), former chair of the House Armed Services Committee, and former Senator Rick Santorum (R-Pa.), rely on three central arguments: that women cannot meet the physical requirements necessary to fight, that they simply don't belong in combat, and that their inclusion in fighting units would disrupt those units' cohesion and battle readiness. Yet these arguments do not stand up to current data on women's performance in combat or their impact on troop dynamics. Banning women from combat does not ensure military effectiveness. It only perpetuates counterproductive gender stereotypes and biases. It is time for the U.S. military to get over its hang-ups and acknowledge women's rightful place on the battlefield.

WOMEN IN A MAN'S WORLD

Women have long served in various auxiliary military roles during wars. Further, the 1948 Women's Armed Services Integration Act created a permanent corps of women in all the military departments. This was considered a step forward at the time, but it is also the origin of the current combat ban. The act limited women's number to two percent of total service members and formally excluded them from combat duties. The exclusion policy was reinforced in 1981, when the U.S. Supreme Court ruled that the all-male draft did not constitute gender-based discrimination since it was intended to increase combat troops and women were already restricted from combat.

Despite this restriction, the share of women in the U.S. armed forces increased in the 1980s and 1990s, from 8.5 percent to 11.1 percent, as a result of the transition to an all-volunteer force in 1973 and high demand for troops. Today, the air force is the most open service for women. Women have been flying in combat aircrafts since 1993, and they now make up 70 of the 3,700 fighter pilots in the service.

In the rest of the military, restrictions on women have also been slipping for some time, albeit more slowly, due to an increase in female enlistment and the public's growing sensitivity to equal labor rights. In January 1994, a memorandum from then Secretary of Defense Les Aspin rescinded the "risk rule" barring women from any positions that could expose them to direct combat, hostile fire, or capture; the rule was replaced by the "direct ground combat assignment rule," which more narrowly tailored the restriction to frontline combat positions.

The main obstacle for women who want to serve their country is an outmoded set of biased assumptions about their place in society.

Recent policy changes have also blurred the distinction between combat and support roles. In 2003, the army began reorganizing units and increasing the number of brigades within each division. Under this system, forward support companies, which provide logistical support, transportation, and maintenance to battalions, are now grouped together on the same bases as combat units. Since women are permitted to serve in such support units, a major barrier designed to keep them away from combat has almost vanished.

The assignment of women to combat-related tasks has further undermined the strength of the ban. Beginning in 2003, for example, so-called Lioness teams were deployed to assist combat units in Iraq searching women for weapons and explosives. Drawing from this model, the military created several other female-only units in 2009, including "female engagement teams." In their first year of operation, these teams conducted over 70 short-term search-and-engagement missions in Afghanistan. Paying lip service to the exclusion policy, the military specified that these units could not contribute to hunt-and-kill foot patrols and should stay at combat bases only temporarily. In practice, however, this meant that female soldiers were required to leave their combat bases for one night every six weeks before immediately returning. Not only did this practice put women at risk

with unnecessary travel in an insecure environment; it also exemplifies the waste and hardship that the preservation of the formal ban imposes on the military.

Meanwhile, the U.S. military is finding different ways to recognize the fact that women now fight in the country's wars. Members of forward support companies and female engagement teams now receive combat pay, also known as "hostile fire" or "imminent danger" pay, acknowledging the threats women regularly face. And 78 percent of the deaths of female U.S. service members in Iraq were categorized as hostile, yet another sign of how American women in uniform regularly put their lives at risk.

In light of all these changes, in 2011 the Military Leadership Diversity Commission recommended that the Department of Defense remove all combat restrictions on women. Although the total number of jobs closed to women is now relatively low, at 7.3 percent, the commission found that "exclusion from these occupations has a considerable influence on advancement to higher positions" and that eliminating the exclusion is essential "to create a level playing field for all service members who meet the qualifications." Echoing this sentiment, Senator Kirsten Gillibrand (D-N.Y.) introduced the Gender Equality in Combat Act in 2012, which seeks the termination of the ground combat exclusion policy. In addition, Command Sergeant Major Jane Baldwin and Colonel Ellen Haring, both of the Army Reserve, filed a lawsuit in May against the secretary of defense and the army's secretary, assistant secretary, and deputy chief of staff claiming that the exclusion policy violates their constitutional rights.

Responding to growing scrutiny, the Pentagon's press secretary, George Little, announced on February 9, 2012, that the Department of Defense would continue to remove restrictions on women's roles. Since then, the military has made a slew of policy revisions and commissioned a series of reviews. In May 2012, for example, the army opened up more than 14,000 combat-related jobs to women. Much of this increase, however, came from officially recognizing the combat-related nature of the jobs conducted by medics and intelligence officers, among others, positions that are already open to women. More substantially, the Marine Corps announced in April 2012 that for the first time, women can enroll and train, but not yet serve, as infantry combat officers. The army has also opened six new combat-related occupational specialties to women. In June 2012, Cicely Verstein became the first woman to serve in one of these newly opened combat support roles when she enlisted as a Bradley Fighting Vehicle systems maintainer. Women such as Verstein can now operate with combat arms units in select positions, yet they are still technically restricted from infantry and special operations roles.

Although the ban still exists on paper, the military is finding various ways to lift it in practice, and so the complete repeal of the policy would not constitute a radical change in operational terms. But it would be an acknowledgment of the contributions that women are already making to U.S. military operations. As Anu Bhagwati, a former Marine captain and now executive director of the Service Women's Action Network, explained in a BBC News interview, "Women are being shot at, are being killed overseas, are being attached to all of these combat arms units. . . . The [combat exclusion] policy has to catch

up to reality.” Indeed, all soldiers, female as well as male, have been given extensive combat training since 2003, when the army altered its basic training procedures in response to the growth of irregular warfare in Afghanistan and Iraq. The main obstacle that remains for women who want to serve their country is an outmoded set of biased assumptions about their capabilities and place in society.

WHY WOMEN CAN KEEP UP

The argument that women are not physically fit for combat is perhaps the most publicized and well-researched justification for their exclusion from fighting units. In her 2000 book, *The Kinder, Gentler Military*, the journalist Stephanie Gutmann summarized the position this way: “When butts drop onto seats, and feet grope for foot pedals, and girls of five feet one (not an uncommon height in the ranks) put on great bowl-like Kevlar helmets over a full head of long hair done up in a French braid, there are problems of fit - - and those picayune fit problems ripple outward, eventually affecting performance, morale, and readiness.”

This argument continues to receive a significant amount of attention in the United States, despite the fact that other militaries across the world have found that with proper training and necessary adaptations, women can complete the same physical tasks as men. In the 1970s, the Canadian military conducted trials that tested women’s physical, psychological, and social capacity for combat roles. The results informed the final decision of the Canadian Human Rights Tribunal to remove Canada’s female combat exclusion. After similar tests, Denmark also lifted its combat ban in the late 1980s.

The physical fitness argument, which tends to focus on differences between average male and female bodies, is also undermined by the fact that women who join the military tend to be more fit than the average American. Additional training and conditioning further decrease the gap between female and male service members, and evidence indicates that women usually benefit substantially from fitness-training programs. More to the point, performance is not necessarily determined by gender; it is determined by other attributes and by an individual’s determination to reach physical prowess. To put it bluntly, there are physically fit, tough women who are suitable for combat, and weak, feeble men who are not.

The U.S. armed services would do a better job recognizing this were it not for the fact that, as critics have pointed out, the military’s physical standards were created to measure male fitness, not job effectiveness. As Matthew Brown, a U.S. Army colonel and director of the Arizona Army National Guard, found in a U.S. Army War College study, “There is no conclusive evidence that all military members, regardless of occupational specialty, unit assignment, age or gender, should acquire the same level of physical fitness.” The U.S. General Accounting Office (now the Government Accountability Office) also admitted in a 1998 report that physical fitness tests are not necessarily a useful gauge of operational effectiveness, explaining, “fitness testing is not aimed at assessing the capability to perform specific missions or military jobs.” To be sure, men and women have different types of bodies, but growing research points to the limitations of having a

single male-centered standard for fitness and equipment. Recently, for example, the army has moved to design body armor for women rather than force them to continue wearing equipment that restricts their movement and cuts into their legs because it was designed for men. With proper training and equipment, women can contribute to missions just as well as men.

BREAKING UP THE BAND OF BROTHERS

Even though the physical argument does not hold up to scrutiny, many in the military establishment continue to instinctively oppose the idea of women serving in combat roles. In a 1993 *New York Times* article, General Merrill McPeak, former chief of staff of the air force, admitted that he had “a culturally based hang-up.” “I can’t get over this image of old men ordering young women into combat,” he said. “I have a gut-based hang-up there. And it doesn’t make a lot of sense in every way. I apologize for it.” This belief had earlier been spelled out in the 1992 report of the Presidential Commission on the Assignment of Women in the Armed Forces, which was established by George H. W. Bush to review the combat exclusion. The commission identified several factors related to having women serve in combat roles that could negatively impact troop dynamics, including the “real or perceived inability of women to carry their weight without male assistance, a ‘zero privacy’ environment on the battlefield, interference with male bonding, cultural values and the desire of men to protect women, inappropriate male/female relationships, and pregnancy -- particularly when perceived as a way to escape from combat duty.”

While campaigning for the Republican presidential nomination this year, Santorum, the former senator, echoed these concerns, arguing that “instead of focus[ing] on the mission, [male soldiers] may be more concerned about protecting . . . a [female soldier] in a vulnerable position.” Others fear that men will not be able to restrain themselves sexually if forced to fight and work in close proximity to women. The conservative Independent Women’s Forum strongly supports the ban because of the “power of the sex drive when young women and men, under considerable stress, are mixed together in close quarters.”

Even as these false assumptions about the inherent nature of men and women persist, many in the military and the general public have changed their minds. In 2010, Admiral Mike Mullen, then chairman of the U.S. Joint Chiefs of Staff, said, “I know what the law says and I know what it requires, but I’d be hard pressed to say that any woman who serves in Afghanistan today or who’s served in Iraq over the last few years did so without facing the same risks as their male counterparts.” Similarly, Bhagwati contends that “as proven by ten years of leading troops in combat in Iraq and Afghanistan, there are women that are physically and mentally qualified to succeed . . . and lead infantry platoons.” Meanwhile, a 2011 survey conducted by ABC News and *The Washington Post* found that 73 percent of Americans support allowing women in combat.

Despite such shifts in opinion, defenders of the status quo argue that lifting the ban would disrupt male bonding and unit cohesion, which is thought to build soldiers’ confidence and thereby increase combat readiness and effectiveness. In 2007, Kingsley Browne, a

former U.S. Supreme Court clerk and the author of *Co-ed Combat: The New Evidence That Women Shouldn't Fight the Nation's Wars*, argued that “men fight for many reasons, but probably the most powerful one is the bonding -- ‘male bonding’ -- with their comrades. . . . Perhaps for very fundamental reasons, women do not evoke in men the same feelings of comradeship and ‘followership’ that men do.” These comments betray the widely held fear that women would feminize and therefore reduce the fighting potential of the military. The Israeli military historian Martin van Creveld has echoed this sentiment, writing, “As women enter them, the armed forces in question will become both less willing to fight and less capable of doing so.” And as Anita Blair, former assistant secretary of the navy, warned, “The objective for many who advocate a greater female influence in the armed services is not so much to conquer the military as conquer manhood: they aim to make the most quintessentially masculine of our institutions more feminine.” By such lights, women fundamentally threaten the unified masculine identity of the military and could never properly fill combat roles because they are inherently incapable of embodying the manly qualities of a soldier.

This argument is intuitive and plausible. It is also dead wrong. It assumes that a key objective of the military is enhancing masculinity rather than national security and that unit bonding leads to better task performance. In fact, a 1995 study conducted by the U.S. Army Research Institute for the Behavioral and Social Sciences found that “the relation between cohesiveness and performance is due primarily to the ‘commitment to the task’ component of cohesiveness, and not the ‘interpersonal attraction’ or ‘group pride’ components of cohesiveness.” Similarly, a 2006 study in *Armed Forces and Society*, written by the scholars Robert MacCoun, Elizabeth Kier, and Aaron Belkin, concluded that “all of the evidence indicates that military performance depends on whether service members are committed to the same professional goals, not on whether they like one another.”

There is significant evidence that not only male bonding but any sort of closeness can actually hinder group performance. In a 1998 study on demographics and leadership, the group management experts Andrew Kakabadse and Nada Kakabadse found that “excessive cohesion may create a harmful insularity from external forces,” and they linked high cohesion to “high conformity, high commitment to prior courses of actions, [and a] lack of openness.” In her analysis of gender integration in the military, Erin Solaro, a researcher and journalist who was embedded with combat troops in Afghanistan and Iraq, pointed out that male bonding often depended on the exclusion or denigration of women and concluded that “cohesion is not the same as combat effectiveness, and indeed can undercut it. Supposedly ‘cohesive’ units can also kill their officers, mutiny, evade combat, and surrender as groups.”

The mechanisms for achieving troop cohesion can also be problematic. In addition to denigrating women, illegal activities, including war crimes, have sometimes been used as a means for soldiers to “let off steam” and foster group unity. In sum, there is very little basis on which to link group cohesion to national security.

STRENGTH IN DIVERSITY

Over the last century, the military has been strengthened when attitudes have been challenged and changed. Despite claims in the 1940s that mixed-race units would be ineffective and that white and black service members would not be able to trust one another, for example, integration proceeded without any major hiccups. A 2011 study of the impacts of racial integration on combat effectiveness during the Korean War found that integration “resulted in improvements in cohesion, leadership and command, fighting spirit, personnel resources and sustainment that increased the combat effectiveness.” Initial research indicates that mixed-gender units could provide similar benefits.

Leora Rosen, a former senior analyst at the National Institute of Justice, found that when women were accepted into mixed-gender units, the groups’ effectiveness actually increased. Similarly, a 1993 RAND Corporation paper summarizing research on sexual orientation and the U.S. military’s personnel policy found that diversity “can enhance the quality of group problem-solving and decision-making, and it broadens the group’s collective array of skills and knowledge.” These conclusions are supported by a 1993 report by the General Accounting Office, which found that “members of gender-integrated units develop brother-sister bonds rather than sexual ones. . . . Experience has shown that actual integration diminishes prejudice and fosters group cohesiveness more effectively than any other factor.” The same report also found that gender homogeneity was not perceived by soldiers to be a requirement for effective unit operations.

It should come as no surprise that elements of the military want uniformity in the ranks. The integration of new groups always ruffles feathers. But the U.S. military has been ahead of the curve in terms of the inclusion of most minority groups. It was the first federal organization to integrate African Americans. And with the repeal of the “don’t ask, don’t tell” (DADT) policy, the military now has more progressive policies toward gay employees than many other U.S. agencies. In fact, DADT was repealed despite the fact that there are no federal laws preventing employment discrimination on the basis of sexual orientation.

In September 2012, one year after the repeal of DADT, a study published by the Palm Center found that the change “has had no overall negative impact on military readiness or its component dimensions, including cohesion, recruitment, retention, assaults, harassment or morale.” The research also found that overall, DADT’s “repeal has enhanced the military’s ability to pursue its mission.” Previous claims about the negative impact that gay service members might have on troop cohesion mirror those currently used to support the female combat exclusion.

Unlike the military’s treatment of other groups, its current policies toward women are much more conservative than those of other federal and state government bodies. Women who choose military service confront not only restricted career options but also a higher chance of harassment, discrimination, and sexual violence than in almost any other profession. The weak record on addressing these issues gives the impression that the military is an unwelcome place and an unsafe career choice for women. In an interview with National Public Radio in 2011, Sergeant Kayla Williams, who served in Iraq, explicitly linked the combat exclusion and harassment: “I believe that the combat

exclusion actually exacerbates gender tensions and problems within the military because the fact that women can't be in combat arms jobs allows us to be portrayed as less than fully soldiers." Fully integrating women could therefore begin to address two major issues for the U.S. military: enhancing diversity and equality and also weakening the masculine culture that may contribute to harassment.

Unsubstantiated claims about the distracting nature of women, the perils of feminine qualities, and the inherent manliness of war hardly provide a solid foundation on which to construct policy. Presumably, some levels of racism and homophobia also persist within the military, yet it would be absurd, not to mention unconstitutional, for the U.S. government to officially sanction such prejudices. The U.S. military should ensure that it is as effective as possible, but it must not bend to biases, bigotry, and false stereotypes.

Just as when African Americans were fully integrated into the military and DADT was repealed, lifting the combat ban on women would not threaten national security or the cohesiveness of military units; rather, it would bring formal policies in line with current practices and allow the armed forces to overcome their misogynistic past. In a modern military, women should have the right to fight.

Why Can't Anything Be Done?

Measuring Physical Readiness of Women for Military Occupations

Paper presented at the 2011 International Biennial Conference of the Inter-University Seminar on Armed Forces and Society

***Chicago, Illinois
21-23 October 2011***

***Dr. William J. Gregor
Professor of Social Sciences
School of Advanced Military Studies
US Army Command and General Staff College
Fort Leavenworth, Kansas 66027***

The opinions expressed in this paper are those of Dr. William J. Gregor and are not the views of the Department of Defense, the Department of the Army, the U.S. Army Command and General Staff College, or the School of Advanced Military Studies.

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

TABLE OF CONTENTS

TABLE OF CONTENTS	2
INTRODUCTION	3
A HOST OF REVIEWS	4
USE OF WOMEN IN THE MILITARY 1982	5
THE ASSIGNMENT OF WOMEN IN THE MILITARY 1992	6
KASSEBAUM-BAKER AND CONGRESS 1997-1998	8
THE ONLY PERFORMANCE STANDARD	10
COMBAT REQUIREMENTS	13
MEASURING PHYSICAL CAPACITY	14
ARMY ROTC APFT PERFORMANCE	17
THE PERFORMANCE GAP	21
MILITARY PERFORMANCE	28
REMEMBERING MILITARY READINESS	28
BIBLIOGRAPHY	32
APPENDIX A: Trends in ROTC APFT Performance 1992-2000	35
APPENDIX B: Discussion of Impact of New 1998 APFT Standards on Cadet APFT Scores (Presented to Cadet Command June 1998).....	37

INTRODUCTION

“After seven years of the all-volunteer armed forces and almost as many years of debate over the role of women in the military, one might ask why anyone would seek to add a single additional syllable to this deeply emotional, frequently acrimonious discussion.” That sentence opened an article entitled “Women, Combat, and the Draft”¹ published in a book entitled *Defense Manpower Planning: Issues for the 1980’s*. The title of the article reflected the then important legal and political argument of whether in reestablishing Selective Service Registration Congress was required to register women as well as men. However, why is the question just as apt today, some 30 years later? Why after some 30 years are the Congress, the Army, and the public still wrestling with the role of women in the military? Just as was the case in 1980, it is unlikely that any new article is going to move the front lines in this policy struggle. Nevertheless, at the risk of becoming yet another casualty in the war of words, it seems useful to revisit the questions posed in “Women, Combat and the Draft,” even if the product is only a litany of what we have failed to learn. Fortunately, a review of the government’s efforts to develop appropriate policy for the assignment of women in the armed forces is not as drab as that, though it is depressing. The story is depressing because the episodic policy reviews have focused more on the social dimensions of the force than on the functional requirements and, thus, the reviews have not advanced the Services’ understanding of how best to train and employ the force. The recently published report of the Military Leadership Diversity Commission, *From Representation to Inclusion: Diversity Leadership for the 21st-Century Military* is yet one more example of social engineering to make the force look right, but will it work right? Hence, the purpose of this article is once again to illuminate the questions that ought to have answers and to ask why can’t anything be done?

¹ William J. Gregor, “Women, Combat and the Draft: Placing Details in Context,” in William J. Taylor, Jr. and Eric T. Olson, ed., *Defense Manpower Planning: Issues for the 1980’s* (Elmsford, NY: Pergamon Press, Inc., 1981), 34.

A HOST OF REVIEWS

When “Women, Combat, and the Draft” was written, there was little evidence upon which to base policy for the assignment of women in the armed forces. The 1977 Department of Defense background study, *The Use of Women in the Military*, set as its goal to avoid emotionalism and to report what information was available and where DoD could expect to be in five years.² However, the study addressed only military manpower requirements and projections for expanding the percentage of women in the military. The issue of individual performance was not addressed. Curiously, in retrospect, *the Use of Women in the Military* established what have become the unchanging boundary of the policy debate. Although the study opened with the usual piety, “To put this study in context, one must remember that the overriding issue is maintaining the combat effectiveness of the armed forces.”³ It immediately stated the reasons for increasing the role of women in the military: the movement in society to provide equal economic opportunity for women and to meet the manpower needs of the all-volunteer force in the face of a declining youth population. The record now shows that the social concern for equality has dominated policy-making to such extent that the collection and evaluation of performance data, individual and collective, has been either overlooked or slighted. The policy reviews that have occurred have been occasioned largely by serious social problems in the training base and field or by flamboyant assessments of military operations. However, in no instance has performance data ever trumped the social concern for progress toward equality. That is not to say that the policy reviews did not invite testimony and collect performance data, only that the issues of performance, especially in the land forces, were discounted when decisions were made. However, in contrast to many of the previous reviews, the Military Leadership Diversity Commission

² Department of Defense, Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), *The Use of Women in the Military*, 2nd Ed. (Washington, D.C., September 1978), 1.

³ *Ibid.*

limited its attention entirely to social concerns, even noting that military eligibility requirements present a structural barrier to service.⁴ Thus, the diversity commission findings are devoted totally to achieving the commission's ideal representation of women and minorities in the military, performance requirements notwithstanding. They want the military to look right, whatever that means?

USE OF WOMEN IN THE MILITARY 1982

A 1976 Government Accounting Office Report, "Job Opportunities for Women in the Military: Progress and Problems," recommended that each Service develop physical standards for its jobs and standards for measuring strength, stamina, and other job requirements.⁵ Two years later an Army study observed, "The current lack of established performance standards, unvalidated critical tasks, and the absence of a system for measuring potential against standards precludes reliable determination of the physical capabilities soldiers of either sex must possess to do their job."⁶ Nevertheless, the Army did not address the matter again until 1982. Unlike the 1977 DoD background study, the 1982 *Women in the Army Policy Review* guided its work by two different principles. First, the review stated that the personnel policies must support fully combat readiness, and second, those policies should maximize the soldier's contribution to the Army's mission.⁷ The policy review recommended implementing a Military Enlistment Physical Strength Capacity Test (MEPSCAT) that would be used to match recruits to the physical demands of their jobs.⁸ The *Women in the Army Policy Review* reported that only 8 per cent of women were

⁴ Military Leadership Diversity Commission, *FROM REPRESENTATION TO INCLUSION: Diversity Leadership for the 21st Century* (Arlington, VA: March 15, 2011), 47.

⁵ Office of the Deputy Chief of Staff for Personnel, *Women in the Army Policy Review* (Washington, D.C., 12 November 1982), 2.

⁶ Department of the Army, *Evaluation of Women in the Army* (Washington, D.C., March 1978), 1-18.

⁷ *Women in the Army Policy Review*, 3.

⁸ *Ibid.*, 9.

capable of performing jobs in the heavy work category; i.e., frequent lifting of over 50 pounds and occasionally lifting of 100 pounds.⁹ Nevertheless, the study observed that 42 per cent of all Army women were assigned to a heavy work MOS. In other words, many of the women in 1982 were assigned to military specialties in which they could not physically perform the required work. Implementing MEPSCAT would have aligned women with appropriate specialties but it would have excluded women from a large number of specialties. The recommendations were not implemented. Setting physical performance standards based on military job requirements interfered with the expansion of the role of women in the military, so the idea was shelved.

THE ASSIGNMENT OF WOMEN IN THE MILITARY 1992

Success in the 1991 Gulf War spawned another review of the assignment of women in the military. The collective success of the military enterprise was seen as evidence that individual performance anywhere on the battlefield had also been superior. While this conclusion is logically insupportable, that did not deter a Democrat Congress intent on using the war outcome to justify increasing social equality. The Congress repealed most of the statutory restrictions on the assignment of women in the Navy, Marine Corps, and Air Force in December 1991. However, the expansion of the role of women was temporarily stalled by a call to appoint a Presidential Commission to investigate changes to policy on the assignment of women in the armed forces. The Commission's report, published November 15, 1992, finessed the intractable problem of female physical performance. The commissioners voted 12 to 0, with one abstention, to recommend that the Services retain gender-specific physical fitness tests to promote general wellness¹⁰ and they voted unanimously, 14-0, to recommend that the Services adopt gender-

⁹ Ibid., 2-16.

¹⁰ *The Presidential Commission on the Assignment of Women in the Armed Forces*, by GEN Robert T. Herres, Ret., Chairman (Washington, D.C., November 15, 1992), 5.

neutral muscular strength/endurance and cardiovascular standards for relevant specialties.¹¹ The commissioners split on entry-level training. By a vote of 8 to 6, with one abstention, the commission supported maintaining gender-specific training. These recommendations along with others such as the recommendation to continue excluding women from combat aircraft positions reflected that commission's somewhat conservative makeup, which showed itself by a slightly greater concern for individual physical performance. All of which was no matter because the report had no impact on Service assignment and training policies.

The Clinton Administration that took office in January 1993 and the new Democrat Congress easily dismissed the Presidential Commission's recommendations and quickly set about the task of revising the definition of direct combat and the related restrictions on the assignment of women. Congress facilitated the redefinition by removing the statutory exclusion of women from combat vessels. However, the 1994 Defense Authorization Act (P.L. 103-160, November 1993) did more than repeal 10 U.S.C. Section 6015.¹² The authorization bill included a provision requiring the Secretary of Defense to establish for any military occupational career field open to both men and women, common relevant performance standards, without differential standards or evaluation based on gender.¹³ For any military occupational specialty requiring muscular strength and endurance and cardiovascular capacity, the statute directed the Secretary to prescribe specific physical requirements on a gender-neutral basis. That never happened. On July 27, 1994, the Secretary of the Army, Mr. Togo West, proposed expanding the number of career fields open to women to 91.2 per cent of the career fields, 67.2 per cent of all Army positions.¹⁴ On July 29, 1994, the Secretary of Defense, Mr. William Perry, announced the opening of some 80,000

¹¹ *Ibid.*, 13.

¹² *1994 Defense Authorization Act* (P.L. 103-160, November 1993), section 541.

¹³ *Ibid.*, section 543.

¹⁴ Secretary of the Army, Mr. Togo West, Memorandum, "Increasing Opportunities for Women in the Army," (Washington, D.C., July 27, 1994), photocopied.

additional positions in all the Services to women. No effort was made to address the question of individual suitability and physical performance standards.

KASSEBAUM-BAKER AND CONGRESS 1997-1998

The next opportunity to review the issue of physical performance standards was occasioned by neither a military success nor a growing shortage of military manpower. Rather it was the result of investigations into incidents of sexual assault at Aberdeen Proving Ground and other training bases. On June 27, 1997, Secretary of Defense William S. Cohen announced the appointment of the Federal Advisory Committee on Gender-Integrated Training and Related Issues. Chaired by former Senator Nancy Kassebaum-Baker, the advisory committee was to assess the current training programs of the Army, Navy, Air Force and Marine Corps and to determine how best to train the gender-integrated, all-volunteer force. In December that year, much to Mr. Cohen's surprise and that of the Services, the Kassebaum-Baker commission recommended separating men and women in basic training and providing gender-specific training.¹⁵ The Congress responded quickly and created its own commission to investigate initial entry training, the Congressional Commission on Military Training and Gender-Related Issues. Therefore, what initially was an investigation into sexual misconduct expanded to address the manner in which military recruits were trained.

Although it was not a formal element of the commission's charter, the political purpose of the commission was to provide a way to avoid implementing the Kassebaum-Baker committee recommendations on gender-integrated training while endorsing the Services response to that committee's other recommendations. Most of the commission's recommendations were unanimous and endorsed the status quo. Only on the issue of gender-integration during Initial

¹⁵ REPORT OF THE FEDERAL ADVISORY COMMITTEE ON GENDER-INTEGRATED TRAINING AND RELATED ISSUES TO THE SECRETARY OF DEFENSE (Washington, D.C.: December 16, 1997).

Entry Training, IET, did the commissioner's split. Five commissioners, Dr. Charles Moskos, LtGen William M. Keys, Ret., Mr. Thomas Moore, and the Chairman, Anita Blair, did not concur on the recommendation to permit the Services to continue conducting basic training in accordance with current policy.¹⁶ Dr. Moskos abstained because he thought the wording of the recommendation implied there were no serious problems with IET. He noted that the trainers' comments indicated that there was something seriously flawed in gender-integrated training. In contrast, commissioners Blair, Moore, and Keys wrote, "that, not only is there evidence of serious problems in gender-integrated training, but there is also substantial evidence that gender-separate training produces superior results."¹⁷ They, however, felt frustrated in their efforts to obtain the data needed to assess the cause. They noted that during the work of the commission the Services indicated that their decisions on gender-integrated training were final and that they were not willing to change. Similarly, the Services were willing to provide information supporting their positions but less forthcoming with information adverse to their positions. These commissioners noted too that the Services provided only a few extremely limited comparative studies and most of the studies were of sociological or psychological issues.¹⁸ In summarizing their findings concerning gender-integrated training, they observed that the evidence needed to judge gender-integrated training was to a large extent missing. So much so, they found it necessary to task the Services to:

Assess, with respect to each service, the degree to which different standards have been established, or if not established are in fact being implemented, for males and females in basic training for matters such as physical fitness, physical performance (such as confidence and obstacle courses), military skills (such as marksmanship and hand-grenade qualifications), and nonphysical tasks required of individuals and, to the degree that differing standards are in fact being implemented, assess the effect of the use of those differing standards.

¹⁶ Congressional Commission on Military Training and Gender-Related Issues Final Report (Washington, DC, 1999), xlii.

¹⁷ *Ibid.*, 191.

¹⁸ *Ibid.*, 195

Despite the Services' position that the standards during IET were the same for men and women the commissioners were convinced that the trainers were informally making accommodations to ensure the training results met Service expectations.¹⁹ In effect, the dissenting commissioners were asking the Services to define the physical performance standards for occupational specialties open to men and women. Thirty-three years after *The Use of Women in the Military*, despite numerous commissions and federal statutory requirements, the military still has no body of objective evidence to demonstrate the effectiveness of their training programs. There are no objective measures to assess the impact of training in a gender-integrated versus a gender-separated environment or to assess whether women are prepared to serve in heavy physical occupations, much less the combat arms.

THE ONLY PERFORMANCE STANDARD

What physical performance evidence was available, the Services in general and the Army in particular have sought mightily to disparage. That data, of course, is the data collected on physical fitness tests. The Congressional Commission on Military Training and Gender-Related Issues made fitness tests a particular point in its findings, recommending the Services take steps to educate service members about the meaning of "physical fitness," and how it differs from job performance standards. The Commission observed that there was widespread misunderstanding about the purposes of the Services' physical fitness test. The tests are, in the words of the Commission, designed to measure physical health and well-being.²⁰ That may be the correct clinical interpretation of Service physical fitness testing, but it is certainly not how the tests are used and it is not consistent with the prodigious amount of Army research devoted to it. The

¹⁹ "It is obvious to any observer of basic training that there are differences in physical performance between men and women. De facto differences in performance (whether or not meeting standards) create the appearance of unequal, or unequally applied, standards. This suggests that standards or testing may be manipulated to permit lower-performing recruits to pass." Congressional Commission on Military Training and Gender-Related Issues Final Report, 198-199.

²⁰ *Ibid.*, xxxiv.

opinions of Army soldiers and officers diverge greatly from the official position but it not from a wanting in education. Their views come from experience with the application of the test. An experience that from enlistment through separation tells them that the Army Physical Fitness Test, APFT, is a test of physical fitness and the fitness implied is “physically fit to perform military tasks.” The APFT does not correlate directly with actual military tasks because the push-up and the sit-up does not measure strength related to tasks. Aerobic capacity measured by the 2-mile run is directly related to physical performance. Additionally, the APFT is important because, if you cannot pass the APFT, your ability to perform your military job does not matter. You will not be promoted; you will not attend Army schools; you will soon be separated.

However, the experience of 10 years of war has made clear that training to APFT standards is insufficient to meet the physical requirements of combat. In November 2010, the Army announced it was preparing a rigorous physical readiness test that addresses functional fitness.²¹ The list of events in what is now being called the Army Physical Readiness Test (APRT) has not yet been finalized. Under consideration is the inclusion of pull-ups as a test element. Additionally, it was thought that aerobic fitness could be assessed using a 1.5-mile run, but commanders have argued for retaining the 2-mile run. According to MGEN Richard Longo, who is responsible for developing the test, “While 1.5 miles measures all we need to measure about your cardiovascular fitness, that other half mile measures the other piece of the heart – the piece that keeps you going at 10,000 feet with an 80-pound ruck.”²² Although the new test is supposed to be gender neutral, that term only means that men and women will perform the same events. Men and women will be assigned scores based upon separate scoring tables. Setting performance standards for women on the pull-up event is especially challenging.²³

²¹ Lance M. Bacon, “A new way to measure fitness,” *Army Times*, November 8, 2010, 16.

²² Lance M. Bacon, “Tough PT Changes,” *Army Times*, September 12, 2011, 17.

²³ *Ibid.*, 16.

Despite the Army's need to develop a better test of physical readiness, the Army Physical Fitness Test does tell us a great deal about the physical potential of Army men and women and it, along with other physical data such as height, weight, and body fat content, does define the physical training potential of the military population. It may not be a worthy surrogate for actual physical performance measurements, but until now, the Army has consistently, avoided developing and applying those standards, so, by default, the APFT score is the standard. What the APFT scores means and how the scores are used is a subject that will be addressed later in this paper. Here, the issue is the Services' insistence, especially the Army's insistence, that the APFT is solely a measure of wellness, not to be confused with job performance. The Services in general, and the Army in particular, do not screen for physical potential before enlistment and they do not match recruits to occupational specialties based upon the physical requirements of the MOS. The Army MEPSCAT, proposed in 1982, would have done that but it was not implemented. The Army opened many physically demanding occupational specialties to women in 1994; e.g., 12C Bridge Crew Member, without establishing gender-neutral physical standards and without establishing a system to screen women recruits for the physical ability to meet the physical requirements of those jobs. Thus, the only measure of a soldier's ability to meet the strength and cardiovascular endurance requirements for advancement from IET, basic training, to advanced individual training, AIT, in the chosen MOS is the APFT. Failure to achieve a 50 on each event of the APFT and an overall score of 150 blocks the recruit's progress from basic training to advanced individual training and, thus, is currently the only assessment of a recruit's strength and endurance needed to be further trained. Introduction of the APRT and up-coming Army Combat Readiness Test will provide a much better assessment. However, until that happens, the APFT is *de facto* the test and the only means by which to assess readiness.

COMBAT REQUIREMENTS

The war against terrorism, in Afghanistan and in Iraq, has reminded the Army of its long-standing mission, sustained land operations. Unlike the swift operations in Panama in 1989 and Iraq in 1991, that were, in effect, sprints the war against terrorism is a long distance event that tests the Army's and the soldier's endurance. Gone too are front lines and a secure rear. The combat support and combat service support units are in amongst the combat units and the combat. For example, the 507th Maintenance Company's losses in the fight at Nasiriyah prompted the Army Training and Doctrine Command and the combat arms centers to work toward revising the basic combat training that combat service and support soldiers receive to ensure more are better prepared for the next fight.²⁴ Nevertheless, the fact that combat service and support units are found among combat units in the operational environment does not mean that women can meet the requirements of service in the ground combat arms because the physical requirements of direct combat, strength and endurance, are greater now than in past wars. In Afghanistan, the common Army infantryman's fighting load is 62 pounds while the average approach march load is 95 pounds. The squad leader, platoon sergeant, even the company executive officer carry similar loads. While an automatic rifleman's fighting load is 79 pounds and the machine gunner carries 120 pounds of equipment and ammunition in his approach march load.²⁵ The average fighting load is 35% of average man's bodyweight but half the bodyweight of an average army women.

The increased weight of the combat load combined with the high altitude in Afghanistan has placed a premium on strength and aerobic capacity and presents a significant challenge to

²⁴ Matthew Cox, "Warrior Spirit: Infantry chief wants to better train support soldiers for the battlefield," *Army Times*, September 22, 2003, 12.

²⁵ U.S. Army Center for Army Lessons Learned, *The Modern Warrior's Combat Load* (December 12, 2003), 112. A historical comparison of combat loads is contained in Joseph Knapik and Kay Reynolds, *Load Carriage in Military Operations: A Review of Historical, Physiological, Biomechanical, and Medical Aspects* (Washington, DC: Borden Institute, 2010).

sustaining performance in continuous operations. FM 22-9 notes at the outset that soldier resources shrink in combat. The combat losses are from not only battle deaths, wounds, non-combat injuries and illness, but also from physical fatigue and sleep loss, which undermines vigilance, slows decision-making, and impairs judgment and general performance.²⁶ Naturally, improving the strength and aerobic capacity of the infantryman is an important element in maintaining performance. So is the ability of the leader to balance the burdens to sustain the capacity of the entire unit. Tasks and load must be distributed within the squad and platoon to prevent overburdening anyone soldier and to ensure the entire unit shares the burden.²⁷ Tasks too must be rotated within the squad and platoon. Soldiers are cross-trained to permit task rotation and to provide relief for persons in critical positions. The leader cannot permit any one soldier to bare a task to exhaustion because that soldier's absence at a critical moment will undermine the performance and jeopardize the survival of the entire squad or platoon. Thus, within direct combat units, soldiers must strive to be not only individually fit but also all soldiers must be equally fit to sustain performance in combat operations. Thus, assessing the suitability of women for assignments in direct combat units requires an assessment of that population's strength and aerobic fitness and the ability to sustain readiness by avoiding injury. As the evidence will show, the physical capacity of women is significantly less than that of men and even more difficult to sustain.

MEASURING PHYSICAL CAPACITY

CPT Joseph Knapik explained in 1989, the elements of fitness tested by the APFT are related to military tasks. CPT Knapik observed:

In a military environment, absolute muscular endurance is important. Typical loads handled by soldiers include artillery shells, sand bags, crates, and weapons.

²⁶ FM 22-9 *Soldier Performance in Continuous Operations* (Washington, DC: HQ Department of the Army, 1983), 1-1 and 1-2..

²⁷ FM 22-9, 2-14.

The weight of these loads stays the same regardless of the individual soldier's strength. Stronger soldiers have a greater capacity for the high intensity, short duration efforts required to lift and carry these loads. Thus, for military purposes it is possible to combine the concepts of muscular strength and endurance since they are highly related on an absolute basis.²⁸

Similarly, the ability to run rapidly is highly correlated with aerobic capacity. An individual with higher aerobic capacity is able to perform sub-maximal physical tasks at a higher rate or for a longer time than an individual with lower aerobic capacity. Therefore, combining muscular strength and endurance with aerobic capacity provides a useful means to assess an individual's ability to perform the tasks normally encountered in the Army. The Army APFT does this by measuring relative strength using the push-up and aerobic capacity using the 2-mile run.

Because the push-up measures strength relative to a soldier's own body weight and weight distribution, the event is not by itself a measure of absolute strength. However, the 2-mile run is a good measure of aerobic capacity and using a formula found in Army FM 21-20 *PHYSICAL FITNESS TRAINING* the run-times for men and women can be converted into a measurement of aerobic capacity, VO_{2max} , measured in milliliters of oxygen per kilogram of weight per minute. Because the push-up is relative to the individual's physical stature, it combined with the aerobic capacity determined by the 2-mile run will not reveal the strength capacity of a population of men and women. However, COL Margarete DiBenedetto noted in her article, "Experience with a Pre-Basic Company at Fort Jackson," that height and weight are significant factors in the ability to use available strength and she developed an index of individual strength with which to compare groups of men and women reporting for basic training.²⁹ Thus, the APFT push-up and 2-mile run results and individual height-weight data when combined provide a good basis with which to compare the physical strength and endurance of men and

²⁸ Joseph Knapik, "The Army Physical Fitness Test (APFT): A Review of the Literature," *MILITARY MEDICINE* 154 (June 1989): 328.

²⁹ Magarete DiBenedetto, "Experience with a Pre-Basic Company at Fort Jackson," *MILITARY MEDICINE* 154 (May 1989): 259-263.

women. The more physically difficult the military task the more relevant the correlation. In other words, the comparison means very little if the task involved is light but it is significant for very heavy tasks.

Another major factor in individual physical performance is body fat content. Aerobic capacity is generally related to body fat while strength is related to fat-free mass. The higher the body fat content the lower aerobic capacity. Inversely, strength is related to total lean mass. A fat free mass of about 50 kilograms, 112 pounds, is the minimum needed to achieve a lift performance of 100 lbs.³⁰ James Vogel in his article "Obesity and Its Relationship to Physical Fitness in the Military," observed that that fact alone could be used to screen recruits for strength capacity at Military Entrance Processing Stations.³¹ That fact also can be used to compare the ability of men and women to perform the tasks of heavy and very heavy category military occupational specialties. In either case, women are at a significant disadvantage when it comes to performing military physical tasks because they have a significantly higher percentage of body fat and generally much lower total lean mass.

The article in *Military Medicine* reporting the results of the 1988 Active Army Physical Fitness Survey clearly stated the Army's position on physical fitness and combat readiness. It stated, "Physical Fitness has always been linked to combat readiness."³² The article went on to say that in 1981, a major Department of Defense study found that the military services could neither accurately measure the fitness of their members nor provide appropriate fitness programs for its occupational specialties.³³ In response, the Secretary of the Army designated 1982 as the Year of Physical Fitness; the Army surgeon general created a task force on physical fitness; and

³⁰ James A. Vogel, "Obesity and Its Relationship to Physical Fitness in the Military," *Armed Forces and Society* 18 (Summer 1992): 506.

³¹ Vogel, 507.

³² John S. O'Conner, Michael S. Bahrke, and Robert G. Tetu. "1988 Active Army Physical Fitness Survey." *MILITARY MEDICINE* 155 (December 1990), 579.

³³ *Ibid.*

the Army Physical Fitness Research Institute and the US Army Physical Fitness School were created. All those efforts were intended to improve the soldier's physical fitness, thereby increasing his ability to succeed on the modern battlefield. Army fitness research cited in this paper is the direct consequence of the Army's concern for establishing the link between its fitness program and physical combat readiness. It is curious, given all the resources the Army has devoted to research and to development of the APFT standards that the Army now declares that the interest in the APFT was solely in wellness. That research has established the relevance of the APFT to military performance and its relationship to Army fitness standards. What the Army has failed to do is identify the physical performance standards for its military occupational specialties.

ARMY ROTC APFT PERFORMANCE

Male and female cadets participate in a summer training event usually a school year prior to commissioning. The training camp was previously called the Army ROTC Advance Camp but is now labeled the Leader Development Assessment Course. Cadets who take the APFT during training camp not a novice military population. They have been trained for at least a year, some for three years, and many have prior military experience in addition to their cadet training. They also come to summer camp knowing that their summer camp performance will have a major impact on their branch and service choices. Consequently, as a population they are experienced and motivated and the results of their tests show that they are physically more capable than any other Army population, with the possible exception of West Point (see the table below).

Table 1 A Comparison of APFT Performance

	Push-Ups	2-Mile Run	
Population	Mean Number	Mean Time (Minutes)	Source
ROTC Men 1992	69	13:03	Gregor, 1992
Army Men (17-21) 1988	53	14:43	Knapik, 1989

Army Men (22-26) 1988	53	14:49	Knapik, 1989
Basic Training Men 1989	45	14:31	DiBenedetto, 1989
ROTC Women 1992	42	15:46	Gregor, 1992
Fitness Company Men	29	15:18	DiBenedetto, 1989
Army Women (22-26)	31	17:43	Knapik, 1989
Army Women (17-21)	30	18:12	Knapik, 1989
Basic Training Women	22	17:24	DiBenedetto, 1989
Fitness Company Women	19	17:44	DiBenedetto, 1989

The APFT performance of ROTC cadets at summer camp provides a sound base upon which to assess the potential for physical training. The observations presented in this paper are made from analyzing the APFT scores from 1992 through 2008, 74,838 records. The results reveal the reasonable limits of any successful training program and the physical limits of the populations.

The ROTC cadet APFT records do not provide a basis for assessing the injury rates of the men and women. However, there are a considerable number of studies of initial entry training from which to assess the relative risk of injury. It is possible that a longer and more gradual period of physical training will reduce the incidence of IET training injuries. Recruits entering IET are not as fit or as strong as the population of ROTC cadets. However, there are a variety of studies relating aerobic fitness and body constitution to both recruit and active duty injury rates. Thus, the data about ROTC cadets can be used successfully to determine whether training can produce in the population of women the physical attributes that are required of men in the combat arms. The other study provides information with which to assess whether successfully trained women can survive the physical demands of close ground combat.

It is necessary to make one additional observation about the relevance of ROTC cadet physical performance data to IET. A new APFT scoring table was introduced in 1999. That table was developed by surveying the general army population and set the maximum score in each

APFT event at the 90th percentile of each age group. Thus, only 10 per cent of men, aged 17-21 in the army at large, are expected to achieve a 13:00 minute time on the 2-Mile Run; 10 per cent of the women in that age group are expected to attain 15:36. In contrast, 30 per cent of the cadet men and women bested those scores prior to 1999. The same is true of Push-Ups. About 76 per cent of the women and 80 per cent of the men exceeded the Army's established maximum score. Consequently, cadets tested after 1998 had less incentive to train to their full potential because the table scores for the events had been lowered. Cadet Command is accustomed to adjusting the relative importance of the APFT in the cadet's camp evaluation and it is extremely difficult to assess the effect those changes had on individual incentive. It is also likely that in the context of the wars in Iraq and Afghanistan cadets seeking active duty were keenly aware of the importance of physical fitness and worked to achieve their full potential. Whatever the impact of the scoring changes on individual incentive, the physical disparity between men and women remained great.

The data on ROTC cadet APFT performance have been collected over the course of 17 years, 1992-2009. The data set obtained from Army Cadet Command has rarely permitted linking height-weight data to individual performance on any event. However, the data provided in 2000 enabled analysis of the relationship between body constitution and aerobic capacity and that data is presented here. Additionally, much of the data was collected year by year and often in a format that differed from the previous year. In 2009, however, Cadet Command provided a decade of data, 1999-2008, in three separate EXCEL workbooks. The inferences drawn here are the product of analyzing 41,599 records from 1999-2008 and 33,239 records obtained previously, 74,838 records. Those records contain the population of men and women commissioned by the U.S. Army through ROTC throughout that period.

The charts below shows the distribution of cadet scores on the 2-Mile Run in 2000, the Push-Up, and the distribution of cadets by weight. The difference in performance is clear. Only 2.9 per cent of the women, 23, were able to attain the male mean score. The strength comparison

is somewhat worse, 1.5 per cent of the women achieved the male mean. Given the difference in stature between the cadet men and women, the difference in absolute strength is very large.

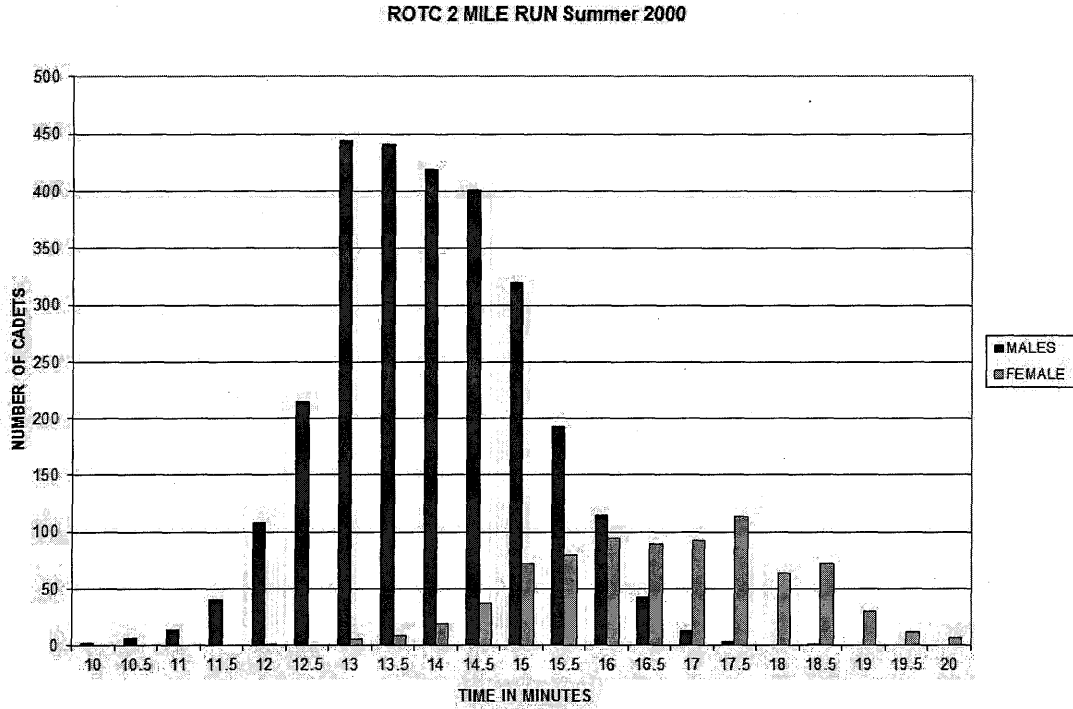


Figure 1 Cadet 2 Mile Run Results FY2000

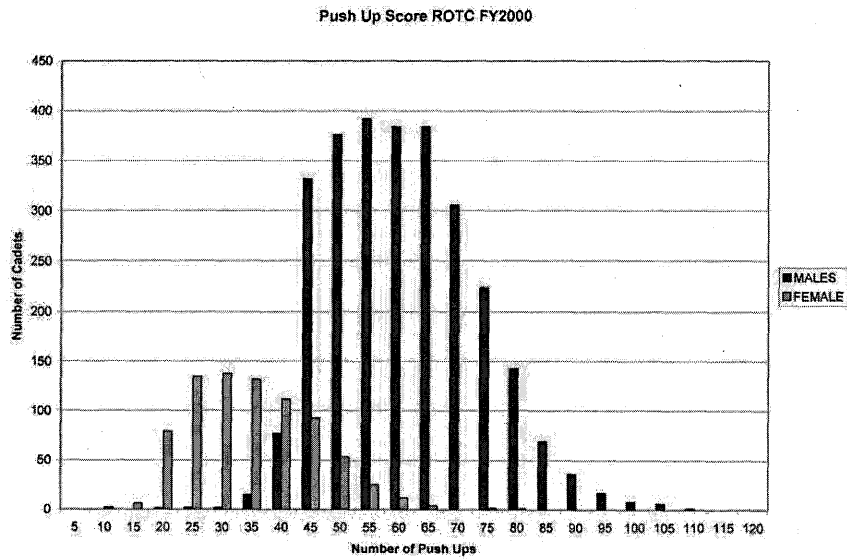


Figure 2 Cadet Push Up Results FY2000

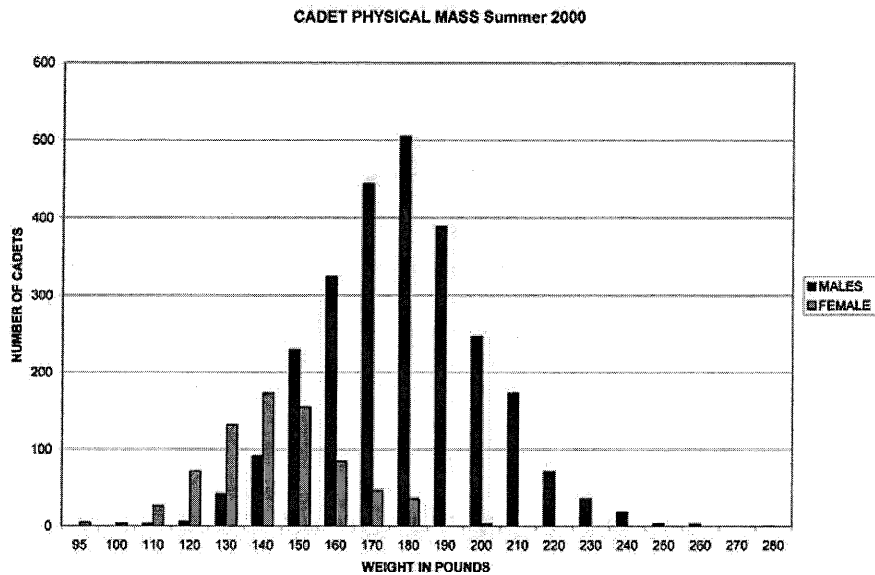


Figure 3 Cadet Mass FY2000

THE PERFORMANCE GAP

The performance gap represents a major obstacle to deriving benefit from common collective training. Consider the simple problem of different stature when it comes to setting the standard for load for training. A woman who is 5 feet 8 inches tall is in the 95th percentile of Army women in stature. Army Regulation 600-9 sets her allowable maximum weight at 150 pounds. The median stature of an Army male is 5 feet 9 inches; his allowable maximum weight is 175 pounds. FM 21-20, Physical Fitness Training suggests march training should begin with a load equal to 20 per cent of the body weight, in this case 30-35 pounds. (Thirty pounds was formerly the weight of a soldier's fighting load; the weight of his battledress uniform, helmet, load bearing equipment, protective mask, canteen, bayonet, and M16 rifle). Applying that standard, 95 per cent of the women to be trained will be over loaded. Now attempt to set the pace of the march given the differences in the level of conditioning. Immediately, the answer seems to be the same as that has been applied to IET physical fitness training: divide the unit into ability groups and set different loads for different populations. However, that is not what it means to train together and as the data from Army ROTC shows, as the men and women improve their

training needs will diverge even more. In contrast, keeping the men and women together can only diminish the training benefit received by men because the load or the march rate or both must be kept within the range of strength and endurance of the women.

The British experience training men and women together supports this conclusion. In 1998, the British Army introduced a job-related gender free selection process, Physical Selections Standards (Recruits) (PSS(R)). In 2002, Ian M.M. Gemmell published a study that observed that after introducing the (PSS(R)), the incidence of injuries among women during recruit training went up dramatically. Therefore, a study of gender differences in the physical demands of recruit training was undertaken. The study observed that male recruits in the mixed gender platoons experienced lower cardiovascular strain compared to their female counter parts.³⁴ However, cardiovascular strain is not only associated with injury but also with performance improvement. Training men with women reduced the benefit from the training. The chart below makes clear why this is the case for all men.

³⁴ Sam D. Blacker, David M. Wilkinson, and Mark P. Rayson, "Gender Differences in the Physical Demands of British Army Recruit Training," *Military Medicine*, 174 (2009), 816.

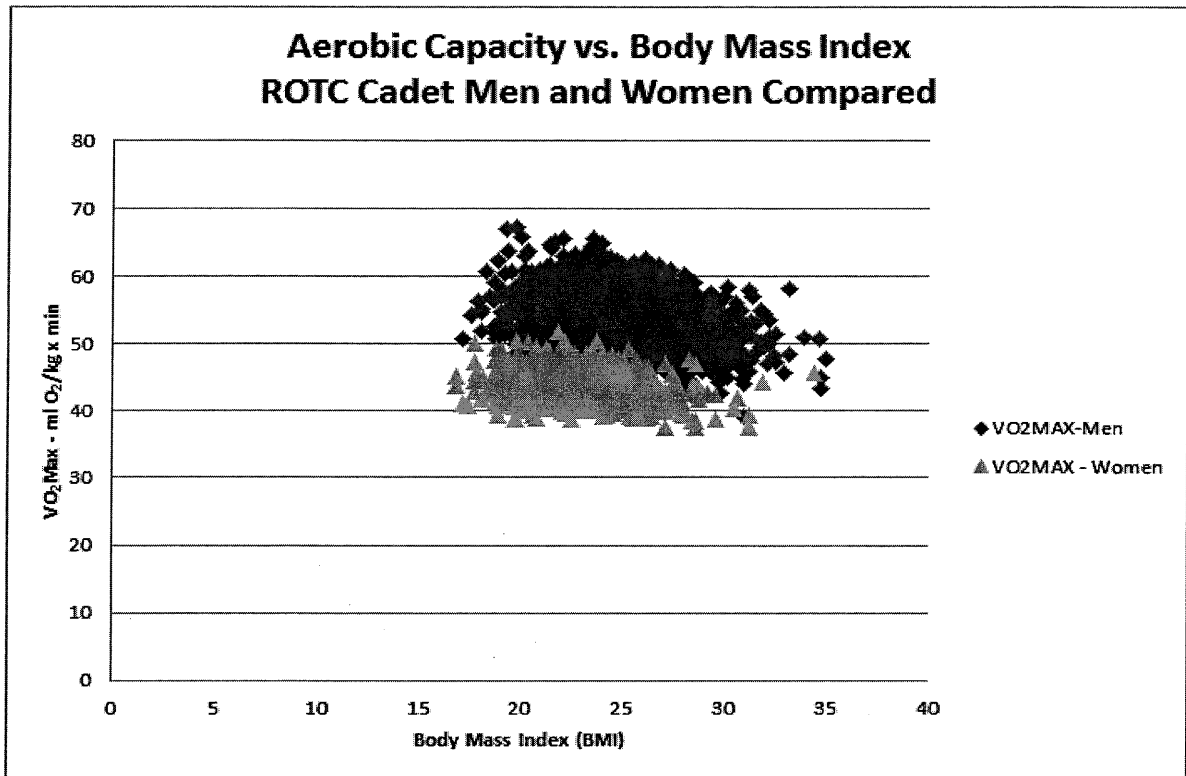


Figure 4 -Aerobic Capacity vs. BMI

If aerobic capacity correlates with body fat percentage and women were similar to men then it could be inferred that women with a lower body mass index would achieve greater aerobic efficiency and could be expected to compete with men. Figure 4 makes clear that that is not the case. Women at all body mass levels fail to achieve aerobic capacities that make them competitive with men. The chart indicates that the aerobic capacity achieved by women regardless of their body composition is less than the capacity of men. As the next chart will indicate there are a few, exceptional women who best the bottom 16% of men, but these rare women are four standard deviations above the female mean, fewer than 1 in a 1000. In this exceptionally fit ROTC Cadet population, considering 74,838 records, not one women achieved the male mean.

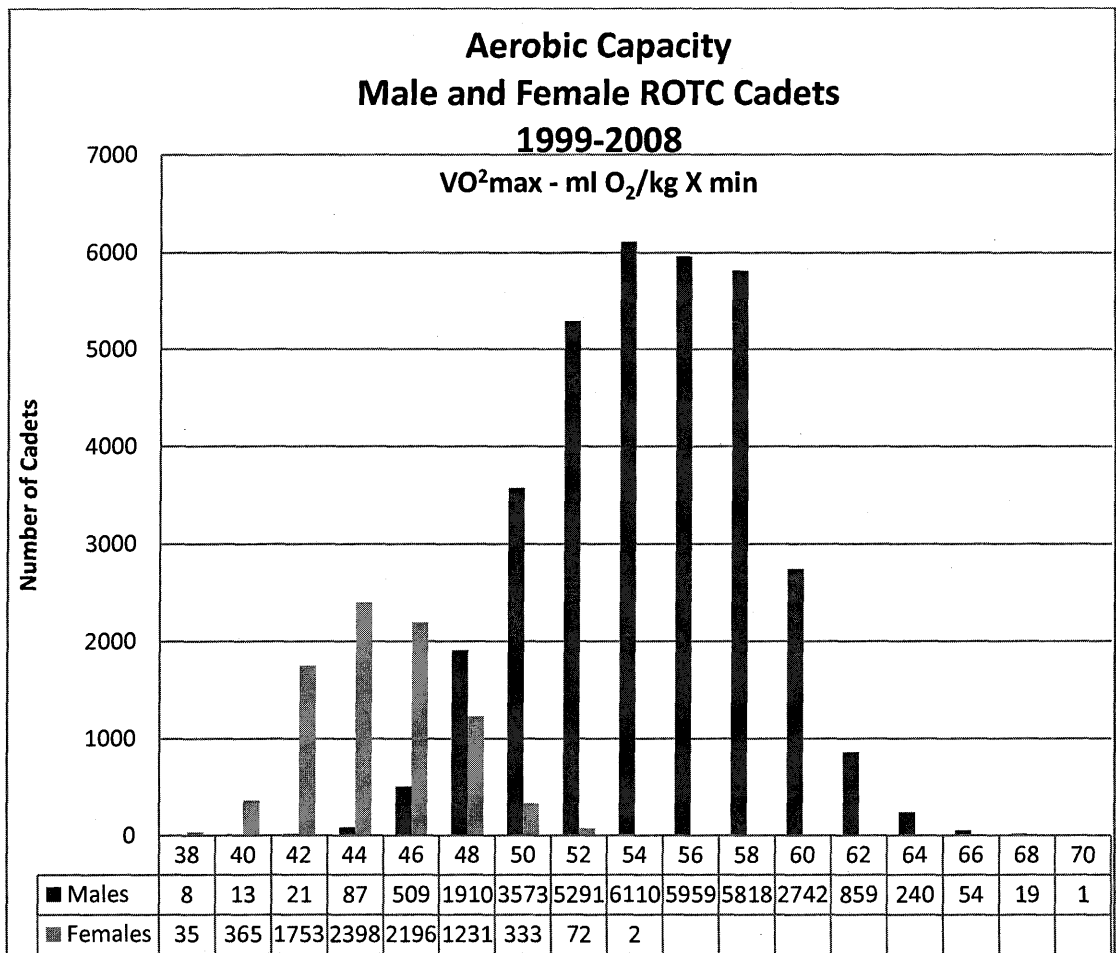


Figure 5: Aerobic Capacity 1999-2008

The average aerobic capacity for cadet males from 1999-2008 was 53.60 ml O₂/kg x min, the female mean was 43.75. The standard deviation for women was 2.5. Thus, the average male aerobic capacity was four standard deviations above the female mean. The best female achieved an aerobic efficiency of 53.1 ml O₂/kg x min, slightly less than the male mean. The standard deviation for men was 3.75, thus a VO₂max of 50 is one standard deviation below the mean. The histogram displayed in Figure 5 groups scores from the nominal number below, so the bar shown as 50 shows all those who scored above 48 up to and including 50. Thus, only 74 women, 74 of 8385 women, achieved an aerobic capacity higher than the lowest 15.8% of men. No training system can close this gap.

The reason men and women cannot truly be trained together is not a matter of attitude; it is physical. The difference in male and female body composition and the components of strength and endurance training are firm obstacles to designing mutually beneficial training events. Aerobic efficiency is related to the body's fat content. Strength is related to total fat-free body mass. Because women have a greater portion of their body composition in fat and do not add muscle mass as readily as men do, training men causes them to improve in absolute terms relative to women. The Natick study designed specifically to improve the load carriage ability of women lasted 24 weeks. At the conclusion of the study, the remaining 32 women had improved their box-weight lift capacity to 81 per cent of the Army male value. However, they added only .9 kilograms of muscle mass and achieved an average level of aerobic fitness below the minimum requirement for Army men.³⁵ The Army officially recognized that strength is related to total fat-free mass when in 1991, it increased the allowable body fat percentage for women from 28 to 30 percent to permit enlistment of larger, albeit aerobically less fit, women. The Army kept the allowable male percentage at 20.³⁶ Can it then follow that these new soldiers can be trained together? Does the evidence provide any support for including women in the ground combat arms, not really?

Numerous newspapers reported that the Natick study results demonstrated that woman could approach the strength of men. However, the more significant finding was that Army Basic Training did not prepare women for heavy lifting. The study recommended an experimental program during AIT to improve lifting capacity so that women might develop the capacity to serve in "very heavy" MOS's.³⁷ To graduate from the basic training men and women must attain

³⁵ U.S. Army Research Institute of Environmental Medicine. *Effects of a Specifically Designed Physical Conditioning Program on the Load Carriage and Lifting Performance of Female Soldiers*, By E. Harman, Peter Frykman, Christopher Palmer, Eric Lammi, Katy Reynolds, and Verne Backus. Technical Report no. T98-1 (Natick, Mass., November 1997) 48, 74

³⁶ HQDA Msg 2120002 August 1991, New Accession Weight and Body Fat Standards.

³⁷E. Harman and others, 78.

50 points in each APFT event. This level of fitness presumably prepares women to enter very heavy occupational specialties as well as men. Yet, the difference between the gender-normed standards and actual performance is quite large. Fifty points on the APFT push-up event represents 32 push-ups for a man but only 13 for a woman. Thirteen push-ups is the IET entrance test requirement for men. If a man does fewer than 13, he is diverted from basic training to the Fitness Training Company to prepare for basic training. It is clear that now that many of the very heavy occupational specialties are open to women, some women have special training needs and basic training does not prepare them for “very heavy” MOS’s. The practice of mixed-gender basic training also does not prepare men fully. If women leave recruit training unprepared for the heavy occupational specialties to which they have been assigned, how is it possible for them to achieve the level of physical performance required by the ground combat arms? It is not.

In this regard, the British experience is insightful. Prior to 1998 female recruits in the British army were trained to a lower standard of physical fitness than men as is the case in U.S. Army IET today. The system was described as “gender-fair.” It was noted, however, that many of those female recruits were subsequently found to be physically incapable of performing the job to which they had been selected. Thus, in April 1998, the British Army introduced a standard set of physical test scores in relation to the career field. It was recognized the applying the “gender-free” standards would reduce the number of women in the more physically demanding career fields. Not only did the number of eligible women decline but also the injury and separation rate soared. The overuse injury rate among women rose from 4.6% to 11.1% while the rate among men was unchanged at 1.5%.³⁸ The study’s author, Ian M.M. Gemmell concluded that women face an excess risk when they undertake the same arduous training as male recruits and such

³⁸ Ian M M Gemmell, “Injuries among female army recruits: a conflict of legislation,” *Journal of the Royal Army Medical Corps* 148 (2002), 24.

training places concerns for health and safety in conflict with equal opportunity goals.³⁹ The British Army formally reviews its policy excluding women from ground combat positions every eight years. The last review was in 2010. The positions remain closed.

The medical literature and the physiological studies of marching, load bearing, and physically demanding military jobs makes clear that women are more prone to injury and separation in recruit training and in physically demanding duty positions. The rate of injury varies from study to study but greatly exceeds the male injury rate. It is not possible here to review all the reports. Many of these studies are listed in the bibliography. One study worth of note is systematic review of studies of stress fracture by Laurel Wentz, et. al.⁴⁰ That study compiles a variety of studies involving number of foreign militaries and athletic teams. The study reveals that women have a greater incidence of stress fracture and that women have anatomical disadvantages that increase the risk of stress fractures. The problem is not one of training.

Women are not only more prone to injury in training. They are also more likely to suffer non-battle injury during active military operations. In a study of disease and non-battle injuries in a brigade combat team (BCT) in Iraq the injury rate among women was twice that for men. The disease and non-battle casualty rate among the 325 women in the BCT was 408.6 per 1,000 Combat-Years versus 244 for men.⁴¹ Half of the women were evacuated for musculoskeletal injuries.⁴² Thus, it is fair to conclude that within ground combat positions the injury rate would be higher even if battle wounds were included.

³⁹ Ibid., 23.

⁴⁰ Laurel Wentz , MS, RD ; Pei-Yang Liu , PhD, RD ; Emily Haymes , PhD ; Jasminka Z. Ilich , PhD, RD, "Females Have a Greater Incidence of Stress Fractures Than Males in Both Military and Athletic Populations: A Systemic Review," *MILITARY MEDICINE*, 176, 4:420, 2011

⁴¹ LTC Philip J. Belmont Jr., MC USA; CPT Gens P. Goodman, MC USA; CPT Brian Waterman, MC USA; LTC Kent DeZee, MC USA; COL Rob Burks, QM USA; MAJ Brett D. Owens, MC USA, "Disease and Nonbattle Injuries Sustained by a U.S. Army Brigade Combat Team During Operation Iraqi Freedom," *MILITARY MEDICINE*, 175(2010), 471.

⁴² Ibid., 469.

MILITARY PERFORMANCE

Given the evidence presented thus far about strength, aerobic efficiency and injury what can be said about the physical performance of individual women in heavy physical occupations during the wars in Afghanistan and Iraq? The answer is, almost nothing, which might also be said about men in those units. Data on unit performance is not a surrogate for data on individual performance. The reason this is true should be obvious from the earlier discussion about sustaining performance during continuous operations. Military leaders will shift tasks not only to maintain individual performance but also to sustain organizational performance. They will not assign tasks to individuals who are unable to perform them, especially in the context of direct ground combat. This is true in all male combat units as well as combat and combat service support units. Assigning women to direct combat units differs from the current situation because the evidence indicates that women cannot be trained to achieve the average aerobic capacity of their male comrades and they do not have the stature to bear the heavy combat loads. An infantry squad leader would be foolish to assign the machine gun to a woman knowing that she would not only tire more quickly but also risk an incapacitating injury. Men in the formation attempting to lighten a woman's load would not be chivalrous; they would be responding to the need to preserve her capacity to participate in the unit's mission. Unfortunately, shifting her load to other squad members would serve to increase the load of the already burdened squad. The effect of task shifting on the overall performance and cohesion of the squad is a subject for another paper. How long a woman in an infantry squad might "ruck-up" and gut it out is a matter for speculation. The data presented in this paper supports the conclusion that whatever the performance of the squad of as a whole, the individual performance of women would be measured by the injury statistics.

REMEMBERING MILITARY READINESS

The discussion in this paper opened by noting that in 1981 when "Women, Combat, and the Draft" was published little was known about the physical requirements of the military

occupational specialties open to women and the ability of women to meet them. The history of the expanding role of women in the Army shows little concern for physical performance and individual combat readiness. The Army has officially argued that it does not know the relationship between performance on the Army Physical Fitness Test and job performance but it has invested much of its research in the APFT and its relationship to training success. The Army has, by its own admission, not developed job performance measures for the military occupational specialties open to both men and women and it does not screen recruits to determine that they possess the requisite strength for their chosen MOS. Yet, the Army's own research indicates that the vast majority of women do not possess the lean mass necessary to achieve the strength requirements for very heavy and heavy physical MOS's. Thus, by default the Army relies on IET and basic training in particular to produce the strength needed to enable training recruits in their MOS and has set "gender-fair" standards for the APFT, which until now have been the only official measurement of strength and aerobic capacity. Consequently, the Army does not know whether female IET graduates are physically ready to perform their duties, let alone ready for the rigors of direct ground combat.

The data presented here does not answer directly the question of whether women are suitable for ground combat positions. The data presented in this paper demonstrates that the Services, especially the Army, have expanded the occupational specialties open to women purely as a part of the social concern for equality and have only paid lip service to combat readiness. The data clearly reveals a very large gap between the physical strength, aerobic capacity and size of Army men and women. Training men and women correctly improves the performance of both groups but it also widens the gap in performance. It is unlikely a woman assigned to a heavy physical occupation possesses the required strength and endurance for those duties, regardless of her APFT score. That said, the data show that absent a gender-free assessment of physical requirements for an occupational specialty, training soldiers to meet actual job and combat requirements has been left up to the individual trainer or the soldier's supervisor. Those trainers

and supervisors, faced with observable differences in physical capacity, informally redistribute the tasks, in effect, establishing personal MOS requirements. What the Army does not admit but which the large body of ROTC data and Army research show, is that the Army could know precisely the training needs of women in the heavy and very heavy occupational specialties currently open to women. Even though assigning women to direct ground combat occupations is clearly infeasible, establishing occupational physical performance standards and testing the physical potential of incoming female recruits would greatly improve the selection of a suitable military occupational specialty.

In light of the evidence collected over the past thirty years concerning the physical performance of men and women in the military, the observations of the Military Leadership Diversity Commission seem bizarre. The commission observed that the discrepancy in the proportion of women among general officers was in part the result of low initial officer accessions and lower retention during the mid-career.⁴³ Establishing a gender-free training system to accommodate the assignment of women to ground combat occupations would reduce the number of women assessed and would increase mid-career attrition. Few if any women would be accessed into the ground combat arms and perhaps none would survive the twenty-five years before being considered for general officer. Yet, those who are committed to making the Army look right will never be swayed by evidence that putting women in the ground combat arms will not work right.

Thirty years ago, “Women, Combat, and the Draft” closed with the observation that the inclusion of women in the military had changed what the very idea of what a soldier was, that no

⁴³ Military Leadership Diversity Commission, *FROM REPRESENTATION TO INCLUSION: Diversity Leadership for the 21st Century Military*, Executive Summary (Arlington, VA: March 15, 2011), 11.

one yet knew the new definition, and that no one was willing to impose one.⁴⁴ The situation is no different today. Why can't anything be done?

⁴⁴ Gregor, 58.

BIBLIOGRAPHY

- Baldi, Karen A. "An Overview of Physical Fitness of Female Cadets at the Military Academies." *MILITARY MEDICINE* 156 (October 1991): 537-539.
- Belmont, Philip J., Jr., LTC MC USA; CPT Gens P. Goodman, MC USA; CPT Brian Waterman, MC USA; LTC Kent DeZee, MC USA; COL Rob Burks, QM USA; MAJ Brett D. Owens, MC USA. "Disease and Nonbattle Injuries Sustained by a U.S. Army Brigade Combat Team During Operation Iraqi Freedom." *MILITARY MEDICINE* 175, 7 (July 2010), 469-476.
- Bielenda, Christine C., Joseph Knapik, and David A Wright. "Physical Fitness and Cardiovascular Disease Risk Factors of Female Senior U.S. Military Officers and Federal Employees." *MILITARY MEDICINE* 158, 3 (March 1993): 177-181.
- Blacker, Sam D., David M. Wilkinson, and Mark P. Rayson. "Gender Differences in the Physical Demands of British Army Recruit Training." *MILITARY MEDICINE* 174, 8 (August 2009). 811-816.
- Buddin, Richard. "Weight Problems and Attrition of High Quality Military Recruits." RAND Note N-2847-FMP (June 1989).
- Carreiras, Helena and Gerhard Kuemmel, ed. *Women in the Military and in Armed Conflict. Wiesbaden, FRG: Verlag fuer Sozialwissenschaften, 2008.*
- Crawford, Kim, PhD; Katelyn Fleishman, MS; John P. Abt, PhD; Timothy C. Sell, PhD; Mita Lovalekar, PhD; Takashi Nagai, MS; Jennifer Deluzio, MS; LTC Russell S. Rowe, MD; LTC Mark A. McGrail, MD; Scott M. Lephart, PhD. "Less Body Fat Improves Physical and Physiological Performance in Army Soldiers." *MILITARY MEDICINE* 176, 1 (January 2011). 35-43.
- Department of the Army. FM 21-20 *PHYSICAL FITNESS TRAINING* (20 September 1992).
- Department of the Army. *WOMEN IN THE ARMY POLICY REVIEW*. Washington, D.C., 12 November 1982.
- DiBenedetto, Magarete. "Experience with a Pre-Basic Company at Fort Jackson." *MILITARY MEDICINE* 154 (May 1989): 259-263.
- Febbraro, Angela R. *Women, leadership and gender integration in the Canadian combat arms: A qualitative study*. Toronto: Defense R&D Canada, December 2003.
- Foulkes, Guy D. "Orthopedic Casualties in an Activated National Guard Mechanized Infantry Brigade during Operation Desert Shield." *MILITARY MEDICINE* 160 (March 1995): 128-131.
- Friedl, Karl E., Jennifer A. Nuovo, Tray H. Patience, and Joseph R. Dettori. "Factors Associated with Stress Fracture in Young Army Women: Indications for Further Research." *MILITARY MEDICINE* 157, 7 (July 1992): 334-338.
- Gebicke, Mark E. GAO, "Military Attrition: DOD Needs to Better Analyze Reasons for Separation and Improve Recruiting Systems," Testimony Before the Subcommittee on Military Personnel, Committee on National Security, House of Representatives, March 12, 1998.
- Gemmell, Ian M. M. "Injuries among female army recruits: a conflict of legislation" *Journal of the Royal Army Medical Corps* 148 (2002), 23-27.

- Goldberg, Matthew S. "Death and Injury Rates of U.S. Military Personnel in Iraq." *MILITARY MEDICINE* 175, 4 (April 2010). 220-226.
- Gordon, Claire C., Thomas Churchill, Charles E. Clauser, Bruce Bradtmiller, John T. McConville, Ilse Tebbetts, and Robert A Walker. *1988 Anthropometric Survey of U.S. Army Personnel: Summary Statistics Interim Report*. U.S. Army Research Institute of Environmental Medicine (March 1989) Technical Report TR89/027.
- Gregor, William J. "Women, Combat and the Draft: Placing Details in Context." in Taylor, William J., Jr. and Schrader, Richard A. ed. *Defense Manpower Planning: Issues for the 1980's* (Elmsford, NY: Pergamon Press, Inc., 1981).
- Harman, E., Peter Frykman, Christopher Palmer, Eric Lammi, Katy Reynolds, and Verne Backus. *Effects of a Specifically Designed Physical Conditioning Program on the Load Carriage and Lifting Performance of Female Soldiers*. U.S. Army Research Institute of Environmental Medicine (November 1997) Technical Report T98-1.
- Headquarters, Department of the Army. Army Regulation 600-9 *The Army Weight Control Program*, (Washington, D.C., 1 September 1986).
- Headquarters, Department of the Army. Interim Change No. 101 to *The Army Weight Control Program*, Army Regulation 600-9. 15 November 1993.
- Hollander, Ilyssa E. MPH and Nicole S. Bell, SeD, MPH. "Physically Demanding Jobs and Occupational Injury and Disability in the U.S. Army." *MILITARY MEDICINE* 175, 6 (October 2010). 705-712.
- Knapik, Joseph and Kay Reynolds. *Load Carriage in Military Operations: A Review of Historical, Physiological, Biomechanical, and Medical Aspects* (Washington, DC: Borden Institute, 2010).
- Knapik, Joseph. "The Army Physical Fitness Test (APFT): A Review of the Literature." *MILITARY MEDICINE* 154 (June 1989): 326-333.
- . "Discharges during U.S. Army Basic Training: Injury Rates and Risk Factors." *MILITARY MEDICINE* 166, 7 (July 2001). 641-647.
- Lester, Mark E., CPT; Joseph J. Knapik; SGT Daniel Catrambone; Amanda Antczak; Marilyn A. Sharp; MAJ Lolita Burrell; Salima Darakjy. "Effect of a 13-Month Deployment to Iraq on Physical Fitness and Body Composition." *MILITARY MEDICINE* 175,6 (June 2010). 417-423.
- Military Leadership Diversity Commission. *FROM REPRESENTATION TO INCLUSION: Diversity Leadership for the 21st Century Military*. Executive Summary, Arlington, VA: March 15, 2011.
- Military Leadership Diversity Commission. *FROM REPRESENTATION TO INCLUSION: Diversity Leadership for the 21st Century Military*. Arlington, VA: March 15, 2011.
- Ministry of Defense, United Kingdom. *REPORT ON THE REVIEW OF THE EXCLUSION OF WOMEN FROM GROUND CLOSE-COMBAT ROLES* (November 2010).
- O'Conner, John S., Michael S. Bahrke, and Robert G. Tetu. "1988 Active Army Physical Fitness Survey." *MILITARY MEDICINE* 155 (December 1990): 579-584.
- Patterson, Harlan S., Thomas W. Woolley, and Wayne M. Lednar. "Foot Blister Risk Factors in an ROTC Summer Camp Population." *MILITARY MEDICINE* 159 (February 1994): 130-135.

- Protzman, Robert R. "Physiologic Performance of Women Compared to Men: Observations of Cadets at the United States Military Academy." *The Journal of American Sports Medicine* 7 no.3 (1979): 191-194.
- Reinker, Kent A. and Susan Ozburne, "A Comparison of Male and Female Orthopaedic Pathology in Basic Training." *MILITARY MEDICINE* 144 (August 1979): 532-536.
- Reynolds, Katy, COL, MC USA (Ret.); Ludmila Cosio-Lima, PhD; LTC Maria Bovill, SP USA; William Tharion, MS; Jeff Williams, BS; Tabitha Hodges, MS. "A Comparison of Injuries, Limited-Duty Days, and Injury Risk Factors in Infantry, Artillery, Construction Engineers, and Special Forces Soldiers." *MILITARY MEDICINE* 174, 7 (July 2009). 702-708.
- Ross, James. "A Review of Lower Limb Overuse Injuries during Basic Training. Part 1: Types of Overuse Injuries." *MILITARY MEDICINE* 158 (June 1993): 410-415.
- Simpson, Michael B., Christopher Young, and Allan L. Burkett. "The Incidence of Permanent Upper and Lower Extremity Profiles in Active Duty Army Officers." *MILITARY MEDICINE* 157 (January 1992): 17-21.
- Snoddy, Robert O'Neil, Jr. and John M. Henderson. "Predictors of Basic Infantry Training Success." *MILITARY MEDICINE* 159 (September 1994): 616-622.
- The Presidential Commission on the Assignment of Women in the Armed Forces*, by GEN Robert T. Herres, Ret., Chairman. Washington, D.C., November 15, 1992.
- Tomasi, Louis F. "The New 1998 APFT Standards." *SOLDIERS* (February 1998): 5-8.
- Vogel, James A. "Obesity and Its Relationship to Physical Fitness in the Military." *Armed Forces and Society* 18 (Summer 1992): 497-513.
- US Army Research Institute for the Behavioral and Social Sciences *The 1995 Gender Integration of Basic Combat Training Study*. By Jacqueline A. Mottern, David A. Foster, and Elizabeth J. Brady. Study Report No. 97-01 (February 1997)
- U.S. Army Research Institute of Environmental Medicine. *Effects of a Specifically Designed Physical Conditioning Program on the Load Carriage and Lifting Performance of Female Soldiers*, By E. Harman, Peter Frykman, Christopher Palmer, Eric Lammi, Katy Reynolds, and Verne Backus. Technical Report no. T98-1, Natick, Mass., November 1997.
- Wentz, Laurel, MS, RD; Pei-Yang Liu, PhD, RD; Emily Haymes, PhD; Jasminka Z. Ilich, PhD, RD. "Females Have a Greater Incidence of Stress Fractures Than Males in Both Military and Athletic Populations: A Systemic Review." *MILITARY MEDICINE*, 176, (2011) 4:420,

Session 5 - Technology, Ethics and War

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

Enhanced Warfighters: Risk, Ethics, and Policy

Prepared for: The Greenwall Foundation

Prepared by: Patrick Lin, PhD
Maxwell J. Mehlman, JD
Keith Abney, ABD

California Polytechnic State University,
San Luis Obispo
College of Liberal Arts
Philosophy Department
Ethics + Emerging Sciences Group

Case Western Reserve University
School of Law
School of Medicine
The Law-Medicine Center

Prepared on: January 1, 2013

Version: 1.0.2

Index

Executive summary	iii
Disclosures	iv
1. Introduction	1
1.1. Purpose	2
1.2. Background	3
1.3. Questions	8
2. What is human enhancement?	11
2.1. Controversies	12
2.2. Working definition	17
2.3. Variables	18
2.4. Technology survey	21
3. Law and policy	28
3.1. International humanitarian law	28
3.2. US domestic law	36
3.3. Operations	38
4. Bioethics	43
4.1. Research model	44
4.2. Medical model	50
4.3. Public-health model	54
5. Risk Assessment	57
5.1. Risk-benefit model	57
5.2. Risk factors	61
6. A hybrid framework	66
6.1. Legitimate military purpose	66
6.2. Necessity	67
6.3. Benefits outweigh risks	67

6.4. The warfighter’s dignity is maintained	71
6.5. Burdens are minimized	71
6.6. Consent	72
6.7. Transparency	75
6.8. Fair distribution of risks and benefits	75
6.9. Superiors are accountable	76
7. Other considerations	77
7.1. Character and virtues	77
7.2. Emotions and honor	80
7.3. Broader impacts	84
7.4. Conclusion	86
8. References	88
Appendix 1: List of acronyms	105
Appendix 2: Authors	106
Appendix 3: Contact information	108

Executive Summary

The United States military is making substantial investments to develop technologies that would enhance the ability of warfighters to complete their missions safely and effectively. Driven by neuroscience, biotechnology, nanotechnology, robotics, and other emerging technologies, this research includes combating sleep deprivation, improving cognitive performance, increasing strength, reducing muscle fatigue, and other enhancements to the human body and mind.

As with other emerging military technologies, such as robotics and cyber-capabilities, human enhancement technologies challenge existing laws and policy, as well as underlying ethical values. But while the implications of human enhancement generally have been widely discussed, little analysis currently exists for the military context—specifically operational, ethical, and legal implications of enhancing warfighters, such as:

How safe should these human enhancements and new medical treatments be prior to their deployment (considering recent controversies such as mandatory anthrax vaccinations)? Must enhancements be reversible or temporary (considering that most warfighters will return to society as civilians)? Could enhancements count as “biological weapons” under the Biological and Toxin Weapons Convention (considering that the term is not clearly defined)?

This report begins an investigation into these and other issues in order to identify problems that policymakers and society may need to confront.

We start with an analysis of international and domestic law, military policy, bioethics, and risk assessments. Then we offer a new framework for evaluating human enhancement technologies in a military context. As an initial model, we also discuss further considerations—related to character and honor, as well as broader social impacts—that can be integrated later into this evaluative framework.

Given a significant lag time between ethics and technology, it is imperative to start considering the issues before novel technologies fully arrive on the scene and in the theater of war. Consider, for instance, the sudden explosion in number of robots in war and the ensuing confusion and controversies over their use. This report, therefore, is intended to help avoid similar ethical, legal, and policy surprises, as well as technology misuses that affect national reputations and real lives.

Disclosures

This report is funded by The Greenwall Foundation, with additional support by California Polytechnic State University's Research and Graduate Programs.

We acknowledge support by the institutions with which we are affiliated, including: California Polytechnic State University's College of Liberal Arts and Philosophy Department; Case Western Reserve University's School of Law and Department of Bioethics; Stanford Law School's Center for Internet and Society; and Australia's Centre for Applied Philosophy and Public Ethics.

We thank our consultants to the project and their respective organizations, who are: Dr. Edward Barrett (US Naval Academy; US Air Force colonel); Dr. Richard O'Meara (Rutgers University; retired US Army Brigadier General and JAG officer); Prof. George R. Lucas, Jr. (US Naval Postgraduate School); and Dr. Thomas Murray (The Hastings Center). Special thanks to Prof. Shannon E. French (Case Western Reserve University's Inamori International Center for Ethics and Excellence) and Prof. Shannon Vallor (Santa Clara University) for their contributions to the discussion. We also thank for their contributions and editorial assistance: Michael Burnam-Fink; Alexander R. LaCroix; Travis Rochelle; Jordan Rowley; and Seth G. Schuknecht.

Our research also has benefited from collaborations with and support from: the Consortium for Emerging Technologies, Military Operations, and National Security (CETMONS); Prof. Braden Allenby (Arizona State University); Brocher Foun-

dation (Switzerland); and the US Office of Secretary of Defense's NeXTech Wargames (Rapid Reaction Technology Office).

Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the aforementioned persons or organizations.

1. Introduction

War is an all-too-human affair, and it perhaps will always require payment in human lives. This is a terrible cost, but one which science and technology hope to ease. History has seen an evolution of defensive and offensive technologies—from shields and body armor to more accurate and longer range artillery—that are aimed exactly at minimizing the human cost, at least to our own side. In the Digital Age today, we are inventing entirely new and imaginative paradigms with military robots, cyber-weapons, and other technologies that promise to replace the organic, soft-bodied combatant and better protect noncombatants as well.

Yet it is difficult to imagine a plausible scenario in which human combatants will never be needed in war, no matter how advanced our technologies may be. No weapon or losses have been so horrific as to deter us from fighting again, as we have learned since World War I, quaintly billed as the “war to end all wars.” Even against daunting odds and fearsome machines, from tanks to flying drones, humans are tenacious and hopeful, refusing to give up so easily.

But therein lies a fundamental problem with how we wage war: As impressive as our weapon systems may be, one of the weakest links—as well as the most valuable—in armed conflicts continues to be warfighters themselves. Hunger, fatigue, and the need for sleep can quickly drain troop morale and threaten a mission. Fear and confusion in the “fog of war” can lead to costly mistakes, such as friendly-fire casualties. Emotions and adrenaline can drive otherwise-decent individuals to perform vicious acts, from verbal abuse of local civilians to

torture and illegal executions, making an international incident out of a routine patrol. And post-traumatic stress can take a devastating toll on families and add pressure on already-burdened health services.

Human frailty is striking and inescapable. Unlike other animals, we are not armed with fangs, claws, flight, venom, fur, or other helpful features to survive a savage world. It is a wonder our naked species has survived at all, if not for our tool-making intellect and resourcefulness. But our tools so far provide limited sanctuary from dangers. For instance, some estimates put the United States government’s investment in each soldier, not including salary, at approximately \$1 million [Shaughnessy 2012], helping to make the US military the best equipped in the world; nonetheless, that soldier is still largely vulnerable to a fatal wound delivered by a single 25-cent bullet.

If humans will always be called upon to fight, then it makes sense to focus efforts on overcoming that frailty. To be sure, military training attempts to address these problems, but it can do only so much. Science and technology again offer hope for whatever challenges we may face, in this case to upgrade the basic human condition. We want our warfighters to be made stronger, more aware, more durable, more adaptive, and so on. The technologies that enable these abilities fall in the realm of human enhancement. As we explain in the following sections, human enhancement technologies are more than mere tools: we are drawing in these tools so close to our bodies that

they become internal to us or, for all practical purposes, integrated with ourselves—and this creates special competitive advantages and, sometimes, special risks.

This report examines risk, ethics, and policy issues arising out of military human enhancements—not necessarily a new class of war technologies but one that is now developing in novel ways. We first explain the purpose for this report, as well as more fully explain what we mean by “human enhancement”; then locate the ethical and policy issues, and survey key military enhancement technologies worldwide. Next, we consider existing legal and ethical norms as we build a new framework to evaluate human enhancement technologies in a military context. Finally, we suggest other areas of ethics research needed to help guide ongoing discussions in this emerging area of science and technology.

1.1 Purpose of Report

Technologies related to national security and defense tend to occur on the edges of innovation, but in so doing, they can raise novel and important ethical and policy issues. As quickly as such technologies advance today and make their way into the public sphere, their impact on society becomes even greater, making their study more urgent.

Human enhancement is one of these rapidly advancing fields. Enhancement technologies can profoundly affect not only the individual soldier and the face of war, but also society at large, particularly as military personnel integrate back into civilian life as veterans. Even before that, there may be integration issues as warfighters rotate out of active-duty deployment and into less-demanding activities, such as management and continuing education and training. Currently, there

are approximately 23 million veterans in the US—or one out of every 10 adults—in addition to 3 million active and reserve personnel [US Census 2011]; therefore, the spillover effect alone into society may be significant, along with possible dual-uses of such technologies for civilian purposes.

Underlying technologies—such as neuroscience, biotechnology, nanotechnology, robotics, artificial intelligence, and more—offer the possibility of enhancing warfighters to help them complete missions safely and effectively. The US Defense Advanced Research Projects Agency (DARPA), for example, supports a host of research projects designed to build a warfighter that can learn better, eat grass, have super-vision, operate without sleep or food, swim like a dolphin, smell with the sensitivity of a dog’s nose, and climb walls like a lizard, among other capabilities. (See section 2.4 for a more detailed survey of these projects.)

The enhancement goal of creating a super-soldier is not unusual: that is essentially what we are doing with military robots, but from an engineering or mechanical starting point rather than a biological one. Indeed, we are beginning to see a convergence in the two approaches where robotics and computer interfaces are integrated with the human body. And where the development and deployment of military robots give rise to ethical issues [Lin et al. 2008], the research and use of military human enhancements do as well.

While there is a substantial literature on ethical, legal, and psycho-social issues raised by similar enhancements, there is little discussion in the context of military use or about the specific programs mentioned above. For instance, there has been some work on stimulant drugs [Mehlman 2009b; Russo et al. 2008; Meijer 2007; Jaeger 2007; Kautz et al. 2007; Roedig 2007; Russo 2007; Schoolmaker 2007]; on ethical and legal issues in

the conduct of enhancement research on warfighters [Mehlman 2009a; Mehlman 2009b, Mehlman et al. 2010]; and on drugs that impair moral responsibility [Ashcroft 2008; Wolfendale 2008; Vincent 2012]. As another researcher in this emerging field, Jonathan Moreno recognizes the need for ethical analyses of military uses of neuro-enhancement, though he does not provide such an analysis himself [Moreno 2006]. Catherine and George Annas argue that soldiers should not be compelled to use enhancements but only do so voluntarily and with the advice of a physician who cannot be ordered to prescribe them [Annas and Annas 2009]. Only recently is research emerging on law and military enhancements generally [Parasidis 2012].

Outside of the military, the use of enhancements is highly controversial, but it is unclear how these objections would apply to the military context, e.g., whether they would be overcome by the special nature of military service and the exigencies of military operations, since enhancements potentially could enable warfighters to achieve their missions more safely, effectively, and economically. On the other hand, the military needs to tread carefully in this area: It is important to avoid causing unnecessary or excessive harm to warfighters by subjecting them to unethical experimentation or requiring them to use unjustifiably dangerous enhancement products. Enhancements could also cause harm to noncombatants, if enhanced warfighters behave in unintended, unethical ways. Furthermore, inappropriate decisions by the military could provoke a negative public reaction that would impair recruitment and the military's ability to fulfill its legitimate responsibilities.

This report, then, is the first to critically examine a broad range of existing and new concerns arising from military human enhancements, serving to fill

the discussion gap. This gap is now visible given how important human enhancement research is potentially to both the military and broader society. The present lack of discussion, though, is perhaps understandable given the experimental nature of many of these innovations. However, some forms of enhancement are already being employed in the field, such as the alertness drug modafinil [Moreno 2006]. Moreover, enhancement technologies currently under study may soon reach the stage at which they would be considered for deployment.

With this report, we intend to help ethics, law, and policy—lagging behind, as is often the case—to catch up with science and technology. Consider, for instance, the explosion in number of robots in war: in its invasion of Iraq, the US had zero ground robots in 2003 and suddenly about 12,000 in 2008 [Singer 2009b, 61]; and its inventory of aerial robots multiplied by 40-fold between 2002 and 2010 [Gertler 2012, 2]. As public backlash today against the “drone wars” by the US demonstrates [Human Rights Watch 2012], it is imperative to start considering their impacts before technologies fully arrive on the scene and are deployed in the theater of war. With this report—and your participation in the discussion, as policymakers, scientists, academics, or just interested public citizens—we can better anticipate ethical, legal, and policy surprises, as well as avoid technology misuses that affect reputations and real lives.

1.2 Background

Since the beginning of human history, we have improved our minds through education, disciplined thinking, and meditation; and we have sought to improve our bodies with a sound diet, physical exercise, and training. But today, something seems to be different. With ongoing work in emerging

technologies, we are near the start of the Human Enhancement Revolution [Savulescu and Bostrom 2009; Allhoff et al. 2010a; Allhoff et al. 2010b].

We are no longer limited to “natural” methods to enhance ourselves or merely wielding tools such as hammers, binoculars, or smart phones. With drugs and devices, we are beginning to alter our biology and incorporate technology within our very bodies, and this seems to hold moral significance that we ought to consider. These technologies promise great benefits for humanity—such as increased productivity and creativity, longer lives, more serenity, stronger bodies and minds, and more—as well as compelling uses for national security.

The issues arising from human enhancement technologies include those related to autonomy and freedom, health and safety, fairness and access, social disruption, and human dignity [Garreau 2006; Selgelid 2007; Allhoff et al. 2010a]. For instance, critics question whether technological enhancements translate into *happier* lives, which many see as the point of it all [President’s Council on Bioethics 2003; Persaud 2006]. Of course, in a military context, general questions of human happiness are not explicitly addressed; rather, typical concerns are about accomplishing the mission and looking to the welfare of the troops, usually in that order. Insofar as context matters in making moral determinations here, the debate is made more complex with the need to account for the values and goods particular to the sphere under consideration [Murray 2007].

These concerns are driving the larger issue of whether and how society ought to regulate human enhancement technologies, which is closely related to how militaries ought to approach the same. For instance, one position is that (more than minimal) regulation would hinder personal freedom or autonomy, infringing on some natural or political

right to improve our own bodies, minds, and lives as we see fit [Naam 2005; Bailey 2005; Harris 2007]. Others, however, advocate strong regulation—and even a research moratorium—to protect against unintended effects on society, such as the presumably undesirable creation of a new class of enhanced persons who could outwit, outplay, and outlast “normal” or unenhanced persons for jobs, in schools, at sporting contests, and so on, among other reasons [Fukuyama 2002; Fukuyama 2006; Friends of the Earth 2006]. Still others recognize that society must adapt to new technological developments in this field [Hughes 2004; UK Academy of Medical Sciences 2012] and seek to strike a middle path between stringent regulation and individual liberty [Mehlman 1999, 2000, 2003, 2004, 2009a, 2009b; Juengst 2000; Greely 2005].

No matter where one is aligned on this issue, it is clear that the human enhancement debate is a deeply passionate and personal one, striking at the heart of what it means to be human. Some see it as a way to fulfill or even transcend our potential; others see it as a darker path towards losing our humanity [President’s Council on Bioethics 2003; Sandel 2009]. And to the extent that the US military is usually on the cutting-edge of science and technology research—having given society such innovations as the Internet, global positioning system (GPS), radar, microwaves, and even the modern computer—this project will examine the associated ethical and policy issues at their root.

Military Interest in Enhancement

The use of human enhancement technologies by the military is not new. Under some definitions—but not necessarily ours, as we explain below—vaccinations would count as an enhancement of the human immune system, and this would place the first instance of military human enhancement at the United States’ very first war, the America

Revolutionary War in 1775-1783: George Washington, as commander-in-chief of the Continental Army, ordered the vaccinations of American troops against smallpox, as the British Army was suspected of using the virus as a form of biological warfare [Fenn 2002]. At the time, the Americans largely were not exposed to smallpox in childhood and therefore had not built up immunity to the disease, as the British had. Biowarfare existed for centuries prior, e.g., in catapulting infected corpses to spread the plague in the Middle Ages [Cantor 2001; Lewis 2009].

More recently, and as a clearer instance of an enhancement, militaries worldwide have turned to amphetamines, though this trend is now shifting to new drugs. Amphetamines were used widely by American, German, British, and other forces in World War II, and again by the US in Korea [Stoil 1990]. Of course, milder and therefore less-controversial stimulants, such as caffeine in coffee and tea, have been used long before that.

Beginning in 1960, the US Air Force sanctioned amphetamines on a limited basis for the Strategic Air Command and, in 1962, for the Tactical Air Command. The US-Vietnam War sparked large-scale amphetamine use, such as by US Air Force and Navy pilots to extend their duty-day and increase vigilance while flying. According to one Cobra gunship pilot, “uppers” were available “like candy,” with no control over how much was used [Cornum et al. 1997]. During the US invasion of Panama (Operation Just Cause), the drugs were administered in smaller doses under much more careful medical supervision; and in contrast to Vietnam, where pilots who used them frequently suffered from nervousness, loss of appetite, and inability to sleep, fewer side effects were reported during Operation Just Cause [Cornum et al. 1997].

The US Air Force continued to dispense “speed” during Operations Desert Shield and Desert Storm, or the first Gulf War. A survey of 464 fighter pilots in that conflict found that, during the six-week operation, 57 percent reported that they took Dexedrine (an amphetamine) at least once, with 58 percent reporting occasional use and 17 percent admitting to routine use; and 61 percent of users believed the drug was essential for completing their missions [Schlesinger 2003]. In 1991, the US Air Force Chief of Staff, General Merrill A. McPeak, banned the use of amphetamines because, in his words, “Jedi Knights don’t need them” [Shanker and Duenwald 2003]. The ban lasted until 1996, when Chief of Staff John Jumper reversed the policy, as long-distance missions were being flown in Eastern Europe [Hart 2003].

In 2002, the US Air Force was dispensing 10 milligrams of amphetamines for every four hours of flying time for single-pilot fighter missions longer than eight hours and for two-pilot bomber missions longer than 12 hours. Asked why military pilots were permitted to use amphetamines when they were prohibited by commercial airlines, Colonel Peter Demitry, chief of the Air Force Surgeon General’s Science and Technology division, explained, “When a civilian gets tired, the appropriate strategy is to land, then sleep. In combat operations when you’re strapped to an ejection seat, you don’t have the luxury to pull over” [Hart 2003].

Amphetamines became controversial in 2002 when four Canadian soldiers were killed and eight wounded in a friendly fire incident in Afghanistan. They were hit by a 500-pound laser-guided bomb dropped from a US Air Force F-16, the pilots of which were returning at 18,000 feet from a 10-hour mission and mistakenly thought they were attracting small-arms fire [St. Louis Post-Dispatch 2003, C12]. When they learned of their mistake,

the pilots claimed that they were jittery from taking Dexedrine for so many hours [Schlesinger 2003]. One of the pilots had been an instructor in the Illinois National Guard and had graduated from the Navy's Top Gun school. The fatalities were the first Canadians to die in combat since the Korean War [Simpson 2004].

Amphetamines continue to be approved for military use, however; a 2005 article by research scientists at the Air Force Research Laboratory states that "the US Air Force has authorized the use of dextroamphetamine in certain types of lengthy (i.e., 12 or more hours) single-seat and dual-seat flight missions. A recent NATO Research and Technology Organization publication discusses amphetamine's significant value as an anti-fatigue measure for aviation personnel" [Caldwell and Caldwell 2005].

In an effort to find a safer alternative to amphetamines, the military is reported to be turning to modafinil, a drug originally used to treat narcolepsy and which is sold under the brand name Provigil. According to Jonathan Moreno, US troops first used modafinil during the 2003 invasion of Iraq. The British press reports that the UK Ministry of Defence purchased 24,000 modafinil tablets in 2004 [Sample 2004].

Research has shown that the drug improves the performance of helicopter pilots in flight simulators [Caldwell, Caldwell, Smythe, and Hall 2000]. Moreno reports on a modafinil study that the Air Force's Office of Scientific Research conducted with 16 volunteers who, over a four-day period, stayed awake for 28 hours, then slept from 11 am until 7 pm: the modafinil group did significantly better on cognitive tests than subjects who took a placebo. Other research showed that modafinil-enabled pilots to remain alert for 40 hours, and experiments at Walter Reed Institute of Research

have been carried out on soldiers who were sleep-deprived for as long as 85 hours [Sample 2004]. We discuss the moral and legal propriety of doing such research on the military later in this report. While the military is actively investigating new alertness drugs like modafinil, it continues to employ the old standby, caffeine. New US army "first strike" rations contain caffeine-laced chewing gum, with each stick providing the equivalent of a cup of strong coffee [Sample 2004].

In addition to using alertness drugs to enhance performance, a long-standing practice among members of the military has been to take dietary supplements. As reported in the journal *Military Medicine*, "a recent worldwide survey showed that over 60 percent of service members are regularly taking some type of dietary supplement. Usually, supplement use is at the advice of the sales clerk or by getting information from magazines or peers. Evidence-based information is rarely available or rarely translated into a form that can be properly used by the warfighter or their commander" [Jonas et al. 2010]. Despite the limited amount of scientific evidence, the military recognizes the potential value of supplements: "Nutritional supplements may indeed be beneficial in certain circumstances. For example, caffeine may provide advantages in military jobs and duties where attentiveness is necessary (e.g., aviators, sentry duty)" [Montain, Carvey, and Stephens 2010].

In 2008, Brookings Institution fellow Peter W. Singer reported an ambitious DARPA goal presented by DARPA program manager Michael Callahan at the agency's 50th anniversary conference in 2007: "making soldiers 'kill-proof'" [Singer 2009a]. Callahan described research that would enable soldiers "to bring to battle the same sort of capabilities that nature has given certain animals," including a sea lion's dive reflex; "products in the pipeline" such as "drugs that will boost muscles

and energy by a factor of 10, akin to steroids ... on steroids,” which Singer says “is jokingly termed the ‘Energizer Bunny in Fatigues.’”; and a long-term \$3 billion initiative entitled the “Metabolically Dominant Soldier,” which bioethicist Jonathan Moreno claims is aimed at developing a super-nutritional pill that, in DARPA’s words, would permit “continuous peak performance and cognitive function for 3 to 5 days, 24 hours per day, without the need for calories” [Moreno 2006]. A 2007 article in *Wired* identified extramural enhancement research projects sponsored by DARPA at the Dana-Farber Cancer Institute in Boston to develop substances to make soldiers more energetic; at Columbia University to enable soldiers to make do with less sleep; and in Ames, Iowa, where agricultural experts are researching bacteria that, once ingested, would enable soldiers to obtain nutritional value from normally indigestible substances such as cellulose [Shachtman 2007a, 2007c].

In 2008, JASON, a group of scientific advisors to the US Department of Defense (DoD), issued a report on “Human Performance” that discussed several types of biomedical enhancements, including the potential use of a class of compounds called ampakines to enhance cognition [JASON 2008]. A central point in the report was that the benefits from military enhancement were not similar to the benefits from performance enhancement in elite sports: “The consequences of gaining a small performance advantage, even if it is highly statistically significant, are likely quite different as regards force-on-force engagements than as regards Olympic competition. In brief, a small performance advantage in force-on-force should generally result in a *small* change in the outcome, while in Olympic competition it can result in a *large* change in the outcome” [JASON 2008, 15].

At the same time, the report acknowledged that a major change in human performance, giving as an

example a reduction in the need for sleep, could have a “dramatic effect” on the “balance of military effectiveness” [JASON 2008, 37]. However, this dramatic effect would occur only if the adversary did not have access to the same benefit, leading the report to emphasize the need to monitor and be prepared to counter-enhancement use by potential adversaries.

Lately, the US military has extended its interest in performance enhancement to genetic technologies. In December 2010, JASON issued a report entitled “The \$100 Genome: Implications for the DoD.” The report outlined an ambitious plan to employ genomic technologies to “enhance medical status and improve treatment outcomes,” enhance “health, readiness, and performance of military personnel,” and “know the genetic identities of an adversary” [JASON 2010].

At the same time, the US military appears to be dead-set against the use of steroids. DoD Directive 1010.1, originally issued in 1984, states that “the illicit use of anabolic steroids by military members” is an offense under the Uniform Code of Military Justice [US Department of Defense 2012a].

The US military’s current interest in biomedical enhancement is a logical corollary to its objective of maximizing the performance capabilities of its members. This performance imperative is reflected in the goals of military training. As an Army training manual states:

War places a great premium upon the strength, stamina, agility, and coordination of the soldier because victory and his life are so often dependent upon them. To march long distances with full pack, weapons, and ammunition through rugged country and to fight effectively upon arriving at the area of combat; to drive fast-moving tanks and motor vehicles

over rough terrain; to make assaults and to run and crawl for long distances; to jump into and out of foxholes, craters, and trenches, and over obstacles; to lift and carry heavy objects; to keep going for many hours without sleep or rest—all these activities of warfare and many others require superbly conditioned troops [Roy et al. 2010].

The recent interest in military performance optimization has led to expansions of the concept of warfighter fitness. In 2005, Army Field Manual 21-20 was replaced by Training Circular 3-22.20, “Physical Readiness Training (PRT),” which describes as its purpose “to develop a more agile, versatile, lethal, and survivable force—while preparing Soldiers and units for the physical challenges of fulfilling the mission in the face of a wide range of threats, in complex operational environments, and with emerging technologies” [Little 2010]. As the deputy commander of the US Army Training and Doctrine Command explained, “the youngest generation has grown up with energy drinks and soda while playing video games on the couch, instead of drinking milk and taking physical education classes in school The Army has seen a major increase in dental problems and bone injuries during basic training. In the last 15 years, average body fat has also increased to 30 percent in the South. ... The challenge is taking young Soldiers entering the Army under these conditions and getting them ready to hump the Hindu Kush,” the 500-mile mountain range between northwest Pakistan and eastern and central Afghanistan [Little 2010].

In 2006, a DoD conference titled “Human Performance Optimization” led to the creation of a dedicated human performance office within Force Health Protection in the Assistant Secretary of Defense Office (Health Affairs) [Land 2010]. The attendees at this conference recognized the need

for a holistic “total force fitness” approach, which subsequently was adopted by the Army in 2008 under the name “Comprehensive Soldier Fitness Program.” This program is described as signaling “the US Army’s attempt to bring science to bear on a complex problem—shaping and accelerating human development and performance. The program is massive in scale and will directly impact three distinct populations—US Army soldiers, their family members, and civilians employed by the Army” [Lester, McBride, Bliese, and Adler 2011].

As these new training initiatives make clear, the goal is to go beyond preparation for the demands of military service and instead enable “functioning at a new optimal level to face new missions or challenges” [Jonas et al. 2010, 9]. Biomedical enhancement is one of the obvious technologies that might be employed in responding to this new performance imperative.

1.3 Questions

This report engages key issues related to military enhancements—from the other-regarding (issues about the impact on others) to self-regarding (issues about the impact on the enhanced individual)—and provides a framework to more fully evaluate them, including the following:

A. Law and Policy

- Could enhanced warfighters be considered to be “weapons” in themselves and therefore subject to regulation under the Laws of Armed Conflict (LOAC)? For instance, must militaries perform a legal review of enhancements as weapons, under Article 36 of the Geneva Conventions, Additional Protocol I? Or could an enhanced warfighter count as a “biological

agent” under the Biological and Toxin Weapons Convention (BTWC)?

- How would other LOAC apply to enhanced warfighters? For example, would a prohibition against torture apply equally to enhanced warfighters who can tolerate greater physical or mental abuse?
- How would enhanced warfighters be viewed by adversaries, and what effect would this have on mission success? For example, would enhanced warfighters be viewed as no longer “human,” and therefore open to inhumane treatment?
- Would deployment of enhanced warfighters tend to lower the threshold for going to war, and does that violate the principle of last resort?
- How might adversaries attempt to negate the effects of enhancements?

B. Operations

- Would having enhanced and unenhanced warfighters together affect unit cohesion? Should enhancements be confined to a small, elite force or used more widely?
- Does being enhanced require a change either way in the length of a warfighter’s service commitment? Is the ability to remain enhanced after leaving the service a legitimate recruiting incentive?
- How should being enhanced affect a warfighter’s prospects of promotion, and how would this be viewed by others who are unenhanced?
- Would official restrictions on warfighter enhancements that are viewed as desirable by command or by the warfighters themselves lead to a black market, or further encourage use of private military contractors (PMCs) more willing to undergo risky enhancement?

C. Civil-Military Relations

- If the enhancements are not reversed, what effects will they have on the ability of the enhanced warfighter to make a successful return to civilian life? How can any adverse effects on the return to civilian life be minimized?
- What effect will this have on the US Department of Veteran Affairs (VA), insofar as enhancements will be regarded as service-connected?
- If there are enhancements available in the civilian market but not in the military, should warfighters be permitted to purchase and use them?
- What are the societal implications if/when military enhancements have a dual-use in civilian applications?
- Given the hostility toward the use of performance-enhancing substances in sports, would their use by the military provoke a negative public or political reaction that would undermine military recruitment, retention, or funding, or otherwise interfere with the mission of the military?

D. Ethics

- Should warfighters be required to give their informed consent to being enhanced, and if so, what should that process be?
- If a standard other than informed consent should be used, what additional protections are needed, and under what circumstances would they supersede a need for consent?
- Can a warfighter successfully object to being enhanced on religious or other grounds?
- Does the possibility that military enhancements will simply lead to a continuing arms race mean that it is unethical to even begin to research or employ them?

- Must the enhancements be reversible? If so, should they be reversed routinely upon discharge? Should the warfighter be able to refuse to have them reversed upon discharge?
- How should the use of enhancements be taken into consideration for purposes of commendation, promotion, and access to specialized training based on aptitude?

E. Safety and Risk

- Legally and ethically, how safe should these technologies be prior to their deployment? Should enhancements that pose longer-term risks be required to be reversible?
- Are there ethical, legal, psycho-social, or operational limits on the extent to which a warfighter may be enhanced as well as types of enhancement, e.g., enhancing mood to induce euphoria, fearlessness, or amnesia?
- Who gets to determine what constitutes an “acceptable risk” of a proposed enhancement?
- When can “military necessity,” as determined by the chain of command, trump a different determination that an enhancement is unacceptably risky? Could enhancements jeopardize the chain of command?

Undoubtedly, there are other important questions arising from military human enhancements. Our goal here is not so much to directly answer each one—a satisfactory discussion of which is beyond the scope of this report—but to provide a framework to begin much-needed discussions about them.

2. What Is Enhancement?

To properly start, we need to offer a definition of “human enhancement”, which has been elusive. Our definition follows that suggested by bioethicist Eric Juengst: an enhancement is a medical or biological intervention introduced into the body designed “to improve performance, appearance, or capability besides what is necessary to achieve, sustain or restore health” [Juengst 1998]. More will be said about this definition shortly.

In explaining what a human enhancement is and what it is not, we might start with ostensive definitions or illustrative examples, as we attempt to arrive at a principled distinction between the two. An enhancement, strictly speaking, is anything that improves our lives or helps us to achieve our goals, including the goal of survival. Thus, healthy foods, exercise, fire, tools such as a hammer, a roof over one’s head, and so on are enhancements in at least the sense that they clearly enhance our lives or make them better.

Notice that these enhancements could also pose dangers, or a net loss in welfare, if used in certain ways: A diet could be fatal to a person who is allergic to its ingredients; exercise can lead to pulled muscles and more serious injuries; fire can burn down communities of homes; hammers can be turned to weapons; and a roof can fall on one’s head in an earthquake. Indeed, it is difficult to imagine what exactly would count as an enhancement, if we were to insist that enhancements deliver only benefits.

But we do not mean “human enhancement” in this general sense. If we did, the notion would become so broad that it is rendered meaningless. Nearly everything we create could count as an enhancement, from Lego blocks to language itself, as long as we derive some benefit from it. Thus, if we are to use the term in a meaningful way—that is, to examine whether novel ethical issues arise from enhancements, as many intuitively suspect—we need to describe a limit to what counts as enhancements, and we need to defend that line, to the extent that there is no general consensus on a definition.

As such, we count the following as examples of human enhancement: an athlete who is stronger with anabolic steroids use; a student who earns higher grades by using Ritalin or modafinil to study more effectively; in the future, a soldier who can run for days on a drug that triggers in humans whatever metabolic processes enable an Arctic sled-dog to run for that long; and in the future, an office worker who is smarter than her peers, given a computer chip implanted in her head that gives instant access to databases and search engines.

In contrast, we do *not* count the following as examples of human enhancement: a muscular dystrophy patient who uses anabolic steroids to regain some strength; an expensive prep school background to enable a student to earn higher grades; a vehicle that can transport soldiers for days; a smartphone that enables a worker to check email and look up information.

But why should this be? Why are some technologies or applications considered as enhancements, while others that seem directed at the same goal are not? It is not enough that we offer a few examples of enhancements and nonenhancements; we need to explain the principle or reasoning for making this discrimination. If we are not aiming at the right targets to being with, our understanding of “human enhancement” would be arbitrary, and the ethical and policy analysis that follows could fail to be relevant or lack the desired force. To that end, we offer the following discussion.

2.1 Controversies

Any distinction that is meant to delimit enhancements instantly takes us down difficult philosophical rabbit-holes. This is why a definition of enhancement has been so elusive, and also why some commentators have denied that a definition can be found. To seriously consider the possibility that some enhancements raise novel ethical issues, though, let us assume that we can define enhancement, even if imperfectly, before we abandon hope for such a project. We will not explore this conceptual maze in much depth here, but only identify a few points of contention to convey a sense of how difficult it is to nail down a definition of enhancement:

A. *Natural vs. Artificial*

As a first approximation of a distinction between nonenhancements and enhancements, we may be tempted to look at the distinction between natural and unnatural (or artificial) [Bostrom and Sandberg 2009; Allhoff et al. 2010a]. That is, medical treatments for the sick, exercise, and education are “natural” activities that humans have been doing throughout history. Insofar as

what is natural is good, these activities are not morally problematic. In contrast, drugs that give us the endurance of sled-dogs and other such enhancements take us beyond “natural” limits of human functioning. In that sense, enhancements are unnatural or artificial aids, and what is unnatural should evoke caution and skepticism—or so the distinction would seem to imply.

Never mind how such a distinction would be morally relevant, it quickly collapses upon reflection. In a general context, we might say that trees and rocks are natural in that they exist independently of human agency or intervention, while houses and computers are artificial in that they depend on our manipulation of materials. But many things we would consider to be natural depend on external manipulation, such as a bird’s nest or a beaver’s dam. If we then retreat and stipulate that external manipulation means *human* manipulation, then nothing created by humans can be considered to be natural. In other words, this move does not get us closer to what human enhancement is, if the definition tracks a natural-versus-artificial distinction: everything created by humans would then be artificial and an enhancement. As mentioned before, this notion seems unreasonably too broad; and it defies common intuitions that education and exercise—which involve printed books, microscopes, athletic shoes, and even high-priced Olympic coaches—are not enhancements but mere activities that benefit humans.

In the alternative, to the extent that humans arise from nature, there is a sense that everything we do is natural. But this conception suffers from being too broad in the opposite direction: nothing we do can be artificial, and so this too does not move us closer to understanding enhancement via the natural-versus-artificial distinction. Where we consider

mass education and high-tech exercise today to be natural, surely these would have been considered as unnatural at earlier times in human history, before the printing press, running-shoe technologies, and so on.

B. External vs. Internal

If the natural-versus-artificial distinction is untenable, then perhaps we can turn to another one: the distinction between an external tool or technology and an internal one [Garcia and Sandler 2008; Allhoff et al. 2010]. For instance, an Internet-enabled smartphone is a mere tool or something less than an enhancement, because it is external to our bodies; it could be held in one's hand, or placed inside a pocket, or connected to a wall charger. But a computer chip implanted in one's head, that delivers the same capabilities as the smartphone, is an enhancement; it is internal to one's body, and this delivers "always-on" or unprecedented access to the tool—and competitive advantage from its benefits—as compared to using it as an external device [Allhoff et al. 2010a; Lin 2010]. Likewise, athletic shoes are not properly enhancements, since they are external devices and not always worn, while anabolic steroids are enhancements because they are consumed or injected into the body.

"Internal" technologies could also be construed to include tools that are closely integrated to one's body, since that too delivers an "always-on" connectivity that does not exist with ordinary external tools. Bionic limbs that deliver super-strength, for instance, are not internal to a body, strictly speaking, yet we may consider them to be enhancements; they are attached to a body and become part of the person's identity, if that's important. Exoskeletons today, then, are mere tools, as they are bulky and cannot be easily

worn for a long stretch of time; but if they were to become much more lightweight and unobtrusive, perhaps wearable like a shirt, then a case could be made for declaring them an enhancement.

The proximity of a device to the body can create a difference in degree that becomes a difference in kind. For instance, compare a person who looks up information on Google's search engine—on either his laptop or mobile device—to another person who looks up the same information through a "Google chip" implanted in her head. We would not say that the former is more knowledgeable or smarter for reciting information he found online; at best, he is resourceful. But while the latter may also just be reciting information she found online, her ability to do so at virtually any time—and seamlessly or transparently to others—would make her appear more knowledgeable, as if she were a savant who has uncanny recollection of facts and trivia (especially, say, in an exam room in which she is not supposed to have a computer).

Similarly, compare a person who uses Google Translate to communicate with the local population on his trip abroad to a person who has a Google translation chip implanted in her head to enable the same communication. The former would be recognized as someone who merely knows how to use a computer, while the latter could very well be mistaken as being fluent in the foreign language (again, say, in a test-taking environment). In other words, when it comes to proximity of a technological aid to the user, closer is generally better. We can thus defend the line between enhancement and mere tool in terms of internal versus external; and

closely held or worn tools are “internal” enough, if the user is rarely without them.¹

A limitation to the external-versus-internal distinction, however, is that it cannot account for a dual-use technology that is internal-only. For instance, the distinction does not speak to any *prima facie* moral difference between anabolic steroids taken by a muscular dystrophy patient and the same drug taken by an Olympic athlete; they are both cases of a pharmacological intervention that is internal to the body. Many critics strongly believe that the former case is morally uncontroversial while the latter is not, and so our distinction ought to be able to sort those cases if possible, at least for further examination.

C. *Enhancement vs. Therapy*

A more capable distinction, then, could be that between therapy and enhancement [President’s Council for Bioethics 2003; Allhoff et al. 2010a]. This distinction easily accounts for the case of anabolic steroids that was so difficult for the external-versus-internal distinction: a muscular dystrophy patient needs anabolic steroids for therapeutic value or medical necessity, while the athlete has more gratuitous and less urgent reasons. So the patient has greater moral justification for using the drug, and this maps to popular intuitions (in case they carry any weight) that therapeutic uses are permissible while

.....
¹ We reserve discussion on whether the neurally enhanced person is in fact more knowledgeable or smarter, or if she even understands the information she is receiving. Insofar as human enhancement is a relational concept—i.e., it requires that the majority of a population is not enhanced—how a subject is perceived by others is relevant to the determination of whether she is enhanced or not.

athletic-performance enhancing uses are questionable.

By “therapy”, we mean an intervention or treatment intended to alleviate a condition suffered by a patient, elevating his or health closer to normal. Of course, “normal” raises a host of questions itself, given the diversity of abilities across the species; for instance, Albert Einstein throws off the curve, as well as across one’s own lifetime, e.g., health decreases with age. So if we may think of normal in terms of a species-typical range of functioning, that understanding may also need to be balanced against what is normal relative to one’s own baseline and trajectory of health.

Thus, if a drug could give an average person the IQ of Einstein, and Einstein does not fall in the species-typical range but exceeds it, then use of the drug in this case is an enhancement and not therapy. However, if Einstein were still alive and suffered a brain injury that reduced his IQ, then his taking of the same drug is not an enhancement, because it serves to restore his abilities back to his normal or baseline level. Enhancement, then, is highly sensitive to context.

A more difficult case to reconcile involves vaccinations: are they an enhancement or therapy [Bostrom and Roache 2008; Allhoff et al. 2010a]? On one hand, the recipient is typically not sick when receiving the vaccination, so there’s no immediate goal of restoring health to previous levels, as in most other cases of therapy. In this respect, a vaccination seems like an enhancement of one’s immune system: it is not in the species-typical range of functioning that humans can naturally resist many pathogens. However, a vaccination could be considered to be preventative therapy: why

should it matter if a therapeutic intervention—that is, designed to restore health back to normal—is administered before or after an illness?

For the purposes of this report, we will consider vaccinations to generally be a case of therapy, not enhancement, insofar as they aim to prevent a diseased condition. But we acknowledge that there may be nuances to this determination, which we lack the space to fully discuss here. For instance, perhaps it is reasonable to say that some therapies (such as vaccinations and accelerated healing) are also enhancements. Further, it may be pertinent whether a disease is caused by a naturally occurring pathogen or a completely engineered one; whether a disease can be caused by something other than biological agents, e.g., a chemical poison; and the related matter of how we define “disease.” Again, context matters.

Notwithstanding difficult cases such as the above, the distinction between enhancement and therapy seems sensible, and so we will adopt it here as the primary heuristic or rule-of-thumb in identifying what is an enhancement and what is not. This distinction, though, is not binary and therefore may not always be sufficient; that is, what is not therapy might not automatically be an enhancement. For instance, the external-versus-internal distinction is still useful to explain why carrying a smartphone is morally permissible—neither therapy nor enhancement, but simply a tool—but integrating it with one’s body is less obviously so, in that it may raise questions about safety, fairness, and other issues. (We discuss these issues in later sections.) Other distinctions may still be relevant, as fuzzy as they may be; for instance, notice that the concept of “natural” slips back into our analysis in the

preceding paragraph, and we will return to it again shortly.

D. Enhancements vs. Disenhancements

If there can be human enhancements, then it also seems possible to have “disenhancements.” An unenhanced person is simply one who is not enhanced, but a disenhanced person is one who undergoes an intervention that makes him or her worse off. This possibility speaks to our previous note that enhancements may pose dangers or lead to a net loss in welfare.

Although it may be difficult to envision the scenario in which a person would be disenhanced, we note that our livestock can be and sometimes are disenhanced. For instance, chickens sometimes have their beaks sheared off, so that they don’t peck other chickens when jammed into confined spaces [Hester and Shea-Moore 2003]; and blind chicken are known to fare better in crowded environments, as they don’t seem to mind as much as sighted birds [Thompson 2008]. Of course, this procedure may be an enhancement or benefit to us as the carnivorous consumer, but likely not a benefit from the chicken’s perspective. On the other hand, it has been proposed that we ought to create blind chickens, because they don’t seem to mind crowded conditions as much, and this is more plausibly a benefit to them [Thompson 2008].

For humans, removing some feature by itself is not automatically a disenhancement, if it delivers a benefit, especially for the subject. Today, for example, drugs are under development that can selectively target and erase memories [Lehrer 2012]. This would have beneficial uses, such as removing tragic memories that cause warfighters to have post-traumatic stress disorder (PTSD)

and rape victims to be unable to live normally without paralyzing fear. That is, their lives would be enhanced by the degradation of their memories. However, there may be more insidious uses of the same drugs that may count as true disenchantments that promise little to no benefit to the subject; for instance, we may threaten a terrorist with a drug that removes the memory of his family or life, if he does not reveal the information we want.

In bioethics, there are infamous cases of patients who want to amputate otherwise-healthy limbs [Dyer 2000; Ryan 2009], or deaf parents who want to select *in vitro* embryos that would lead to deaf babies [Dennis 2004]. From an outsider's perspective, this seems to be voluntary disenchantment and perhaps unethical. But from the patient's perspective, it is a case of therapy in that the limbs seem alien and are therefore unwanted. This and other cases, as well as the preceding ones, are also difficult cases in understanding what an enhancement is.

E. Enhancement vs. Engineering

The last distinction above suggests that more precise terminology may be needed here. Indeed, the US military appears to be shying away from the term "human enhancement" in favor of "performance optimization" and less evocative language; some of its past projects were called Metabolic Dominance and Peak Soldier Performance [Burnam-Fink 2011]. While we recognize that different terminology may be also suitable, we nonetheless will refer to the same technologies primarily as "human enhancements" for the following reasons:

First, they appear to be functionally equivalent: optimizing performance and enabling superior metabolism are both about enhancements to the

human body; as we have defined enhancement, they are not meant to be therapeutic and are intended to exceed typical human limits. Second, "enhancement" is the primary language used by the ethics literature on the subject, and we would like our terminology to be consistent with that body of work, to avoid any confusion. Third, the military's apparent preference for "performance optimization" over "enhancement" could be viewed as a public-relations choice about avoiding the specter of ethical concerns that has been associated with human enhancements; in contrast, the very purpose of this report is to examine such ethical concerns. The perception is that "optimization", "peak performance", and similar phrasing do not immediately suggest the familiar charge of hubris that we are pushing science beyond what is natural or ethical—in this case, to create a superhuman.

But one person's Superman is another's Frankenstein's monster. This, then, raises another issue in terminology, that we perhaps ought to use the more neutral "human engineering" rather than value-laden "human enhancement" [Allhoff et al. 2010a]. That is, "enhancement" seems to imply a net benefit to the individual, for instance, resulting in increased endurance, greater concentration, or some other desired good desired. Yet it is often unclear whether the technologies we have used to improve aspects of ourselves truly deliver a net benefit, especially without long-term studies of such effects. For instance, anabolic steroids can help athletes gain strength, but they may also cause serious health conditions; therefore, if anabolic steroids can do more harm than good, then it seems premature to label them as enhancements as opposed to, say, poisons. Similarly, alcoholic drinks seem to deliver the benefit of enhancing one's mood, but too much can lead to a painful hangover, liver

damage, addiction, and so on, making it an open question of whether they are properly enhancers or not. As enhancements are usually context-specific and intended for a particular purpose, it may be misleading to discuss “enhancements” in general, absent some larger account of their role in human flourishing (e.g., see section 7.1).

Even less clear are the benefits and risks posed by emerging and speculative technologies that have not been studied nearly as much as alcohol, anabolic steroids, caffeine, and other familiar drugs. As rapidly advancing as modern science seems to be, there is still much that we do not know. But what we know for certain is that biological and neurological (as well as environmental or ecological) systems are highly complex and interconnected, making it very difficult for us to accurately predict the effects of any given drug or technological intervention in our bodies. In the face of that uncertainty, we should refer to such technologies as instances of human engineering, rather than enhancements—or so the argument goes.

Even so, this report will stay with the notion “human enhancement”, with the understanding that the term does not necessarily imply a *net* benefit to the individual (much less to the larger society of which they are a part, especially given the likelihood of unintended consequences). Of course the term does signal a promise of delivering *some* benefit to the individual—otherwise why would we give a drug or apply a technology to a person, if we did not expect those benefits?—but we remain agnostic or neutral on the question of whether it results in an overall positive gain, all things considered. That open question is important to note but, for us, not at the expense of breaking from the traditional language of the ethics literature, especially since it is understood that human enhancements may raise

significant ethical concerns, including health risks—and this is the point of examining the subject in the first place.

Again, we won’t more fully explore the controversies here for space considerations, but they are important to acknowledge as we perform an ethical, legal, and policy analysis of human enhancements.

2.2 Working Definition

Given the preceding, we will operate under the working definition that **an enhancement is a medical or biological intervention to the body designed “to improve performance, appearance, or capability besides what is necessary to achieve, sustain or restore health”** [Juengst 1998].

By this definition, since an enhancement does not aim to prevent, treat, or mitigate the effects of a disease or disorder, a vaccination or immunization—even if it makes individuals’ immune systems better than normal—would typically not qualify as an enhancement, as it merely aims to sustain health. Similarly, a drug to improve cognitive function in persons with below-normal cognitive ability ordinarily would not be considered an enhancement. But a pharmacological agent that improved their cognitive function to such a degree that the individual exceeded population-norms for cognitive functioning clearly would qualify as an enhancement. An intervention, such as caffeine and modafinil, also might be regarded as an enhancement if it improved the cognition of someone with normal cognition to start with, even though the resulting cognitive performance remained within population norms. Again, it seems relevant to also consider the baseline

state and potential of the subject, as difficult as they are to measure.

The concept of normality, as we mentioned, is itself elusive; what is “normal” for one population may be quite abnormal for another. If an average professional basketball player began to suffer from scoliosis and lost four inches from his current height, he would still be far taller than normal for an adult male human. And even within populations, the term retains some ambiguity. In some cases, normality refers to the frequency with which a trait or capability occurs within a population, presuming a normal distribution. In regard to height, the convention is to regard individuals who are more than two standard deviations below the mean height of the population as being of short stature [US Department of Health and Human Services 2008]. Such an approach may leave unresolved the dilemma between understanding normality as individual-relative versus population-relative, however; consider two eight-year old boys, both shorter than 98 percent of their peers. One’s shortness is genetic; the other has parents of typical height but a glandular problem. Is giving a human-growth hormone (HGH) to them enhancement in both cases, therapy in both cases, or enhancement in one but therapy in the other?

To make things even more complicated, sometimes what is considered normal may have little to no relationship to the distribution of a trait. For example, normal eyesight is deemed to be 20/20, but only about 35 percent of adults have 20/20 vision without some form of correction [Johnson 2004]. Standards of normality also may vary from place to place and time to time, and can be expected to change as the use of enhancements increases. For example, body shapes that were considered healthy a hundred years ago are now considered obese; and the

advertising and pharmaceutical industries are notorious for taking conditions heretofore taken as normal (such as body odor or discolored teeth) and presenting them as conditions requiring treatment. Furthermore, the concepts of disease and disorder themselves may be hard to pin down. Before 1973, the American Psychiatric Association regarded homosexuality as a mental disorder [American Psychiatric Association 2012]. The tendency seemingly only grows to regard increasingly more health-states as diseases and increasingly more interventions as treatments.

Insofar as enhancement is related to health, and health is related to normality, of course we would prefer that these concepts were all clearly defined. As a foundational notion, what is normal may still lack precision, so it may be tempting to discard the notion and therefore the interconnected chain of concepts. But there are good reasons to retain “normal” in our conceptual toolbox [Sparrow 2010], at least as a useful rule-of-thumb, even if imprecise.

In short, the distinction between health-oriented and enhancing interventions will not always be clear, and invariably there will be borderline cases. The difficulty of clearly identifying what counts as an enhancement complicates the task of determining the conditions, if any, in which it would be ethical to research or use enhancements in the military. Nevertheless, the above working definition is sufficient to allow us to draw some conclusions about the ethical and legal propriety of military enhancement.

2.3 Variables

A number of variables affect the analysis of the ethical and legal issues raised by military

enhancement. First is the matter of perspective: Obviously a primary perspective is that of the warfighter who is being asked to serve as a research subject or who is going to take the enhancement. There are different types of warfighters, with potentially different needs and concerns. Career enlistees presumably might be expected to care less about the impact of enhancements on their return to civilian life than enlistees who did not intend to make the military their career. Members of the Reserves and the National Guard might be legitimately concerned about whether an enhancement would be a boon or a handicap in their civilian jobs. Warfighters who engaged in direct combat might be more willing to take risky enhancements than service personnel or operators of drones and other remote weapons.

Special-operations personnel in particular are known to be risk-takers, including in the area of increasing their mission effectiveness, such as by intense training. This might make it necessary to protect them from voluntarily agreeing to take potentially dangerous enhancements. On the other hand, if these troops are sent on more dangerous missions than regular troops, their willingness to take greater risks to improve their performance would be understandable, and this could be grounds for treating them differently. However, caution should be exercised in policy choices that create class divisions—for instance, special treatment or different rewards—within a military, to the extent they cause dissension in the ranks.

Another perspective is that of the other members of the enhanced warfighter's unit, who will share to some degree in the benefits and burdens of the enhancement use.

A third perspective is that of the warfighter's superiors. There are a number of different types of superiors, each of whom has a somewhat different role and therefore a somewhat different viewpoint, including the immediate commander who is considering whether or not to give a warfighter an enhancement in the field; the officer in charge of a military research project who is considering whether to enroll a warfighter as a subject; military policy-makers deciding whether to embark on an enhancement research program; and the officers supervising enhancement research programs. The perspective of the warfighter's superior also will be affected if he or she is a physician or other type of health care professional, since health care professionals are subject to both military and professional normative regimes.

Finally, there is the perspective of third parties such as family members, members of military outside of the warfighter's unit, civilians with whom the warfighter interacts, the government, the public, and the nation. As we will see, a critical question is what should happen if these perspectives are in conflict, for example, if consideration of the welfare of the individual warfighter points the resolution of an ethical or legal issue in a different direction than the welfare of the unit, the military service, or the state.

A second variable that must be considered when addressing the ethical and legal concerns raised by military enhancements is the risk or other adverse consequences associated with their use. These may be physical or mental health risks, such as those sometimes attributed to anabolic steroids and other drugs, including addiction. In contrast to the US, for example, the Danish military does not give amphetamine to its combat troops; not only do the Danes question the

use of a drug that can impair judgment, but they are concerned about the potential for addiction [Nielsen 2007]. The degree of health risk may depend on whether the enhancement effect is permanent, long-term, or temporary, and on whether or not it can be reversed.

There may be adverse effects on third parties, such as family members who are impacted by the warfighter's adverse health effects, as well as harm to family and other relationships. One factor in terms of relational effects is likely to be how significantly the warfighter's characteristics are altered by the enhancement, and how perceptible the enhancements are. A radical change in someone's appearance or behavior could have serious social consequences, and the negative effects would be even greater if the change were so extreme that it provoked repugnance or horror.

Another harmful effect could be the reaction of adversaries. The more far-ranging the change brought about by the enhancement, the greater the risk that an adversary might view enhanced warfighters as no longer being really "human," and therefore treat them worse as prisoners. Adversaries also might use harmful methods to combat or reverse their enemy's enhancement effect.

A final element of risk is uncertainty. The less that is known about an enhancement, the less it has been properly studied, the more difficult it is to engage in the key ethical and legal process of balancing risks and benefits. At the same time, it is important to understand that no biomedical intervention is completely safe—all are accompanied by risks. A substance as safe as pure water can be deadly, as a Sacramento radio station learned when a participant died after drinking too much of it during a contest called "Hold your

pee for a Wii," in which a Nintendo console was awarded to the person who could consume the most water without going to the bathroom [Nevius 2007]. The question is not how safe an intervention is but whether its risks are outweighed by its benefits. When the US Food and Drug Administration (FDA) approves a new drug or device as "safe," for example, what the agency really is saying is that it considers the health hazards of the product to be acceptable in view of the potential health benefits.

In the case of civilian medical care, for the most part the benefits that must be balanced against the risks are benefits that accrue to the patient. (The main exception is medical interventions to promote public health, about which more will be said in section 4.3.) But should warfighters be thought of primarily as private patients, or as defenders of society? If the latter, there are potential benefits of enhancement not only to the warfighter but also to the warfighter's unit, mission, service, and nation. One of the major challenges presented by military enhancement is determining how to balance the benefits to these third parties against the risk of harm to the warfighters themselves.

Another third variable is the legal status of the enhancement and the additional legal complexities that "military necessity" can create. Is the enhancement a drug, a medical device, or some other technology, such as a behavioral or psychological intervention? From a legal standpoint, these are subject to very different regulatory schemes, with drugs and devices governed by complex FDA rules and behavioral and psychological interventions essentially unregulated.

A product that the FDA has approved for enhancement use may be deemed to present fewer ethical concerns than a product that is

unapproved or still experimental, since approval bears on what is known about its safety and efficacy, as discussed later in connection with the administration of pyridostigmine bromide (PB), botulinum toxin (BT), and anthrax vaccine to combat troops during the Gulf War [Fitzpatrick and Zwanziger 2003].

Another important regulatory issue for enhancement drugs is whether they are controlled substances or otherwise illegal when used for non-medical, enhancement purposes. As noted earlier, the DoD prohibits the use of anabolic steroids by warfighters; accordingly, a commander who ordered her subordinates to take steroids would raise different ethical and legal concerns than one who ordered them to use a product that was not known for being subject to abuse, such as modafinil. Another variable is whether the enhancement product is supplied by the military or, like many dietary supplements, purchased privately by the warfighter. In the former case, the issue is whether warfighters need protection from their superiors; in the latter, the issue is whether the military ought to protect warfighters from their own poor choices.

A fourth variable that bears on ethical and legal issues is the type of characteristic or set of characteristics sought to be enhanced. For example, a drug that altered certain aspects of a person's mental state, such as mental acuity, personality, or emotions, might be deemed more problematic than a drug that increased strength or endurance, on the theory that the mental enhancement was more likely to affect the person's sense of self. The same would be true for a drug that altered or blocked a person's memory or reduced their capacity to make moral judgments, which have been raised as objections to the proposed prophylactic use of beta block-

ers such as propranolol to prevent post-traumatic stress disorder (PTSD) in combat troops [Henry et al. 2007; Wolfendale 2008]. Finally, one of the most important factors is whether warfighters are serving as human subjects in formal research projects, in which case the warfighters may be entitled to refuse to participate as human research subjects, or if instead the warfighters are given the enhancement to enable them to carry out their mission more successfully in the course of deployment, in which case a refusal to cooperate is likely to be viewed as disobeying an order. As we will see later, the experience during the Gulf War demonstrates that the distinction is not always clear.

2.4 Technology Survey

To help guide our discussion and ensure it is grounded in reality and not science fiction, we need to take a closer look at the human enhancement projects recently or currently pursued by militaries worldwide. Inasmuch as the US military is the most transparent about its research projects as well as the most heavily invested, most but not all of our examples are projects based in US, drawn from open-source or unclassified information. These should provide a sufficient sense of the many kinds of human enhancement projects in progress, as we expect any foreign-based projects to be similar or at least not radically different.

This survey of projects can be presented in any number of sensible ways. We will conduct our brief visit through this landscape with the following taxonomy. These are not mutually exclusive categories, but they are meant to highlight key areas of focus and trends, sometimes overlapping: enhancements to (a) physical capabili-

ties, (b) cognitive capabilities, (c) the senses, and (d) metabolism. We will also survey (e) some dual-use research that could be used for enhancement, including relevant projects in basic science.

To be sure, many of these projects may not yet qualify as enhancements, as understood above. For instance, some are exoskeletons or other external equipment that are cumbersome or not yet of a size that can be closely and continually worn. But we include these in anticipation that such equipment could be scaled down and intimately connected with the human body that they would qualify as enhancements in the future. This anticipation is supported by current trends: we are seeing our everyday tools—computers, telephones, books, etc.—shrink and metamorphose into things that previous generations would not recognize, e.g., Google’s digital glasses; and we are often reminded that the mobile phones we carry today have more processing power than the Apollo rockets that flew to the moon.

Other projects we discuss here are not so much devices as they are unusual training techniques, e.g., exploiting neuroscience and other technologies much more advanced than typical exercise equipment or teaching methods. These blur the line between mere “natural” training and more advanced methods that plausibly count as enhancements. We won’t attempt to nail down a sharper definition than already provided above, and so we leave open the possibility that these and other projects may fall into the realm of enhancement, if not now then perhaps someday later.

Finally, the following is far from an exhaustive survey of projects but only a sampling of recent and current ones. Rather than a comprehensive

catalog of all such projects, the purpose here is merely to have specific enhancements in mind as we continue a discussion about their ethical, legal, and policy implications, as well as to convey the ambitious range of enhancements that militaries are pursuing.

A. *Physical capabilities*

- **Strength:** Several research organizations are developing exoskeletons to increase human strength and endurance, e.g., ability to carry payloads of 200 pounds and to sustain a run at 7-10 miles per hour. These include Lockheed Martin’s HULC, Raytheon’s XOS, UC Berkeley’s BLEEX, and other projects [UC Berkeley 2012; Lockheed Martin 2012]. Note again that anabolic steroids are prohibited by the US military as a method to enhance strength, unless prescribed by a physician [Ray 2012].
- **Mobility:** Similar to the above exoskeletons, other external devices can aid mobility. With military funding, University of West Florida researchers are developing an aquasuit that emulates the locomotion of dolphins, penguins and turtles [Sherer 2008]. In DARPA’s Reconfigurable Structures program, the Z-Man is another bio-inspired project that is developing Geckskin, an adhesive fabric that can enable humans to climb walls like geckos, spiders, and other animals do [DARPA 2012g]. In the same program, robotics work is advancing, including a model with a running speed of a cheetah [DARPA 2012a, 155; DARPA 2012c]. From Germany, the Gryphon wingsuit reimagines the parachute as a wearable and powered glider system [Fallon 2009].
- **Protection:** Body armor, from medieval to

modern times, are not enhancements but mere tools, by our understanding of “enhancement.” But again, as tools are closely held and integrated with the body, this determination becomes debatable. Researchers at BAE Systems and MIT’s Institute for Soldier Nanotechnologies are working on liquid armor as well as dynamic, flexible fabric that can stiffen into armor (or a splint, etc.) when needed [BAE 2012a; MIT 2005].

- **Virtual capabilities:** Perhaps the greatest protection against harm on the battlefield is to not be on the battlefield at all. Militaries use drones, missiles, and other weapons for this and other compelling benefits [Shane 2012]. These tools do not replace the human warfighter—they are not true proxies—but other projects may. Similar to the Hollywood movie of the same name, DARPA’s Avatar program seeks to “develop interfaces and algorithms to enable a soldier to effectively partner with a semi-autonomous bi-pedal machine and allow it to act as the soldier’s surrogate” [DARPA 2012a, 123]. Relatedly, DARPA’s Bits to Behavior via Brains (B3) program is investigating how virtual-world activities impact real-world neural mechanisms [DARPA 2012a, 45].

B. Cognitive capabilities

- **Awareness:** Related to virtual capabilities is the need for greater situational awareness, e.g., better communication, data integration from different sources, threat identification, coordinated efforts, and so on. The Future Combat Systems program had sought to develop an ambitiously networked and multi-purpose battlesuit, the Future Force
- **Warrior suit** [Hanlon 2004]; while that program was cancelled in 2009, the idea lives on in other projects. DARPA’s Cognitive Technology Threat Warning System (CT2WS) is a computer-assisted visual aid that instantly identifies threats that warfighters might only subconsciously see, given that only a fraction of our visual data is consciously registered [DARPA 2012a, 168]; and the agency’s Soldier Centric Imaging via Computational Cameras (SCENICC) seeks to develop electronic contact lenses that do similar work [DARPA 2012f]. (DARPA’s Neovision2 is a related project in computer vision, though designed to be mounted on vehicles, not to enhance human visual processing [DARPA 2012a, 169].)
- **Attention:** Caffeine has been long used as an attention stimulant, especially in war. In the American Civil War, the Union had cut off supply lines to coffee and tea to strategically deprive the South of caffeine [Brecher 1972]. In modern warfare, pilots and other combatants use amphetamines—sometimes called “go pills” by the US military—to increase focus, though to possible serious side-effects [Emonson and Vanderbeek 1995]. As a safer alternative, the US and other militaries are using or exploring the use of modafinil and other drugs, which are already used illicitly to enhance academic and workplace performance [Caldwell et al. 1999].
- **Memory:** Where DARPA’s Human Assisted Neural Devices program seeks to strengthen and restore memories [DARPA 2012a, 50], other research aims to produce drugs and treatments that can erase memories, e.g., horrific ones that cause post-traumatic stress disorder (PTSD) [Lehrer 2012].

- **Planning:** Artificial intelligence is enhancing decision-making and planning. DARPA's "Deep Green automatically infers the commander's intent and produces a plan from the commander's hand-drawn sketches to facilitate rapid option creation, and plan recognition and understanding capabilities ensure the commander's intent is fully represented in the system" [DARPA 2012a, 246].
- **Learning:** Learning that surpasses results from traditional methods may perhaps count as an enhancement, especially new techniques that exploit advances in neuroscience and cognitive psychology. This is the goal of several research programs at DARPA, such as: Neurotechnology for Intelligence Analysts, Accelerated Learning, Education Dominance, Augmented Cognition, and Training Superiority programs [DARPA 2012a, 163].
- **Language:** Real-time language translation would bridge a major communications gap with foreign-language speakers; this is the goal of DARPA's Boundless Operational Language Translation (BOLT), Robust Automatic Translation of Speech (RATS), TRANSTAC, and other programs [DARPA 2012a, 87]. Today, there's even mobile software for that, DARPA's TransTalk app [Pierce 2011]; and Google offers a consumer version with its Translate app, as well as potentially for its Project Glass, an augmented-reality digital display that is worn like eyeglasses [Gilbert 2012].
- **Communication:** DARPA's Joint Tactical Air Controller (JTAC; formerly, Joint Terminal Attack Controller), or hologram goggles, also resembles eyeglasses but, instead of

language translations, facilitates direct communication between pilot and aircraft [DARPA 2012a, 205]. In the UK, BAE Systems' Q-Sight is a flight helmet that enhances situational awareness as well as control of the aircraft, including targeting through eye movements [BAE 2012b]. Some projects seek to enable communication through thought alone, such as the brain-computer interface work—or "synthetic telepathy"—funded by the US Army Research Office, in collaboration with University of California at Irvine, Carnegie Mellon University, and University of Maryland [Bland 2008].

C. Human senses

To add more nuance and details to this survey of enhancement projects, we can also consider cognitive-related projects as enhancements to particular senses; as such, they all increase situational awareness.

- **Sight:** DARPA's Functional Materials program aims to develop telescoping contact lenses [DARPA 2012a, 156]. We already mentioned DARPA's CT2WS, as well as the Google's Project Glass; the latter is not funded or related to military applications, but it may have dual-use applications (see below).
- **Smell:** Chemical sensors are increasingly important, since they can detect explosives and other materials. DARPA's RealNose project, for instance, seeks to mimic a dog's sense of smell (olfactory system) for "greater sensitivity to a wide range of new and old chemical agents and will serve to protect troops and infrastructure" [DARPA 2012h].

- **Hearing:** Defence Research and Development Canada seeks to develop an electronic pass-through hearing protection (EPHP) that filters out environmental noises while enhancing verbal signals [Burrell and Abel 2009]. In Scotland, similar projects aim to enhance acoustic speech with cochlear implants. Scientists in China, too, are working on technologies to enhance speech [Shao and Chang 2007].
- **Touch:** Defence Research and Development Canada is also developing a tactile cueing system for pilots to detect motion without visual or auditory cues [Bouak, Kline, and Cheung 2011]. Apart from military research, citizen-scientists and artists have been experimenting with magnetic implants in their bodies to detect electromagnetic fields, such as those created by electronic devices [Popper 2012; Ungerleider 2012; Cole 2012]; these are potentially dual-use technologies as well (see below).
- **Taste:** Insofar as smell is closely related to taste, enhancements to the olfactory system (smell) could be considered as enhancements to the gustatory system (taste). “Electronic tongues” can identify flavors, bitterness, spiciness, saltiness, sweetness, and other tastes. Moreover, the US military is interested in using the human tongue to sense or “see” the surrounding environment, as sonars can, for 360-degree “vision” even at night [Associated Press 2006].

D. Human metabolism

To add more nuance and details to this survey of enhancement projects, we can also consider physical- and cognitive-boosting projects as

metabolic enhancements, in case the distinction is relevant.

- **Endurance:** DARPA’s Metabolic Dominance program—later called Peak Soldier Performance—includes many umbrella goals, one of which is to boost human endurance, both physical and cognitive [Shachtman 2007c]. Stanford researchers, for instance, have developed CoreControl, a glove that regulates body temperature to enhance resistance to heat and cold, thus increasing endurance and preventing the associated loss of cognitive functions [Shachtman 2007a]. Dietary supplements, such as quercetin, are also being investigated for cognitive-enhancing effects under stress [Priprem et al. 2008].
- **Food:** Relatedly, US and UK scientists are researching genetic and cellular (mitochondrial) enhancements to enable soldiers to run for long distances and to survive longer without food, e.g., as Alaskan sled dogs are able [Alexander 2010]. As a substitute for human food, DARPA’s Crystalline Cellulose Conversion to Glucose (C3G) program could someday enable warfighters to eat otherwise indigestible materials, such as grass [DARPA 2012i].
- **Sleep:** Besides the basic need for food, our need for sleep is a severe limiting factor in battlefield performance; and this is why alertness drugs are much sought after. DARPA-funded researchers are also investigating light and magnetic therapies to safely maintain wakefulness [Shachtman 2007b]. Nature already provides proofs-of-concept that mammals can operate without much sleep: whales and dolphins would drown if they were to truly sleep, but instead

only half of their brains sleeps at a time [Hecker 1998]; and giraffes reportedly need less than two hours of sleep a day, taking highly efficient “power naps” throughout the day [Tobler and Schwierin 1996].

- **Health:** Though we had defined enhancement as more than restoring health, some technologies or treatments that accelerate healing are arguably enhancements, given the speed of regeneration [Wang 2008]. That is, a difference of degree may become a difference in kind. Other research aims at blocking pain [Wang 2008].

E. Dual-use research

The following is a sample list of military-funded research projects in therapeutics or healing, but they may also have dual-use applications as enhancements. The basic-science projects we include here also could have either therapeutic or enhancement uses, or both.

- **Stress:** Enabling Stress Resistance; Neuroscience Technologies; Detection and Computational Analysis of Psychological Signals (DCAPS; formerly called Healing Heroes); US Army’s investigation of selective serotonin reuptake inhibitors (SSRI) as treatment for PTSD [McKeon et al. 2009; DARPA 2012d; DARPA 2012a, 63].
- **Circulatory:** Blood Pharming; Tactical Biomedical Technologies [DARPA 2012a, 165, 169].
- **Metabolism:** Metabolic Flexibility and Suspended Animation, e.g., hibernation [Roth Lab 2012].
- **Environmental:** Maintaining Combat Perfor-

mance; Warrior Web; Rapid Altitude and Hypoxia Acclimation (RAHA) [DARPA 2012a, 162; DARPA 2012e]; thermotolerance research by US Army and others [Carter and Calais 2009].

- **Toxins and radiation:** US Army’s protective bioscavenger research (human serum butyrylcholinesterase) [Saxena et al. 2009]; radioresistance research by Bulgaria’s Military Medical Academy and others [Kindekov et al. 2009].
- **Prosthetics:** Revolutionizing Prosthetics; Reliable Neural Interface Technology (RENET) [DARPA 2012a, 171].
- **Diagnostics:** Autonomous Diagnostics to Enable Prevention and Therapeutics (ADEPT; formerly part of Synthetic Biology) [DARPA 2012a, 45; 2012c]; HEADS helmet research by UK’s BAE Systems [BAE 2012c].
- **Drug delivery system:** Feedback Regulated Automatic Molecular Release (FRAMR) [DARPA 2012j].
- **Basic science:** BioChronicity; Unconventional Therapeutics; In Vivo Nanoplatforms; Biological Adaptation, Assembly and Manufacturing; BioDesign; Living Foundries [DARPA 2012b; DARPA 2012a, 196, 4, 165, 166].
- **Non-military:** Not all research projects related to human enhancement are funded by the military, though many are. For instance, we mentioned Google’s Project Glass previously, as well as citizen-scientists and artists—i.e., “biohackers”—who conduct their own enhancement research, often on their own bodies. Life-extension is a major

area of civilian enhancement research [de Grey and Rae 2007], though surprisingly we have not seen the military take much interest here. As such, we do not engage a suite of ethical questions related to the quest for longer life [Williams 1973; Garreau 2006; Harris 2007; Allhoff et al. 2010a].

3. Law and Policy

With that background, we begin with a discussion of the primary legal issues—in both international law and US domestic law—that are relevant to military enhancements, as well as possible implications on military operations. To the extent that ethics underwrites law and policy, we can better understand the former by looking at the latter as the real-world implementation of ethics. In later sections, we will focus more on an ethical analysis, as distinct from a legal one.

3.1 International Law

What are the provisions in international law that may bear upon military human enhancements? Should enhancement technologies, which typically do not directly interact with anyone other than the human subject, be subject nevertheless to a weapons legal-review? That is, is there a sense in which enhancements could be considered as “weapons” and therefore subject to legal instruments such as the Biological and Toxin Weapons Convention? How do norms related to human-subject research and medical ethics impact military enhancements?

These are some of the most important questions for military enhancements as they relate to international law [Lin 2012a]. Conceptually, we will divide international law into two categories: the first is international humanitarian law (IHL), also known as the laws of armed conflict (LOAC), and the second is composed of international agreements related to biomedical research. Because these are well-discussed conventions,

we will only list them here and discuss them later in more detail as needed.²

Under IHL, the main instruments of interest here are:

- Hague Conventions (1899 and 1907)
- Geneva Conventions (1949 and Additional Protocols I, II, and III)
- Biological and Toxin Weapons Convention (1972)
- Chemical Weapons Convention (1993)
- Rome Statute of the International Criminal Court (1998)

Under international biomedical laws—which we discuss more in the next section—the main instruments of interest here are:

- Nuremberg Code (1947)
- Declaration of Geneva (1948)
- Declaration of Helsinki (1964)

As it concerns new technologies, Article 36 of the Geneva Conventions, Additional Protocol I, specifies that:

In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party

.....
² Though not a signatory to some of these conventions, the US nevertheless has an interest to stay within international norms, e.g., to not trigger international condemnation.

is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party [1977].

But does Article 36 apply to human enhancement technologies? That is, should they be considered as a “weapon” or “means or method of warfare” in the first place? Unlike other weapons contemplated by IHL, enhancements usually do not directly harm others, so it is not obvious that Article 36 of Additional Protocol I would apply here. If anyone’s safety is immediately at risk, it would seem to be that of the individual warfighter, thereby turning the debate into one about bioethics. To that extent, warfighters, whether enhanced or not, are not weapons as typically understood.

Yet in a broader sense, the warfighter is not only a weapon but perhaps a military’s best and oldest weapon. Warfighters carry out missions, they sometimes kill enemies, and they represent one of the largest expenditures or investments of a military. They have cognitive and physical capabilities that no other technology currently has, and this can make them ethical, lethal, and versatile. The human fighter, engaged in hand-to-hand combat, would be the last remaining weapon when all others have been exhausted. So in this basic sense, the warfighter is undeniably a weapon or instrument of war.

Still, should Article 36 be interpreted to include warfighters themselves as weapons subject to regulation? There could be several reasons to think so. First, other organisms are plausibly weapons subject to an Article 36 review. Throughout history, humans have employed animals in the service of war, such as dogs, elephants, pigeons, sea lions, dolphins, and

possibly rhinoceroses [Knights 2007; Beckhusen 2012; US Navy 2012]. Dogs, as the most commonly used animal, undergo rigorous training, validation, and inspections [US Department of the Army 2005]. If a military were to field a weaponized rhino in an urban battlefield that contains innocent civilians, we would be reasonably worried that the war-rhino does not comply with Article 36, if rhinos cannot reliably discriminate friends from foe, e.g., a rhino may target and charge a noncombatant child in violation of the principle of distinction. A similar charge would apply to autonomous robots in such a general environment in which distinction is important, as opposed to a “kill box” or area of such fierce fighting that all noncombatants can be presumed to have fled [Lin et al. 2008].

If autonomous robots are clearly regulatable weapons, then consider the spectrum of cyborgs—part-human, part-machine—that exists between robots and unenhanced humans. Replacing one body part, say a human knee, with a robotic part starts us on the cybernetic path. And as other body parts are replaced, the organism becomes less human and more robotic. Finally, after (hypothetically) replacing every body part, including the brain, the organism is entirely robotic with no trace of the original human. If we want to say that robots are weapons but humans are not, then we would be challenged to identify the point on that spectrum at which the human becomes a robot or a weapon.

The inability to draw such a line may not be a fatal blow to the claim that humans should be treated as weapons; after all, we cannot draw a precise line at which a man who is losing his hair becomes “bald”, yet there’s clearly a difference between a bald man and one who has a head full of hair [Stanford 2011]. But a simpler solution

may be to say that humans are weapons, especially given the reasons offered previously.

As it applies to military enhancements, integrated robotics may be one form of enhancement, but we can also consider scenarios involving biomedical enhancements such as pharmaceuticals and genetic engineering. Again, on one end of the spectrum would stand a normal, unenhanced human. One step toward the path of being fully enhanced may be a warfighter who drinks coffee or pops amphetamines (“go pills”) as a cognitive stimulant or enhancer. Another step may be taking drugs that increase strength, erase fear, or eliminate the need for sleep. At the far, more radical end may be a warfighter so enhanced that s/he no longer resembles a human being, such as a creature with four muscular arms, fangs, fur, and other animal-like features. If a war-rhino should be subject to Article 36, then so should this radically enhanced human animal, so it would seem. And to avoid the difficult question of drawing the line at which the enhanced human becomes a weapon, a more intuitive position would be that the human animal is a weapon all along, at every point in the spectrum, especially given the previous reasons that are independent of this demarcation problem.

If we agree that enhanced human warfighters could properly be weapons subject to Article 36, what are the implications? Historically, new weapons and tactics needed to conform to at least the following:

- Principle of distinction
- Principle of proportionality
- Prohibition on superfluous injury or unnecessary suffering (SirUS)

To explain, first, the principle of distinction demands that a weapon must be discriminating enough to target only combatants and never noncombatants [Geneva Additional Protocol I 1977; Sassòli 2003]. Biological weapons and most anti-personnel landmines, then, are indiscriminate and therefore illegal in that they cannot distinguish whether they are about to infect or blow up a small child versus an enemy combatant. Unintended killings of noncombatants—or “collateral damage”—may be permissible, but not their deliberate targeting; but to the extent that biological weapons today target anyone, they also target everyone. (If they don’t target anyone in particular but still kill people, then immediately they would seem to be indiscriminate.) However, a future biological weapon, e.g., a virus that attacks only blue-eyed people or a certain DNA signature [Hessel et al. 2012], may be discriminate and therefore would not violate this principle (but it could violate others).

Second, the principle of proportionality demands that the use of a weapon be proportional to the military objective, so to keep civilian casualties to a minimum [Geneva Additional Protocol I 1977; Cohen 2010]. For instance, dropping a nuclear bomb to kill a hidden sniper would be a disproportionate use of force, since other less drastic methods could have been used.

Third, the SirUS principle is related to proportionality in that it requires methods of attack to be minimally harmful in rendering a warfighter *hors de combat* or unable to fight [Coupland and Herby 1999]. This prohibition has led to the ban of such weapons as poison, exploding bullets, and blinding lasers, which cause more injury or suffering than needed to neutralize a combatant.

However implausible, we can imagine a human enhancement that violates these and other provisions—for instance, a hypothetical “berserker” drug would likely be illegal if it causes the warfighter to be inhumanely vicious, aggressive, and indiscriminate in his attacks, potentially killing children. (For the moment, we will put aside enhancements that are directed at adversaries, such as a mood-enhancing gas to pacify a riotous crowd and a truth-enhancing serum used in interrogations; the former would be prohibited outright by the Chemical Weapons Convention in warfare [The Royal Society 2012], partly because it is indiscriminate, and the latter may be prohibited by laws against torturing and mistreating prisoners of war.) The point here is that it is theoretically possible, even if unlikely, for a human enhancement to be in clear violation of IHL.

But let us assume that the human enhancement technologies generally conform to these basic principles. (If they do not, then there’s already strong *prima facie* reason to reject those technologies as unlawful under IHL; those are the easy cases that do not need to be examined here.) Given this assumption, are there other, less-obvious international laws that could prohibit military enhancements? We will discuss three possible areas of concern:

A. Biological Weapons

First, the opening discussion of this report’s section on whether enhancements are weapons is relevant not only to Article 36 of Additional Protocol I but also arguably to the Biological and Toxin Weapons Convention (BTWC). The first article of the BTWC states that:

Each State Party to this Convention undertakes never in any circumstances to develop,

produce, stockpile or otherwise acquire or retain: (1) microbial or **other biological agents**, or toxins whatever their origin or method of production, of types and in quantities that have no justification for prophylactic, protective or other peaceful purposes; (2) weapons, equipment or means of delivery designed to use such agents or toxins for hostile purposes or in armed conflict [1972; emphasis added].

Whether or not they are properly weapons, are military human enhancements “biological agents” in any reasonable sense? The BTWC is silent on this question, though it does anticipate unforeseen developments in genetic engineering, biotechnology, synthetic biology, and other scientific fields [BTWC 1972, Additional Understandings of Article I]. The usual assumption is that these “agents” are both limited to roughly being microbial in size and to biological substances that are directed at adversaries, not directed to the enhancement of one’s own military personnel. This assumption, unfortunately, is not explicit enough in the BTWC; that is, it does not define what a biological agent is. As a result, it is still an open question of whether the BTWC applies to human enhancement technologies.

To answer this open question, let’s try to better understand what a “biological agent” is. This seems to mean an agent that is biological in nature (e.g., anthrax virus), as opposed to purely chemical (e.g., chlorine gas) or physical (e.g., a falling object); and an agent is a substance or actor employed for some effect or purpose (e.g., LSD is a psychotropic agent). But in a broader but consistent sense, agents can be persons too (e.g., a government spy is a “secret agent”). If so, then enhanced warfighters can be agents. Even if we reject this understanding and stipulate

that biological agents must be nonperson substances—an interpretation that is not explicit in the BTWC—we can still consider the enhancement technology itself as an agent, apart from the warfighter it enhances.

Again, insofar as the BTWC does not specify that biological agents must be of the kind that directly harms adversaries, then some human enhancements—such as anabolic steroids for increased strength—would seem to count as biological agents: they are substances employed for some effect and are biological in nature. They would serve “hostile purposes” in that they create a warfighter more capable of defeating adversaries and fulfilling military missions; so these enhancements would at least indirectly harm others.

With respect to scale, it is difficult to see why size would matter for the BTWC, which again is not explicit on the issue. If we understand the BTWC to be interested in only microbial-sized agents—and returning to the position that humans can be agents—then consider a hypothetical process that can shrink a human soldier to the size of bacteria, such as in the theatrical film *Fantastic Voyage* [Shickel 1966; Internet Movie Database 2012]: if size matters, then the BTWC would seek to regulate the microscopic soldier, but not the full-sized soldier who has the exact same capabilities. Why the difference in concern here? It may be that the microscopic soldier can be stealthier, infiltrate more places, and so on, but none of these concerns is cited in the BTWC as a motivating reason for regulation.

Related to enhancements, the BTWC arguably would have something to say about bioengineered insects and animals, for instance, that are used as weapons. Like pathogens, insects and most animals do not obey human orders and

would therefore be unpredictable and indiscriminate as a weapon; and tiny attack-insects do not seem significantly different in kind than microscopic organisms also designed for attack. One possible difference is that microorganisms typically harm us from the inside-out, and somehow this could be less humane and more frightening than biting us or attacking our bodies from outside-in. Yet we can also envision bioengineered animals that operate from the inside-out too, as tapeworms and mosquitoes do (at least the disease they transmit into our bloodstreams). So if it’s not unreasonable to think that bioengineered insects would be subject to the BTWC, then size does not matter for the BTWC, or at least the interest is not limited to microscopic organisms.

As for other qualifiers in the BTWC, some enhancements could be noncompliant in that they have no “prophylactic, protective or other peaceful purposes” [BTWC 1972, Article I.1]. A hypothetical berserker drug could be an example: its only obvious function is to make a person a fiercer, rampaging combatant. This is to say that, under some plausible understanding of the BTWC, at least some possible warfighter enhancements could count as “biological agents” and therefore subject to the BTWC. If the BTWC intends or ought to rule out enhancements under its purview, then its language needs to be made more explicit.

B. Inhumane Weapons

Contributing to the above problem with the BTWC—i.e., what counts as a “biological agent”—is also a lack of specificity on the motivating reasons for the BTWC in the first place. That is, the convention is unclear on *why* we should want to prohibit biological and toxin weapons. But there are some clues. In the

preamble to the BTWC, state parties to the convention declare they are:

Convinced of the importance and urgency of eliminating from the arsenals of States, through effective measures, such dangerous **weapons of mass destruction** as those using chemical or bacteriological (biological) agents...

Convinced that such use would be **repugnant to the conscience of mankind** and that no effort should be spared to minimize this risk [1972; emphasis added].

That is, biological agents, such as highly infectious bacteria or viruses, are difficult to control in their propagation and therefore are indiscriminate to use as a weapon. Anthrax spores, for instance, may be carried by the wind and can infect a child or entire populations just as easily and likely as a soldier. This would be a clear violation of the principle of distinction in IHL.

If this were the only motivating reason for the BTWC, then perhaps we can conclude that human enhancements are not the biological agents that the convention intends to address; enhancements are generally not infectious or “weapons of mass destruction.” But this cannot be the only reason. In its categorical prohibition of biological and toxic weapons, the BTWC does not distinguish between infectious and non-infectious ones. For instance, a poison dart that can be used only once in a precisely targeted attack would still be banned, even though it is not a weapon of mass destruction, given that it is a toxin and especially if there were no “prophylactic, protective or other peaceful purposes” for the poison.

To explain why the prohibition is categorical, we can examine the next clue, that the BTWC is motivated by “the conscience” of humanity. That is, some methods of killing are more insidious and repugnant than others. Biological and toxin weapons, then, are of special concern, because they are usually silent, invisible, and indiscriminate ways of killing people—often with horrific, painful medical symptoms over the course of several days or weeks.

But is any of this relevant to human enhancements? Again, enhancements usually do not directly harm others, much less kill people in “repugnant” ways. Even if we say that enhancements indirectly harm others, they do not typically do so in ways more repugnant than conventional means, since an enhanced warfighter is still bound by IHL to never use certain weapons and tactics against adversaries.

Like the “weapons of mass destruction” clue, that a biological agent is “repugnant to the conscience of mankind” also does not seem to be a necessary requirement, just a sufficient one. Consider that some poisons or pathogens may kill quickly and painlessly, such as those administered in death-penalty executions: they seem to be much more humane than conventional means, such as shooting bullets and dropping bombs that render an adversary *hors de combat* through massive, bloody injury to human bodies and brains. Nevertheless, these “clean” poisons are prohibited by the BTWC and elsewhere, such as the Hague Conventions. So, even if human enhancements are not repugnant in the same ways that anthrax or arsenic may be, and even if they are not weapons of mass destructions, they could still fall under the authority of the BTWC, again since the convention is not explicit on its motivating reasons.

In any event, enhancements could be repugnant in different ways. We previously mentioned the possibility of creating a “berserker” drug, as well as a warfighter so enhanced that s/he no longer resembles a human being, such as a creature with four muscular arms, fangs, fur, and other animal-like features. If this sounds far-fetched, we need only look at the history of warfare to see that intimidating adversaries is a usual part of warfare. From fierce Viking helmets, to samurai armor designed to resemble demons, to tigers and sharks painted onto warplanes, to ominous names for drones such as “Predator” and “Reaper”, scaring adversaries can demoralize and make them easier to defeat. This suggests that it may not be so irrational nor inconsistent with customary practices to design enhancements to be inhuman and therefore perhaps inhumane.

Further, biomedical research is presently ongoing with “chimeras”, or animals composed of genes or cells from other organisms not involved with the reproduction of those animals. These may include animals created with human genes, for instance, in order to grow transplantable organs *in vivo* and for research to find medical cures [Ekser 2012]. Manipulation of human embryos, too, can lead to human-animal chimeras, though this possibility has caused much ethical concern and debate [Greely 2003; di Melo-Martin 2008], so much so that US legislation had been proposed to prohibit this line of research, calling it an affront to human dignity as well as an existential threat [Human Chimera Prohibition Act 2005].

Not all enhancements, of course, are as fanciful as a human-chimeric warrior or a berserker mode, nor are we suggesting that any military has plans to do anything that extreme. So most, if not all, enhancements will likely not be as obviously inhuman. Nonetheless, the “con-

sciousness of mankind” is sometimes deeply fragmented, especially on ethical issues. So what is unobjectionable to one person or culture may be obviously objectionable to another. Something as ordinary as, say, a bionic limb or exoskeleton could be viewed as unethical by cultures that reject technology or such manipulation of the human body. This is not to say that ethics is subjective and we can never resolve this debate, but only that the ethics of military enhancements—at least with respect to the prohibition against inhumane weapons—requires specific details about the enhancement and its use, as well as the sensibilities of the adversary and international community. That is, we cannot generalize that all military enhancements either comply or do not comply with this prohibition.

Beyond the BTWC, inhumanity as a prohibitory reason is a common theme that underlies IHL. In the preamble to the first Hague Convention:

Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the **laws of humanity** and the **requirements of public conscience** [1899; emphasis added].

Known as “the Martens Clause”, this basic principle is found throughout the laws of armed conflict, such as the Geneva Conventions and its Additional Protocols and opinions issued by the International Court of Justice [Geneva Additional Protocol I 1977; Ticehurst 1997]. As one would expect, much debate has occurred on what the “laws of humanity” and “requirements of public

conscience” are, especially related to the actual or even threatened use of nuclear weapons. And the same debate could be applied to emerging technologies, from attack drones [Human Rights Watch 2012] to human enhancements.

We will not engage that lengthy and unresolved debate here, except to note that a prohibition against inhumane weapons and methods is a fundamental principle, sometimes explicit and sometimes implied, that underwrites the laws of war and therefore relevant to an ethics assessment of military enhancements. This is also to say that an ethics assessment of new weapons, such as military enhancements—the purpose of this report—seems to be legally required by IHL, at least in the context of the Martens Clause if not also Article 36 of the Geneva Conventions, Additional Protocol I.

C. *Inhumane Treatment*

The concept of inhumanity is important to clarify, not just for the legal evaluation of weapons but also for the ethical limits on how combatants may be treated. The prohibition on torture [Universal Declaration of Human Rights 1948, Article 5; Reidy 2003; Garcia 2008], for instance, presumes certain facts about the human condition, such as the kinds of treatment that cause pain, how much pain a person can withstand, how much sleep a person needs, and so on. For instance, if our tolerance for pain were dramatically elevated, then what used to count as torture yesterday may no longer be so torturous today, and therefore such behavior may now be morally permissible.

More generally, ethics itself also presumes a similar set of facts about the human condition, for instance, that we are fairly susceptible to being killed. These facts inform our ethics, for in-

stance, when self-sacrifice is permitted or prohibited and, again, what kinds of action toward others are unethical. If we change these presumed facts about human bodies and minds, then ethical prohibitions and permissions may also be affected. This gives us reasons to believe that an ethical code of behavior for robots could very well be different from how humans ought to behave; for instance, robots—to the extent that they have no instinct for self-preservation, cannot feel pain, etc.—may be permitted to sacrifice themselves in more trivial scenarios than human ethics might allow [Lin, Bekey, and Abney 2008; Ingram and Jones 2010].

At the beginning of this report’s section, we suggested that there is a continuum from a fully human animal to a cybernetic organism to a fully robotic machine. This spectrum is perhaps defined by how many human body parts we replace with mechanical ones, ranging from zero to all. Enhanced warfighters, then, could fall somewhere in the middle of this continuum. If “robot ethics” is different from human ethics, at least where relevant facts about humans and robots differ, then it seems that “cyborg ethics” too would diverge from human ethics where there’s a relevant difference in the construction and abilities between cyborgs and humans. Though not all enhanced persons are cyborgs, e.g., if the enhancements are genetic, pharmacological, or otherwise not robotic, we can also reasonably conclude that ethics for enhanced persons generally may be different from the standard human ethics.

So it becomes an interesting question of whether it would still be illegal or inhumane to whip a prisoner of war, or deprive him of food or sleep, if the individual can better withstand a whipping or does not have the same food or sleep requirements that normal people typically do.

These actions possibly would not cause pain or suffering, or at least as much of it, to the enhanced subject; therefore, it would be difficult to count those actions as torture.

Beyond prisoners of war, questions about inhumane treatment could be directed at how we treat our own enhanced warfighters. For instance, drill sergeants may be tempted to push an enhanced soldier harder than other ones without augmented strength and endurance, and perhaps reasonably so. But where there are prohibitions on what military trainers are permitted to do, we may need to reevaluate those rules where an enhancement might change the pre-suppositions about human limits that motivated those rules in the first place.

3.2 US Domestic Law³

The international law considered above speaks primarily to what militaries can do with their enhanced warfighters, but there is also a prior question of whether militaries are permitted to enhance their personnel in the first place. This is more a question for bioethics and related law than for the considered IHL. Here, we will briefly outline some of the key US domestic laws and regulations that would apply to military enhancements. In section 4, we will draw upon this background to more fully discuss the international standards in bioethics previously mentioned.

But here we ask, what does US domestic law say about requiring enhancements for our own

³ We thank Michael Burnam-Fink, Alexander R. LaCroix, and Seth G. Schuknecht for their discussion in this section.

military personnel? To answer that question, we can look at actual legal cases in the US that are closely related, if not directly about, human enhancements. While we had excluded vaccinations as a type of human enhancement in the definitional section of this report above—because they are designed to sustain health, not provide capabilities beyond it—we also acknowledged that this understanding was contentious: in some sense, a vaccination seems to be an enhancement of the immune system, especially considering that the patient is not sick at the time of the immunization. At the least, even if not enhancements themselves, vaccinations seem to be closely related and can inform a study on how US law might deal with military enhancements.

The US military has been vaccinating troops since 1777 [US Department of Defense 2011]. There are currently thirteen vaccinations used by the military mandated for trainees alone: mandatory vaccinations include influenza, hepatitis A and B, measles, poliovirus, rubella, and yellow fever, among others [Grabenstein et al. 2006]. The standard military policy for the mandatory administration of pharmaceutical agents is the same as the policy applied to civilians [Russo 2007]: pharmaceuticals need to be approved by the US Food and Drug Administration (FDA) for their intended use before they are mandatorily administered; and absent FDA approval, a Presidential waiver or informed voluntary consent must be obtained for the administration of an investigational drug (IND) [Russo 2007]. The US Supreme Court has held that mandatory vaccinations of FDA-approved drugs do not violate the US Constitution [Jacobson v. Commonwealth of Massachusetts 1905, 18-19]. Mandatory vaccination programs in the military have been challenged in court [United States v. Chadwell 1965], but they were rarely subjected

to substantial legal challenges until 2001, directed at the Anthrax Vaccine Immunization Program (AVIP).

A. Federal Law

As an important catalyst for US law related to vaccinations, AVIP—established in 1997—had roots in Operation Desert Shield in 1990, at which time the US military worried about biological and chemical weapons that Saddam Hussein was rumored to have possessed. At the time, the DoD argued that the informed consent requirement for the administration of INDs was impractical [*Doe v. Sullivan* 1991]. The requirement was feasible during peacetime, but the DoD urged that it posed significant obstacles to the safety of troops and mission accomplishment in wartime [Brown 2006]. In response to pressure from the DoD, the FDA promulgated Rule 23(d), otherwise known as the Interim Final Rule:

i. 21 CFR 50.23(d), or Interim Final Rule

Rule 23(d) allows the DoD to waive the informed consent requirement, if it is not feasible to obtain consent in a particular military operation, subject to conditions [Brown 2006]. Most importantly, the waiver must be limited to “a specific military operation involving combat or the immediate threat of combat” [*Doe v. Sullivan* 1991, 1374]. Upon receiving the request for waiver from the DoD, the FDA must evaluate it and grant the waiver “only when withholding treatment would be contrary to the best interests of military personnel and there is no available satisfactory alternative therapy” [*Doe v. Sullivan* 1991, 1374]. This rule was challenged in 1991, in *Doe v. Sullivan*, but the federal court held that 23(d) was constitutional and within the scope of the FDA’s authority [*Doe v. Sullivan* 1991, 1381].

ii. 10 USC §1107(f)

In 1998, in response to the ruling in *Doe v. Sullivan*, the US Congress enacted 1107(f). This statutory provision requires the DoD to obtain informed consent from soldiers before administering an IND (including an approved drug for an unapproved use) and provides that the President can waive said requirement [10 USCA § 1107 (West)].

iii. Executive Order 13139

President Clinton unified both rule 23(d) and 1107(f) in 1999 with Executive Order 13139, a guideline for waiving informed consent within the context of military operations [Brown 2006, 942]. According to the order, to use an “investigational drug” or a “drug unapproved for its intended use,” the Secretary of Defense must obtain informed consent from each individual service member [Executive Order No. 13139 1999]. However, a Presidential waiver can overcome this requirement, but it can only be obtained upon a written determination that obtaining consent is: (1) not feasible; (2) contrary to the best interests of the member; or (3) is not in the interests of national security [Executive Order No. 13139 1999].

iv. DoDD 6200.2

The Department of Defense Directive (DoDD) 6200.2, like Executive Order 13139, synthesized several sources of authority governing the use of INDs for military health protection [US Dept. of Defense 2000]. It defines an IND as a “drug not approved or a biological product not licensed by the FDA,” or alternatively, as a “drug unapproved for its applied use” [US Dept. of Defense 2000, 2]. Further, it provides that the DoD must prefer

products approved by the FDA for use as countermeasures over INDs [US Dept. of Defense 2000, 3]. However, “when, at the time of the need for a force health care protection countermeasure against a particular threat, no safe and effective FDA-approved drug or biological product is available, DoD Components may request approval of the Secretary of Defense to use an investigational new drug” [US Dept. of Defense 2000, 3]. If the Secretary of Defense determines that obtaining informed consent is not feasible, contrary to the best interests of the member, and is not in the interests of national security, s/he can then request a waiver from the President [US Dept. of Defense 2000, 4].

B. Military Law

Military law operates in conjunction with federal civil law, but it focuses on matters germane to the military alone. In addition to the Constitution, US military law is governed by the Uniform Code of Military Justice (UCMJ). In the context of military vaccinations, the issue is about the lawfulness of the order to take the vaccination. The DoD’s successful defense strategy of the legality of the AVIP throughout the anthrax cases was straightforward on this account: the vaccine was determined by the FDA to be safe and effective for use against inhalation anthrax, and under military law the legality of an order to take the vaccine was a question of law for a judge to decide, not a question of fact for determination by a jury [Katz 2001].

Under the UCMJ, disobedience of a direct and lawful order from a superior officer is punishable under articles 90 or 92. Article 90 prohibits willfully disobeying a superior commissioned officer [10 USC § 890 (1994) (UCMJ Art. 90)], and article 92 prohibits failing to obey an order or regulation [10 USC § 890 (1994) (UCMJ Art. 92)]. A soldier

who refused to take the anthrax vaccination was court-martialed, where the DoD would file two interlocutory motions: (1) that the lawfulness of the order should be decided as a question of law; and (2) that all the evidence regarding the safety, efficacy, and necessity of the vaccine should be excluded because the legal authority of an order is not based on the safety of the vaccine [Ponder v. Stone 2000]. The DoD did this in every challenge to the AVIP, and in every challenge to the AVIP in military court they were successful [Katz 2001].

A strong, but rebuttable, presumption is that a military order is lawful when someone is charged with willful disobedience of a lawful order [US Government, Manual for Courts-Martial 2010; Katz 2001], and the lawfulness of a military order is an interlocutory order to be decided on by a judge, not a jury [US v. New 1999; Perry v. Wesely 2000]. What this effectively does is foreclose a legal challenge to the scientific efficacy of a vaccine on procedural grounds.

We will not delve much into the legal challenges to AVIP specifically, except to use the preceding as background as we discuss the different bioethics models in the next section. Again, while these legal issues were involved with actual cases involving vaccinations, we can plausibly extend them to anticipate how they would address technologies and procedures that are more clearly human enhancement than therapy.

3.3 Operations

Beyond the demands of international and domestic law, military enhancements likely will have important policy implications. We will examine here some of those implications on military operations themselves. (The broader

impact of enhancements on society at large has been well discussed in literature elsewhere [Allhoff et al. 2010].) Cognitive and physical human enhancements can significantly help a military achieve its missions, operate more efficiently and perhaps ethically, as well as a host of other benefits. We won't explicate these, as their applications are fairly clear. Rather, we will focus on unintended problems that may be caused by enhancements.

A. *Morale and Unit Cohesion*

Assuming that enhancements are not adopted by all warfighters at once—for instance, they are rolled out selectively or slowly for safety, economic, or other reasons—there would instantly be an inequality among the ranks. Some warfighters will be privileged (or unlucky?) to be enhanced, while others remain “normal.” In broader society, we see that uneven access to technology creates a gap between the haves and the have-nots, such as the Internet divide [Rozner 1998]; and this translates into a difference in quality of life, education, earnings, and so on. It is therefore not unreasonable to expect a similar effect within the military.

At the unit level, enhancements may increase any dissension between warfighters. A mix of enhanced and unenhanced warfighters within a single unit may affect morale and unit cohesion. To be sure, similar worries had been voiced related to the integration of different ethnic groups, religions, and sexual orientation in the military [Canaday 2001]; but where these differences do not intrinsically imply different levels of capabilities or merit that would matter operationally, human enhancements do.

By definition, an enhanced warfighter would be stronger, faster, or otherwise better-abled than

normal counterparts. This means they could accept riskier roles and have lower support requirements, for instance. Further, because enhanced warfighters represent a significant investment of research and effort, they may not be subjected to the hard work of fighting or other “mundane” uses. Compare this to Allied airborne troops in World War II who were pulled from the lines after the D-Day invasion of Normandy, rather than being required to slog through France and the Hürtgen Forest in Germany [O'Meara 2012b].

The asymmetry of needs and capabilities could cause resentment of the unenhanced as a drag on capabilities and operational efficiency of the enhanced, as well as resentment by the unenhanced of the superior abilities and (likely) superior status of the enhanced. To some extent, we already witness this when militaries switch their dependence from soldiers to “special operators” such as Navy SEALs [O'Meara 2012b]. The asymmetry also could create a sense of entitlement among the enhanced and undermine an *esprit de corps*, much as some superstars do on sports teams.

B. *Command*

Morale is relevant to confidence in command. Thus enhancements could create novel difficulties for the command structure, particularly if commanders were unenhanced and were seen as physically—or, worse, intellectually—inferior to those they command.

As one firsthand perspective, according to retired US Army Brigadier General Richard O'Meara, a social contract exists between troops and leaders that places the burden of defining the goals of a mission on the leaders and the burden of accomplishment on the troops [O'Meara 2012a,

2012b]. But while the troops have the responsibility to accomplish goals which command has set forth, they also have a right to demand that leaders make informed decisions, even if difficult ones, and to do so in a way that warfighters recognize as legitimate. It is a recipe for disaster when those further down the chain of command are continually second-guessing and evading their orders. If human enhancement exacerbates that lack of confidence in leaders, it could undermine the strategy and tactics of command.

Physical enhancements may be less problematic in this regard than cognitive ones, at least with respect to challenges to command. When the troops are generally less educated, less interested in strategy, and more concerned with communal rather than individual rights and values, command can worry less about the potential disobedience that could result from enhancements. According to O’Meara [2012b], the primary responsibility of typical enlisted soldiers is to know at all times what their superiors desire of them; their well being, even their survival, may well depend on it. Therefore, the rank-and-file are typically extremely sensitive to the wishes of command and, even when those wishes are not officially communicated, there is an expectation that a soldier will “get it” and learn to read the signs and comply, or disregard at their peril. Further, military culture is based on the assumption that the decisions of leadership are entitled to greater weight based on superior knowledge and judgment. Diffusing the power to make decisions strikes at the heart of the legitimacy of leadership; and so cognitive enhancements pose dangers to received military models that mere physical enhancements do not.

C. Service, Pay, and Promotion

Perhaps we should think about enhanced warfighters as we do with other specially trained operators, such as the Army’s Special Forces or Navy SEALs. That is, military policy could be to keep the enhanced separated from the unenhanced, in special or elite units; this would reduce any friction between the two groups.⁴

However, this segregation may merely telescope the problem out to a broader level, shifting tension from within units to among different units: If special units are given access to enhancements, or otherwise treated or rewarded differently—assuming we can even think of enhancements as rewards—then other units may feel slighted. Indeed, O’Meara [2012b] identifies this kind of inequity as a primary cause for dissension in the ranks presently. So it matters to military policy who gets enhanced and when, if not also how and why.

But as we alluded to above, and discussed in the definitional section of this report, it may be an open question of whether a particular enhancement may be a benefit to the individual. Leaving disenchantments aside, some or many enhancements pose side-effect risks; for instance, we still do not adequately understand the role of sleep and long-term effects of sleep deprivation, even if we can engineer a warfighter to operate on very little or no (true) sleep, as some animals are already capable of doing. So depending on one’s perspective, an enhance-

.....
⁴ There may be other reasons to segregate the enhanced, as well as countervailing reasons. We mention a few throughout this report, though if a policy to segregate were under serious consideration, it would need much further study.

ment could be a reward or benefit, or it could be an undesired risk, as some believed about anthrax vaccinations [Wasserman and Grabenstein 2003; Berkelman, Halsey, and Resnik 2012].

How, then, should enhancements affect the service commitment of military personnel? Insofar as an enhancement is costly to develop and represents an investment, it may be reasonable to expect the enhanced warfighter to commit to longer service. But if an enhancement is seen more as a risk, then perhaps a shorter length of service is appropriate for the enhanced. Similar decisions may need to be made with respect to pay, promotions, and so on. For instance, if promotions and “danger pay” may be used to incentivize volunteers, enhanced soldiers could be better positioned and more likely to accept dangerous missions in exchange for those benefits.

D. Lessons from the “Drone Wars”

On the mission side of operations, human enhancements may elicit a backlash that hinders the mission and therefore detracts from the value of enhancements for the military. This kind of blowback is already seen with the US government’s use of unmanned aerial vehicles (UAVs) in the so-called “drone wars”: While the US views its target strikes as appropriate—if not ethically required—to the extent that it is taking American military personnel out of harm’s way in a presumably just campaign, adversaries see drones as a cowardly and dishonorable proxy for a military afraid to engage face-to-face with human resistance. This sentiment seems to fuel resentment and hatred toward the US, which in turn helps to recruit more terrorists [Foust 2012; Plaw 2012]. Similarly, if military enhancements are regarded by adversaries as cowardly or abominable, they may be counterproductive to

the larger war for “hearts and minds”, even if the enhancements work as designed.

Another criticism of the drone wars that may be applied to military enhancements is the charge that these technologies, by better ensuring the survival and success of our own military personnel, serve to make war more risk-free and therefore a more palatable option [Lin, Abney, and Bekey 2008; Lin 2010; Lin 2011; Human Rights Watch 2012]. That is, we may be tempted into choosing a military option during a political conflict, rather than saving war as the last resort as demanded by just-war theory. This ethical imperative is reflected in Civil War General Robert E. Lee’s observation: “It is well that war is so terrible; otherwise, we would grow too fond of it” [Cooke 1876, 184; Levin 2008]. As war becomes less terrible—at least for our own side—our natural aversions to it may be lessened as well.⁵

This criticism leads to other related charges such as that drones are making it easier to wage war secretly, thus subverting democratic requirements, e.g., any due-process afforded to targets who are US citizens and the War Powers Resolution of 1973 [50 USC §1541-1548]. To the extent that enhancements can make it easier for military teams to covertly conduct missions and penetrate enemy lines, it would likewise be easier to conduct illegal operations, such as assassinations and cross-border attacks without the permission of the receiving nation-state.

.....
⁵ While reasonable on the face of it, this position could perhaps be fatally flawed, if it implies that war ought to be as terrible as possible in order to create maximum disincentive against choosing war over other options [Lin, Bekey, and Abney 2008].

Other relevant issues echo from robotics and other technologies, especially for neural devices and other cognitive enhancements. For instance, if hacking is a concern for UAVs and other robotics, then hacking is also a plausible concern for neural devices implanted or connected to the human brain [Denning 2009; Martinovic 2012]. As another example, where we struggle with questions of moral and legal responsibility with (future) autonomous robots, we may face the same questions sooner with human minds enhanced by either electronic or other biochemical means [Vincent 2012].

The above discussion certainly does not exhaust all the legal and operational issues that will arise from military human enhancements; and so more research is needed in this area. In the following section, we introduce another complication in the analysis of enhancement technologies: the impact to the warfighters themselves.



B | Center for
Technology Innovation
at BROOKINGS

B | Center for
21ST Century Security and Intelligence
at BROOKINGS

Compromised By Design? Securing the Defense Electronics Supply Chain

John Villasenor

November 2013

Executive Summary

John Villasenor is a nonresident senior fellow in Governance Studies and the Center for Technology Innovation at Brookings. He is also a professor of electrical engineering and public policy at UCLA.

Electronic “chips” are found everywhere—not just in critical defense systems, but also in the broader infrastructure for power, finance, communications, and transportation. All of these systems function effectively only when the electronic circuits at their heart can be trusted to operate as intended.

Unfortunately, ensuring trust has become much more difficult in recent years. Concern over the growth of counterfeit electronics (parts that have been harvested from discarded systems, relabeled, and sold as new to unsuspecting buyers) has grown in recent years.¹ These parts can fail prematurely, with potentially disastrous consequences. Thanks to recent congressional attention, improved detection methods, and heightened screening requirements for parts destined for defense systems, however, the threat of counterfeits is being actively addressed.

Yet the supply chain is almost completely unprotected against a threat that may turn out to be more significant in the long term: Chips could be *intentionally* compromised during the design process, before they are even manufactured. If placed into the design with sufficient skill, these built-in vulnerabilities would be extremely difficult to detect during testing. And, they could be exploited months or years later to disrupt—or exfiltrate data from—a system containing the compromised chip.

As chips have gotten more complex and design teams have grown larger and more globalized, the opportunities to insert hidden malicious functionality have increased. If the history of cybersecurity has taught us anything, it is that these opportunities will be exploited. The prudent question, therefore, is not “will intentionally compromised hardware end up in the defense electronics supply chain?” but “how do we maintain security when it inevitably does?” This paper aims to help frame the discussion regarding how best to respond to this important and underappreciated aspect of cybersecurity.

Introduction: The Growth of Chip Complexity

It is almost impossible to overstate the importance of integrated circuits—or “chips”—to modern society. We rely on them to run the Internet, the power grid, the financial markets, laptop computers and mobile phones, food distribution networks, medical equipment, and an essentially endless list of other networks and devices. Defense systems are similarly reliant on chips. To be effective, the military needs the ability to gather, analyze, and move information, and, when necessary, to move troops, supplies, and weapons systems. In today’s military, chips are essential to all of these tasks.

Thanks to continued advances in electronics, chips have become staggeringly complex. In accordance with Moore’s Law,² the number of transistors—the basic electrical elements used for implementing logic functions—that can be built into a single chip has roughly doubled every two years since the 1960s. In the early 1970s, a chip with several thousand transistors was considered large; today, chips with well over a billion transistors are routine. These advances are directly responsible for many of the conveniences we have learned to take for granted, including the ability to perform Internet searches and to carry smartphones having more computational power than a roomful of 1960s-era computers.

Yet this complexity, which has so enhanced our capabilities, also complicates trust. In the days when the largest chips had only a few thousand transistors, it would have been nearly impossible for a rogue designer to maliciously compromise circuitry without being detected. Post-manufacturing testing could explore nearly all of the functionality of the chips in that era. In addition, the people performing the testing had often been deeply involved in the design. Just as a longtime resident can know every street and building in a small town, the designers of a small chip could know the role of every circuit, and easily recognize any unauthorized changes.

A cyberattack launched using a chip containing compromised circuits could:

- Exfiltrate data while making the chip appear to function normally
- Corrupt data within the chip
- Stop the chip from functioning

Hardware-based cyberattacks:

- Harder to conduct than software attacks, since far fewer people have the necessary skills and access
- Harder to defend against, since replacing corrupted hardware can be extremely difficult and expensive

Today’s chips have become so complex that no single person can understand every detail of a chip’s design. Even the fastest automated testing methods would take many years to exhaustively test everything that a modern large chip can do. To avoid this obviously impractical outcome, testing is done on a statistical basis. A small fraction³ of all possible inputs are provided to the chip, and the resulting observations are used to infer behavior even for inputs that weren’t

specifically tested. Historically, this approach has worked quite well. The laws of probability ensure that a properly designed suite of tests will be extremely effective at

identifying accidental design flaws. However, intentionally hidden alterations inserted by a skilled designer would be much harder to find. In particular, latent functionality configured to be dormant until triggered months or years later would be invisible to testing protocols built on the assumption that all of the possible behaviors of a chip can be easily explored during the verification process.

Hardware: A Gaping Cybersecurity Exposure

For all the attention paid in recent years to cybersecurity, it remains largely software-focused, both in terms of the techniques employed and the expertise of the people and companies working in the field. This is a blind spot; hardware represents a gaping and exploitable hole in the current approach to cybersecurity. While software cybersecurity remains critically important, a complete cybersecurity strategy now requires consideration of hardware as well.

The varied means of attack illustrate how hardware-level vulnerabilities can be exploited to completely sidestep software-based security countermeasures. For example, a team of university researchers recently demonstrated that carefully chosen alterations in portions of a chip involved in encryption processing could allow an attacker to extract encryption keys.⁴ These compromises did not require the addition of any additional circuitry, but instead involved introducing subtle modifications in the electrical behavior of certain transistors in the chip. The modifications would not be noticed by an unsuspecting observer, and would not be detectable by any of the software running on the chip. But an attacker with specific knowledge of where to look could exploit them to decrypt data coming off the chip.

Another possible attack, demonstrated in a 2011 paper,⁵ involves corrupting the circuitry responsible for governing the data movement within a chip. When activated, the effect would be similar to turning all of the traffic lights in a city to red: Data movement on the chip would simply grind to a halt. Inspection of the software running on the system containing the chip would not provide any substantive insight into the nature of the problem. In some systems, a software-based inspection would not even be capable of identifying the chip as the location of the problem. And even if the chip were suspected, software-based techniques would typically be unable to pinpoint the circuitry within the chip responsible for the problem, and would be unable to rectify it.

Yet another attack involves intentional corruption of data within a chip. The corruption could be designed to act only after receipt of a specific externally delivered trigger. Or, it could be triggered automatically by the arrival of a pre-programmed date, by use of the chip in a defense facility of particular interest, use in close proximity to certain GPS coordinates, or any combination of the above. The presence of the malicious circuitry would be nearly impossible to detect during pre-deployment testing.

The impact of a hardware-based cyberattack could range from relatively modest to catastrophic. In some cases, an attack might cause a relatively minor malfunction that, while undesirable, could initially be handled using existing protocols put in place to

detect and mitigate non-intentionally-caused failures.⁶ However, many attacks would be far more alarming. An attack aimed at enabling exfiltration and/or eavesdropping of defense/intelligence communications networks would pose grave national security concerns. An attack targeting navigation or control systems in defense aircraft, ships, submarines, or land vehicles could put lives at risk.

Attacks targeting systems for industrial control or manufacturing automation could cause extensive physical and economic damage, and even lead to loss of life. In 1999, the failure of a control system—believed to have been accidental—contributed to a gas pipeline rupture that killed three people and injured multiple others.⁷ In early 2003, a computer worm left safety monitoring systems in an Ohio nuclear plant inoperable for several hours.⁸

Unfortunately, the lists of potential attacks and U.S. defense (and non-defense) systems that could be targeted are almost infinitely long. And, as explained above, detecting intentional compromises is much harder than in the past due to the size of today's chips. But how would a rogue designer insert malicious functionality in the first place? To answer that question, it's helpful to consider how chips are designed and how the dynamics of the global semiconductor market have changed in recent years.

Building a Chip

While the details of creating a new chip are complex, at the highest level the process involves a small number of basic steps. The first is to specify what the chip will need to do. This includes identifying the functions it will perform, how fast it will need to perform them, and constraints on power, size, and cost. The next step is design, which entails mapping the desired functionality first into a set of logical operations, and then into the corresponding electrical circuits. A description of the completed design is then sent to a semiconductor manufacturing facility for fabrication into an actual, physical chip. The chip is subjected to testing, and if there are no problems identified, it can be shipped to customers and incorporated into a product.⁹

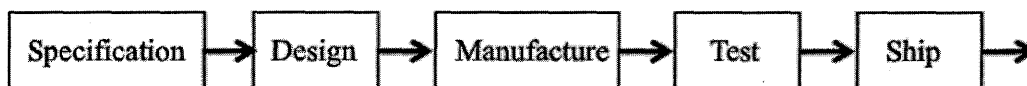


Figure 1: Steps in creating a chip

In the earliest days of the semiconductor industry, a single company would often do the specification, design, manufacturing and testing. Some companies, including IBM, Intel, Samsung, and Texas Instruments still operate in this manner.¹⁰ However, the costs of building manufacturing facilities—sometimes referred to as fabrication facilities, or more commonly, “fabs”—have gone from extremely expensive to stratospheric. Back in the 1980s, a fab could cost over \$200 million in 2013 dollars.¹¹ As technology advances enabled the production of chips with ever-smaller features, semiconductor manufacturing equipment got more advanced—but also more expensive. Fab costs, which are closely

coupled to the purchase prices of manufacturing equipment, have risen accordingly. In late 2012, Samsung broke ground on a new fab in Xian, China that will cost \$2.3 billion¹² initially and involve a total investment of \$7 billion¹³ by the time the project is completed. Gartner, Inc. has projected that “[i]ncreasing costs of manufacturing equipment will drive the average cost of semiconductor fabs to between \$15 billion and \$20 billion by 2020.”¹⁴

The prohibitive costs of building an in-house manufacturing capability spurred the growth of services enabling semiconductor companies without their own fabs to send their designs to an external facility known as a “foundry” for manufacturing. (A “foundry” is the term used for a fab—or, equivalently, a manufacturing facility—that can be accessed on a contract basis by companies that have designed a chip, but need access to external manufacturing capacity in order to build it.) This made it possible to start a semiconductor company without incurring the immense capital costs associated with constructing a fab. Both Qualcomm,¹⁵ which was founded in 1985, and Broadcom,¹⁶ which was founded in 1991, were built using this model. Today, the ranks of “fabless” semiconductor companies include not only giants like Qualcomm, Broadcom, AMD,¹⁷ and Nvidia, but also hundreds of smaller companies, some with only a few dozen employees.

Manufacturing: A Recognized Security Concern—But Only Part of the Problem

As the semiconductor industry grew through the 1990s, cost pressures pushed more and more manufacturing of American chip designs to offshore foundries. While this raised some concerns related to intellectual property security among American semiconductor companies focused on the consumer and business markets, it created a particularly acute challenge for those designing chips for sensitive defense and intelligence systems. Classified designs, of course, could not be shipped overseas for manufacturing. And even unclassified designs can contain sensitive information that could raise national security concerns if improperly disclosed.¹⁸

To address this, in 2004, the Department of Defense and the National Security Agency jointly funded a “trusted foundry” at a preexisting IBM semiconductor manufacturing facility in Vermont.¹⁹ The Trusted Foundry Program is administered by NSA’s Trusted Access Program Office²⁰ and aims “to ensure that mission-critical national defense systems have access to leading-edge integrated circuits from secure, domestic sources.”²¹ The program has now grown to over 50 accredited suppliers,²² including over a dozen trusted foundries as well as companies offering design, test, and other services.²³ While the original focus of the program was manufacturing, there are now multiple participating suppliers focused purely on providing trusted design services.

While the Trusted Foundry Program has been vitally important for enabling the secure production of the most sensitive chips, it is used for only a small fraction of the chips in defense systems. When purchasing computers, routers, navigation and communications

equipment and most other electronics hardware, The Department of Defense (DoD) is heavily reliant on the commercial supply chain—and therefore exposed to any associated vulnerabilities.

In addition, from the standpoint of someone wishing to insert malicious functionality, the manufacturing stage—which has been the traditional focus of chip security concerns—is not the weak link in the chain. Designs are provided to manufacturers as descriptions of the shapes and locations of all the silicon and metal structures that must be built into the chip. It is possible, but very expensive and time consuming, to reverse engineer the full functionality of a chip from the information provided to a manufacturer. Attempting to insert malicious functionality by directly modifying the description of on-chip structures would be difficult, and in some (though not all)²⁴ cases would create easily detectable defects. The task facing an attacker is much easier if he or she can get access at an earlier stage of the supply chain, when the design is still being created.

The Chip Design Ecosystem: More Globalized, More Complex

The semiconductor industry has evolved significantly over the last decade. In 2003, according to an analysis by PwC, global semiconductor expenditures for consumption (as opposed to production) were \$166 billion, in non-inflation-adjusted terms.²⁵ In 2003, Japan accounted for about 23% of this total, and each of the Americas, Europe, China, and the rest of the world (ROW) contributed between 18% and 20%.²⁶ By 2012, the global market had grown to slightly under \$300 billion. China had risen to the dominant position, with 52.5% of the total.²⁷ The Americas, Europe, and Japan accounted for only 12.4%, 11.1%, and 7.3% respectively.

Revenues related to semiconductor production have also experienced both growth and offshore migration, including to countries that are not U.S. alliance partners. In China, chip manufacturing (including packaging and testing) revenues increased from \$3.7 billion in 2003 to over \$24 billion in 2012.²⁸ Chip design revenue to companies in China was \$130 million in 2000 and \$540 million in 2003.²⁹ By 2012, it had grown to \$9.87 billion³⁰—an amount, after adjusting for inflation, over 14 times higher than in 2003 and over 56 times higher than in 2000.³¹ India has experienced rapid growth in its semiconductor design market as well. According to a 2011 report from the India Semiconductor Association, semiconductor hardware design revenues in India were over \$1.4 billion in 2010 and expected to grow to over \$2 billion by 2012.³²

Statistics aside, anyone who has spent significant time interacting with semiconductor companies in recent years has seen plenty of anecdotal evidence that design has become a much more globalized—and less American—endeavor. The same cost pressures that pushed semiconductor manufacturing offshore during the 1990s have been acting during the past decade to push design offshore. In the early 2000s, fabless American semiconductor companies would often perform much of the circuit design in house and outsource the manufacturing to a foundry in Asia. Today, American companies do more

and more of their design overseas. In some cases, this is due to outsourcing. Alternatively, American companies will sometimes establish an overseas branch or acquire an overseas company with the specific goal of getting access to foreign design talent. This makes eminent economic sense. Salaries for highly skilled designers in China, India, Eastern Europe, and South Korea are lower than in the United States. And in global economic terms, the inflow of design revenue in these and other places has become an important contributor to economic growth and prosperity.

For highly sensitive American defense and intelligence applications, these changes raise challenges. It is extremely unlikely that the U.S. government would want the design of chips destined for use in such applications outsourced to China. And, to be fair, it is equally unlikely that the Chinese government would be comfortable entrusting the design of similarly sensitive chips to an American company. However, the global nature of the semiconductor design industry can leave governments with less choice than they might like over such matters. Chip design has become so globally interconnected that, for all but the most narrowly tailored applications and systems, there is no longer any economically practical way to avoid complex international supply chains.

Furthermore, it is the complexity of today's chip design ecosystem, as much as its international nature, that complicates the security picture. Just as a city has residential, commercial, and industrial areas, the real estate on a chip is partitioned into different sections, often called "blocks," with different functions. Some blocks of a chip are devoted to memory. Another block might be used to decode a JPEG file into an image that can be displayed on a screen. Often, a chip will contain a programmable block of circuitry that can be instructed using software to do many different things at different times as the processing needs change. In addition, in a chip, just as in a city, plenty of real estate is devoted to facilitating movement. The analog of city streets in a chip is the network of microscopic interconnecting metal lines that allow data to move within and among the various blocks.

A company leading the design of a large chip will typically rely not only on its own engineers, but will also obtain a significant fraction of the design by purchasing blocks created by other companies. Purchased blocks arrive not as physical pieces of silicon, but as files of computer-like code expressing the logic that will eventually be converted into a physical circuit. The lead company assembles all of these design files into a single description of the entire chip. Next, the company performs a set of computer simulations aimed at confirming that the chip will behave as expected. If problems are found, the design is modified and checked again through another round of simulations. When all of the identified problems have been ironed out, the design is sent to a fabrication facility for manufacturing.³³

Why Design Corruption Is a Growing Threat

There are multiple security exposures in this process. One concern lies in the large number of organizations and people involved in the design of a single large chip. In addition to the lead company, there are companies subcontracted to provide design

services or provide pre-made designs. Some subcontracted companies may in turn farm out some of their work through a further level of subcontracting. Pieces of the chip design are stored and exchanged using a myriad of computers and networks, some of which may be insufficiently protected against external intrusions—potentially allowing an attacker to hack in and corrupt a previously healthy portion of the logic. There is also the risk of an insider threat among the dozens or hundreds of engineers with access to the design.

The overwhelming majority of designs will emerge from this process without any intentionally introduced flaws. But there are over 5000 new chips designed each year³⁴ in a globe-spanning ecosystem involving thousands of companies and hundreds of thousands of people. To conclude that chip designs will *never* be intentionally compromised flies in the face of much of what the last twenty years of cybersecurity have taught us. Design corruption is a very real, growing threat for multiple reasons:

1. The laws of statistics guarantee that there are people with the skills, access, and motivation to intentionally compromise a chip design

Globally, hundreds of thousands of people are employed as chip designers.³⁵ The overwhelming majority of them aim to produce the best, most effective designs possible. But some small percentage could be induced to intentionally compromise a design. Even if only 1/10 of 1% of chip designers would consider corrupting a chip for financial gain, competitive advantage, or other reasons, that still corresponds to hundreds of people with exactly the right skills and access. It would defy logic to assume that none of them will ever try.

2. A skilled attacker could compromise a design in a manner minimizing the chance of detection

An unskilled attacker who engaged in wholesale replacement of commonly used functionality within a critical block of a chip would easily be detected during the pre- or post-manufacturing testing. But a skilled attacker with access to detailed information about the targeted chip would be well positioned to identify the locations within the chip where carefully placed malicious circuitry would likely remain undetected. As noted above, chips have become so complex that testing is only partial. Malicious circuitry designed to lay dormant and avoid impeding the normal operation of the chip would be extremely difficult to detect using verification protocols aimed at accidental, as opposed to intentional, design flaws.

3. The threat of attribution is not a sufficiently strong disincentive

One of the most commonly articulated arguments against the risk of intentionally compromised chips holds that the attacker would easily be identified once the flaws were discovered. Thus, the argument goes, the near certainty of getting caught would dissuade a would-be attacker from actually compromising a chip in the first place. This logic is incorrect for several reasons.

First, a skilled attacker could introduce a flaw with plausible deniability (see point 4 below). Second, just as occurs in traditional software cyberattacks, the source of a design attack could be disguised. Consider a subcontracted design company with weak network security, allowing an outsider to break in and introduce malicious circuitry into the design. Years later, the compromise might be identified and traced to the company. The company's executives and designers—many of whom may have moved on to other jobs by that point in time—could be interviewed, and all would express bafflement about how the chip could have been corrupted. The computer records that might help identify a network intrusion as the attack vector might be long gone, and the attacker would be off the hook.

More complex combinations of the above are also possible. For example, an insider intent on corrupting a design could create a software vulnerability in his or her employer's networks, and then introduce the design corruption by accessing those networks from an anonymized, off-site location.

Despite all of the above, a thorough after-the-fact investigation of a malicious design alteration may eventually be able to identify the person who introduced it and prove that it was intentional. All of that, however, would occur downstream, potentially years after the design was corrupted. As long as the attacker believed—perhaps incorrectly—that he or she was unlikely to get caught, the threat of attribution would be a weak disincentive.

4. A skilled attacker could introduce a flaw with plausible deniability

Consider a skilled attacker who builds a back door into the circuitry in a chip, with the intent of exploiting it years later to exfiltrate data or configure the chip to impede its function. What would happen if the back door were discovered and studied, and the designer who inserted it identified and confronted? He or she could respond to accusations of tampering by characterizing the back door as a feature that was introduced to assist in testing early prototypes of the chips. The intent, he or she could claim, was to remove it before high volume manufacturing. But when there are a billion transistors to think about, some things fall through the cracks. Leaving the back door open, the attacker could say, was simply a mistake—albeit one that other people could later attempt to exploit for malicious purposes.

To make matters more complex, honest designers sometimes make mistakes that can increase the vulnerability of a chip to being called into service in a later hardware-based cyberattack. In some cases, disentangling genuine mistakes from intentional changes intended to weaken security could be extremely difficult.

5. An attacker could afford to cast a wide net, knowing that only a tiny fraction of the corrupted chips would end up in systems of interest

In many cases, a single design is used to manufacture millions of identical copies of a chip, only a tiny fraction of which might end up in systems of interest to an attacker. Knowing this, the attacker could cast a wide net by modifying the design to provide an

extremely small back door in every one of these systems, with the full knowledge that most of the back doors would never be exploited. The back door could be designed so that a software- or firmware-based probe could be used later to query the chip and obtain information regarding the system in which it is installed.

Growing Recognition of the Threat

While the globalization of semiconductor design is a more recent trend, U.S. government concerns about the threat of intentionally corrupted chips were initially spurred by the offshore migration of manufacturing that occurred during back in the 1990s and early 2000s. In October 2003, then-Deputy Secretary of Defense Paul Wolfowitz wrote in a widely distributed unclassified memo that the “country needs a defense industrial base that includes leading edge, trusted commercial suppliers for critical integrated circuits used in sensitive defense weapons, intelligence, and communications systems.”³⁶ A few months later, at DoD request, the Defense Science Board convened a task force on “High Performance Microchip Supply.”³⁷ The task force’s report, published in 2005, offered a blunt conclusion:

The Department of Defense and its suppliers face a major integrated circuit supply dilemma that threatens the security and integrity of classified and sensitive circuit design information, the superiority and correct functioning of electronic systems, system reliability, continued supply of long system-life and special technology components.³⁸

The report also stated that “[t]rust cannot be added to integrated circuits after fabrication; electrical testing and reverse engineering cannot be relied upon to detect undesired alterations in military integrated circuits.”³⁹

In 2007, the Defense Advanced Research Projects Agency (DARPA) launched⁴⁰ the Trust in Integrated Circuits⁴¹ program to address the challenges of “ensuring IC hardware works as intended and ensuring the design and fabrication process can be trusted.”⁴² This was followed, in 2011, by DARPA’s Integrity and Reliability of Integrated Circuits⁴³ program and by an Intelligence Advanced Research Projects Activity (IARPA) program in Trusted Integrated Circuits.⁴⁴

These programs have generated important technological advances in the ability to secure the chip supply, and in some cases have addressed not only manufacturing concerns but also design vulnerabilities. But the threat landscape, particularly with regard to chip design, is continuing to evolve, and the scale of the challenge is immense. Even if sufficiently effective technological solutions are found, they will have limited impact in the absence of a broader national strategy and an accompanying set of policy responses.

In 2013, there have been encouraging signs that supply chain integrity is gaining visibility within the broader cybersecurity dialog. In February, President Obama issued an Executive Order⁴⁵ requiring, among other things, that NIST develop a “Cybersecurity Framework” addressing risks to critical infrastructure. While the scope of the Order is

very broad, a discussion draft of the framework, while primarily focused on software, identified the supply chain as one of the areas needing increased attention,⁴⁶ observing that “[s]upply chain risk management, particularly in terms of product and service integrity, is an emerging discipline characterized by diverse perspectives, disparate bodies of knowledge, and fragmented standards and best practices.”⁴⁷

A June 2013 report from Semiconductor Research Corporation and the Computing Community Consortium, “Research Needs for Secure, Trustworthy, and Reliable Semiconductors,” focused specifically on emerging challenges to ensuring hardware integrity.⁴⁸ The report identified an “overarching need . . . for research in ‘Design for Security,’ ” and identified seven specific research challenges requiring attention.⁴⁹ The report concluded that “[n]ow is the time to launch a collaborative program of research with industry and government support in ‘design for security.’ ”⁵⁰ Technology measures are a critical aspect of the solution. But a complete solution will also require reframing the policy discussion to better reflect today’s electronics industry.

Addressing the Threat: Some Guiding Principles

Against this backdrop, here are some principles that should guide the development of a more comprehensive strategy to secure the defense electronics supply chain against intentionally compromised hardware:

- **Many of the most significant security exposures today lie in chip design; focusing on manufacturing alone is no longer sufficient**

The current approach to supply chain security still largely reflects a decade-old view that the most significant vulnerabilities lie in manufacturing. This is no longer the case. As chip complexities have increased, the vulnerabilities have expanded upstream in the supply chain, to include design. Corrupting a chip during manufacturing first requires determining which subset of the millions or billions of transistors to target. To obtain this knowledge, a malicious manufacturer would either need to engage in a lengthy and time consuming reverse engineering effort, or, more likely, to obtain inside information from someone directly involved in the design process. By contrast, design-stage corruption is easier, less expensive, and offers a far broader range of opportunities to an attacker.

- **Outsourcing is only part of the concern**

To the extent that offshore outsourcing can reduce the ability to maintain oversight into design integrity for defense electronics components, it is a concern. But given the complexity of today’s design environment, it would be a mistake to assume that onshore designs are of necessity less vulnerable to corruption than those created offshore. Intentional design corruption requires either 1) the ability to hack into the computer systems used to hold a design as it is being created, or 2) access to an insider willing to insert malicious circuitry. Geography offers little protection against either of these attack vectors.

- **Trust should not be assumed**

Electronics components destined for defense systems are often subjected to an initial screening process aimed at ensuring they are not counterfeit, and that they operate as designed. Once those tests are passed, trust is assumed. However, a skilled attacker could embed latent malicious functionality and triggering it long after the system containing the part was deployed. To address this, systems should be designed to actively assess trust of their components throughout their service lifetimes.

- **Most of the global semiconductor market is outside the U.S.**

A 2009 White House-directed “Cyberspace Policy Review” concluded with respect to supply chain concerns that, somewhat circularly, “[t]he best defense may be to ensure U.S. market leadership through continued innovation that enhances U.S. market leadership and the application of best practices in maintaining diverse, resilient supply chains and infrastructures.”⁵¹

However, the U.S. now accounts for less than 13% of the global market for semiconductor consumption.⁵² Semiconductor production (including design) is also largely based outside the U.S. A practical supply chain security strategy needs to reflect the economic reality that most defense systems, and nearly all non-defense systems, will need to rely on chips that were designed and built, at least in part, outside the U.S.

- **The security of the commercial supply chain needs more attention**

The greatest supply chain security exposure for defense applications comes not from the small fraction of chips designed and manufactured uniquely for defense systems, but from the massive inflow of commercial chips into those systems. While the designers of chips for sensitive defense and intelligence applications have long recognized concerns raised by the hardware supply chain, in the commercial world cybersecurity attention is almost never directed to the supply chain. In order to better protect DoD systems as well as the broader non-defense infrastructure, it is critical to raise the level of attention regarding the commercial chip supply as potential cyberattack vector.

- **Design practices within the semiconductor industry should be changed**

Design practices in the semiconductor industry should be modified to specifically recognize and address the potential for malicious hardware insertion. Companies engaged in chip design should adopt a greater level need-to-know partitioning of information.⁵³ When subcontracting design work to external parties, companies should do diligence not only the design skills of the contractor, but their procedures for ensuring design security.

- **Responses to hardware-based cyberattacks should be formulated in advance**

To the extent that there is not already a well-formulated plan that would guide the response to a significant, hardware-based cyberattack on defense and/or other critical

infrastructure systems, one should be developed. The governmental entity that would be charged with overseeing the response to a hardware attack should be identified preemptively, and procedures should be put in place for reporting an attack and for engaging with the appropriate companies and governmental organizations.⁵⁴

Making Hardware Cybersecurity a Reality

Addressing hardware cybersecurity will require action at multiple levels on the part of various governments around the world. Much can be done to improve the level of government awareness and preparedness regarding hardware-based cybersecurity threats. In the U.S., the government generally and DoD in particular should expand its current software-centered approach to give more attention to the threat from maliciously modified hardware. Currently, awareness of the threat is high within some subsets of the DoD research community (notably DARPA and IARPA), but at the operational level, cybersecurity everywhere—both within and outside of government—is still considered a software-only concern. An indirect measure of this can be found by looking at the companies that provide cybersecurity products aimed at government and corporate customers, the overwhelming majority of which appear to be focused on software. Other important government steps include development of a response plan for addressing significant hardware-based attacks, and procurement incentives aimed at encouraging suppliers to actively mitigate hardware cybersecurity risks.

Companies that make chips have a role to play as well. As noted above, they can change their design practices to reduce the likelihood of chip corruption. They can more carefully vet their design suppliers and more carefully track design changes as a chip is developed. Chipmakers should also consider building countermeasures into chips that can help identify and respond to an attack as it occurs. Importantly, many of these effective countermeasures could be introduced at negligible additional cost.⁵⁵

Companies that provide design tools—i.e., the software products used by hardware designers during the chip design process—can also give increased attention to the potential for malicious design insertion. Design tools can be improved to 1) enable more careful tracking of design changes, 2) enable increased testing for intentionally introduced flaws, and 3) recognize that different blocks within a chip may have different levels of trustworthiness. The coming years will also see significant private-sector opportunities to provide hardware cybersecurity expertise for government and corporate customers. This will require a very different set of technical skills than those used in software cybersecurity environments.

Conclusions

Too often, we wait for catastrophe to spur change. As there has not yet been a string of publicly disclosed examples of defense hardware with malicious design alterations, it is hard to spur interest in investing significant effort to address the inevitability of intentionally compromised hardware. But given the critical role of chips in nearly every

defense system, there are good reasons to be proactive as opposed to purely reactive with respect to hardware cybersecurity.

None of this means that hardware cybersecurity will require the same level of effort and expense that has been directed to software cybersecurity. Software has always been, and will remain, the more significant vulnerability. But the commonly held view that software is the *only* vulnerability is out of step with the reality of how today's systems are designed and built.

At the 2011 Aspen Security Forum, retired Gen. Michael Hayden, who formerly headed both the CIA and NSA, said with respect to compromised hardware, "Frankly, it's not a problem that can be solved . . . This is a condition that you have to manage."⁵⁶ Unfortunately, this dark assessment is accurate. But while the bad news is that the problem can't be completely solved, the good news is that there is enormous opportunity to more proactively manage and mitigate the risks.

Acknowledgements

I would like to thank Chris Daverse of Defined Business Solutions, Peter Singer of Brookings, and Ian Wallace, also of Brookings, for providing comments on an earlier draft.

¹ Recycled components are the largest, but not the only, types of counterfeit parts. See John Villasenor & Mohammad Tehranipoor, *Chop Shop Electronics - Clever Counterfeiters Sell Old Components as New, Threatening Both Military and Commercial Systems*, IEEE SPECTRUM, Sept. 20, 2013, <http://spectrum.ieee.org/semiconductors/processors/the-hidden-dangers-of-chopshop-electronics>.

² Intel co-founder Gordon Moore was one of the first people to observe the exponential growth in chip complexity. In 1965, while he was at Fairchild Semiconductor, he wrote a paper suggesting a doubling every year. See Moore, Gordon E., "Cramming More Components onto Integrated Circuits," Reprinted from ELECTRONICS, Vol. 38, No. 8, pp. 114-17 (Apr. 9, 1965)," PROCEEDINGS OF THE IEEE, Vol. 86, No. 1, pp. 82-85 (Jan. 1998), available at <http://www.cs.utexas.edu/~fussell/courses/cs352h/papers/moore.pdf>. In 1975, he offered a revised prediction of "a doubling every two years, rather than every year." See 1965 - "Moore's Law" Predicts the Future of Integrated Circuits, COMPUTER HISTORY MUSEUM, <http://www.computerhistory.org/semiconductor/timeline/1965-Moore.html> (last visited on Oct. 5, 2013).

³ The number of possible inputs is so large that exhaustive testing is impossible.

⁴ Zeljka Zorz, *Researchers create undetectable layout-level hardware Trojans*, HELP NET SECURITY, Sept. 17, 2013, <http://www.net-security.org/secworld.php?id=15589>.

⁵ Kim, Lok-Won & Villasenor, John D., *A System-On-Chip Bus Architecture for Thwarting Integrated Circuit Trojan Horses*, IEEE TRANSACTIONS ON VLSI SYSTEMS, Vol. 19, No. 10, pp. 1921-1926 (Oct. 2011), available at <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5556060>.

⁶ In the ensuing investigation, it would of course be alarming to discover that the failure was intentionally caused. Such a discovery could require an expensive and complex response to rid similar systems of the corrupted parts.

⁷ Ryan Singel, *Industrial Control Systems Killed Once and Will Again, Experts Warn*, WIRED.COM, Apr. 9, 2008, <http://www.wired.com/threatlevel/2008/04/industrial-cont/>.

⁸ Id.

⁹ In the interest of a simple, readable description, this high-level characterization of how a chip is made omits many details and potential variations.

¹⁰ This does not mean that companies such as TI manufacture all of their chips in house. For example, TI has outsourced some manufacturing to UMC. See United Microelectronics Corp. Press Release, UMC

Recognized for Excellence by Texas Instruments, Apr. 27, 2009,
<http://www.umc.com/English/news/2009/20090427-1.asp>.

¹¹ See Thomas C. Hayes, *Texas Instruments In Gamble*, N.Y. TIMES, Mar. 29, 1986,
<http://www.nytimes.com/1986/03/29/business/texas-instruments-in-gamble.html>, noting that “Texas Instruments’ efforts have included spending \$200 million to build wafer-fabrication plants at its headquarters in the Dallas suburb of Richardson and in Miho, Japan.” While the article does not break down the cost across the two plants, one of the plants must have cost at least \$100 million. Using CPI numbers published by the Social Security Administration at *Average CPI By Quarter And Year*, SOCIAL SECURITY ADMINISTRATION, <http://www.ssa.gov/oact/STATS/avgcpi.html> (last visited on Sept. 1, 2013), \$100 million in 1986 corresponds to about \$210 million in 2013 dollars.

¹² Samsung Press Release, *Samsung Breaks Ground for Memory Manufacturing Complex in China*, Sept. 12, 2012, <http://www.samsung.com/global/business/semiconductor/news-events/press-releases/detail?newsId=12043>.

¹³ Id.

¹⁴ Johnson, B., Freeman, D., Christensen, D. & Wang, S.T., *Market Trends: Rising Costs of Production Limit Availability of Leading-Edge Fabs* [Abstract] (ID: G00238123), GARTNER, INC. (Sept. 1, 2012), available at http://www.gartner.com/DisplayDocument?doc_cd=238123.

¹⁵ *Who We Are: History*, QUALCOMM INC., <http://www.qualcomm.com/who-we-are/history/story> (last visited on October 5, 2013).

¹⁶ *Facts at a Glance*, BROADCOM CORP.,
<http://www.broadcom.com/docs/company/BroadcomQuickFacts.pdf> (last visited on Oct. 5, 2013).

¹⁷ AMD’s manufacturing arm was spun off in 2009 to form GLOBALFOUNDRIES. See *About GLOBALFOUNDRIES*, GLOBALFOUNDRIES INC., <http://www.globalfoundries.com/about/> (last visited on Oct. 5, 2013).

¹⁸ A separate, related issue that arises for all chips, not only those designed specifically for defense applications, is the intellectual property contained in the chip design. Chip designs are prime targets for economic espionage.

¹⁹ McCormack, Richard, *\$600 Million Over 10 Years For IBM’s ‘Trusted Foundry’ Chip Industry’s Shift Overseas Elicits National Security Agency, Defense Department Response*, MANUFACTURING & TECHNOLOGY NEWS, Vol. 11, No. 3 (Feb. 3, 2004), available at <http://www.manufacturingnews.com/news/04/0203/art1.html>.

²⁰ *Trusted Access Program Office (TAPO)*, NATIONAL SECURITY AGENCY, July 1, 2013,
<http://www.nsa.gov/business/programs/tapo.shtml>.

²¹ *DoD Trusted Foundry Program*, ACQUISITION COMMUNITY CONNECTION, Oct. 23, 2011,
<https://acc.dau.mil/CommunityBrowser.aspx?id=480636>.

²² Ortiz, Catherine, *DOD Trusted Foundry Program – Ensuring “Trust” for National Security & Defense Systems* [PowerPoint presentation], TRUSTED FOUNDRY PROGRAM, at pg. 17 (June 20, 2013), available at <http://www.ndia.org/Divisions/Divisions/SystemsEngineering/Documents/Past%20Meetings/06-20-12%20Division%20Meeting/04%20-%20Trusted%20Foundry%20Program,%20Ortiz,%20NDIA%20SE%20Talk%2020120620%20cjo2.pdf>.

²³ For a map showing companies and services provided, see id. For a list of trusted foundries, see id. at pg. 18.

²⁴ Given sufficiently detailed knowledge of the design, there is a class of intentional circuit corruptions that could be introduced without creating easily detectable defects. See., e.g., *Researchers create undetectable layout-level hardware Trojans*, supra note 4.

²⁵ See *Continuing to grow: China’s impact on the semiconductor industry 2013 update* [Report] (the “2013 PwC Report”), PRICEWATERHOUSECOOPERS LLP, at pg. 3 (Sept. 2013), available at <http://www.pwc.com/gx/en/technology/chinas-impact-on-semiconductor-industry/assets/china-semicon-2013.pdf>.

²⁶ Id.

²⁷ Id.

²⁸ Id., at pg. 13. The chart separately considers “manufacturing” and “packaging and testing.” In 2003, manufacturing and packaging/testing comprised a total of 44.6% of \$8.3 billion, corresponding to \$3.7 billion. In 2012, manufacturing and packaging/testing comprised a total of 43.3% of \$56.3 billion, corresponding to \$24.4 billion.

²⁹ Id., at pg. 17.

³⁰ Id.

³¹ The PwC figures for chip design revenue to companies in China of \$130 million in 2000 and \$540 million in 2003 are not inflation adjusted. Using CPI numbers published by the Social Security Administration at *Average CPI By Quarter And Year*, SOCIAL SECURITY ADMINISTRATION, <http://www.ssa.gov/oact/STATS/avgcpi.html> (last visited on Sept. 27, 2013), \$130 million in 2000 corresponds to \$174 million in 2012 dollars. \$540 million in 2003 corresponds to \$679 million in 2012 dollars. By contrast, 2012 revenue was \$9.87 billion, over 14 times as high as the inflation-adjusted 2003 amount. The inflation-adjusted ratio of 2012 revenues to 2000 revenues is about 57.

³² *Study on semiconductor design, embedded software and services industry* [Report], INDIA SEMICONDUCTOR ASSOCIATION & ERNST & YOUNG GLOBAL LTD., at pg. 24 (Apr. 2011), available at http://deity.gov.in/hindi/sites/upload_files/dithindi/files/Semiconductor06April11_020511_0.pdf. The report states that total “semiconductor design industry” revenues were \$7.5 billion in 2010. However, that includes \$6.076 billion for embedded software development. “VLSI design” and “board design” were reported as \$944 million and \$476 million respectively; adding these figures gives \$1.42 billion. For the projected 2012 total of \$10.6 billion, the VLSI design and board design components were \$1.33 billion and 6.72 million respectively.

³³ The full design process involves not only ensuring that the design logic is correct, but that it will function as expected when, after a process called “layout,” it is mapped into the structures of the target semiconductor process.

³⁴ Ramamoorthy, Ganesh, *Market Trends: ASIC and ASSP Chip Design Start Trends, Worldwide, 2013* [Abstract] (ID: G00246635), GARTNER, INC. (Aug. 16, 2013), available at <http://www.gartner.com/id=2575215>.

³⁵ Countries with very large numbers of people involved in chip design include the United States, India, and China. In China, for example, the 2013 PwC Report, *supra* note 25 at pg. 17, noting that “The total number of employees in the IC design sector increased by 6% in 2012 to about 112,500 . . .”

³⁶ *Defense Science Board Task Force On High Performance Microchip Supply* [Report], OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY AND LOGISTICS, at pg. 87 (February 2005), available at <http://www.acq.osd.mil/dsb/reports/ADA435563.pdf>.

³⁷ Id., at pg. 83.

³⁸ Id., at pg. 5.

³⁹ Id., at pg. 3.

⁴⁰ Sally Adece, *Contracts awarded for DARPA's Trust in Integrated Circuits program*, IEEE SPECTRUM, Dec. 6, 2007, http://spectrum.ieee.org/tech-talk/semiconductors/devices/contracts_awarded_for_darpas_t.

⁴¹ The program is now referred to as the “Trusted Integrated Circuits program.” See *Microsystems Technology Office: Trusted Integrated Circuits (Trust)*, DARPA, [http://www.darpa.mil/Our_Work/MTO/Programs/Trusted_Integrated_Circuits_\(TRUST\).aspx](http://www.darpa.mil/Our_Work/MTO/Programs/Trusted_Integrated_Circuits_(TRUST).aspx) (last visited on Oct. 5, 2013).

⁴² *TRUSTED in Integrated Circuits Program BAA07-24 Industry Day Notice* (Solicitation No. SN07-24), DARPA, Mar. 16, 2007, <https://www.fbo.gov/index?s=opportunity&mode=form&id=56160bfade0d4086dbd06abb45668b18>.

⁴³ Adam Rawnsley, *Can Darpa Fix the Cybersecurity 'Problem From Hell?'*, WIRED.COM, Aug. 5, 2011, <http://www.wired.com/dangerroom/2011/08/problem-from-hell/>. DARPA's description of the program is at *Microsystems Technology Office: Integrity and Reliability of Integrated Circuits (IRIS)*, DARPA, [http://www.darpa.mil/Our_Work/MTO/Programs/Integrity_and_Reliability_of_Integrated_Circuits_\(IRIS\).aspx](http://www.darpa.mil/Our_Work/MTO/Programs/Integrity_and_Reliability_of_Integrated_Circuits_(IRIS).aspx) (last visited on Oct. 5, 2013).

⁴⁴ Michael Cooney, *US intelligence group wants to change the way chips are made*, NETWORK WORLD, Nov. 1, 2011, <http://www.networkworld.com/news/2011/110111-iarpa-chips-252624.html>.

⁴⁵ Exec. Order No. 13,636, 78 Fed. Reg. 11,739 (Feb. 19, 2013), available at <http://www.gpo.gov/fdsys/pkg/FR-2013-02-19/pdf/2013-03915.pdf>.

⁴⁶ *Discussion Draft of the Preliminary Cybersecurity Framework* [Report], NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (Aug. 28, 2013), at pg. 11, available at <http://www.acuta.org/wcm/acuta/legreg/083013b.pdf>.

⁴⁷ Id., at pg. 13.

⁴⁸ *Research Needs for Secure, Trustworthy, and Reliable Semiconductors* [Report], COMPUTING

COMMUNITY CONSORTIUM & SEMICONDUCTOR RESEARCH CORP. (June 2013), available at <http://www.src.org/emerging-initiative/cybersecurity/>.

⁴⁹ Id., at pg. 16.

⁵⁰ Id., at pg. 19.

⁵¹ *Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure* [Report], THE WHITE HOUSE, at pg 34 (May 29, 2009), available at http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf. The report document is undated and does not clearly identify specific agencies involved in writing the report at White House direction. However, a May 29, 2009 White House press release accompanying the report's release states that "In February 2009, President Obama directed the National Security Council (NSC) and Homeland Security Council to conduct a 60-day review of the plans, programs, and activities underway throughout government that address our communications and information infrastructure (i.e., 'cyberspace'), in order to develop a strategic framework to ensure that the U.S. government's initiatives in this area are appropriately integrated, resourced, and coordinated" and that "The review team's report to the President contains five main chapters, outlined below, and includes a near-term action plan for U.S. Government activities to strengthen cybersecurity." See The White House Press Release, *Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure*, May 29, 2009, <http://www.whitehouse.gov/the-press-office/cybersecurity-event-fact-sheet-and-expected-attendees>.

⁵² According to the 2013 PwC Report, *supra* note 25, at pg. 3, the Americas as a whole accounted for 14% of the global semiconductor consumption in 2011 and 12.4% in 2012. The report does not specify the subset of those percentages attributable specifically to the United States.

⁵³ The recommendation for more effective information partitioning was also discussed in a 2011 Brookings paper. See Villasenor, John D., *Ensuring Hardware Cybersecurity* [Report], THE BROOKINGS INSTITUTION, No. 9 (May 2011), available at <http://www.brookings.edu/research/papers/2011/05/hardware-cybersecurity>.

⁵⁴ Some of the text in this recommendation was previously published in a 2011 Brookings paper. See *id.*

⁵⁵ Ways to actively monitor the chip behavior and identify suspicious activity were discussed in a 2011 Brookings paper (see Villasenor, *Ensuring Hardware Cybersecurity*, *supra* note 53) and, in more detail, in a 2010 Scientific American article (see Villasenor, John, *The Hacker in Your Hardware* [Preview], SCIENTIFIC AMERICAN, Vol. 303, No. 2, pp. 82-87 (Aug. 4, 2010), available at <http://www.scientificamerican.com/article.cfm?id=the-hacker-in-your-hardware>).

⁵⁶ Rawnsley, *supra* note 43.

About the Centers

B | Center for **Technology Innovation** at BROOKINGS

Founded in 2010 and led by Director Darrell West, the Center for Technology Innovation (CTI) in Governance Studies at Brookings focuses on delivering research that impacts public debate and policymaking in the arena of U.S. and global technology innovation. At CTI, our research centers on identifying and analyzing key developments to increase innovation; developing and publicizing best practices to relevant stakeholders; briefing policymakers about actions needed to improve innovation; and enhancing the public and media's understanding of technology innovation.

B | Center for **21ST Century Security and Intelligence** at BROOKINGS

The Center for 21st Century Security and Intelligence (21CSI) in Foreign Policy at Brookings was created to address the key issues shaping security policy over the coming decades. The Center seeks to answer the critical questions emerging in defense, cybersecurity, arms control, and intelligence in an all-encompassing manner, seeking not just to explore important new policy challenges but also how they cross traditional fields and domains. Under the leadership of Peter W. Singer, one of the world's leading experts on changes in warfare, the Center focuses on delivering cutting-edge research, analysis and outreach that shapes public understanding and official decision-making across a broad range of security issues.

The views expressed in this piece are those of the author and should not be attributed to the staff, officers or trustees of the Brookings Institution.

Nuclear Blindness: An Overview of the Biological Weapons Programs of the Former Soviet Union and Iraq

Christopher J. Davis
Johns Hopkins University Center for Civilian Biodefense Studies

The demise of the biological weapons capability of the United States in 1969 and the advent of the Biological and Toxin Weapons Convention in 1972 caused governments in the West to go to sleep to the possibility of biological weapons development throughout the rest of the world, as technically knowledgeable workers were transferred and retired, intelligence desks were closed down, and budgets were cut. By 1979, despite the Sverdlovsk anthrax release, a senior British government policy official described any biological weapons threat as nebulous. President Nixon's biological weapons disarmament declaration in 1969 had conveyed the impression that biological weapons were uncontrollable and that the U.S. program had not been successful in producing usable weapons (when in fact the opposite was true). Add to this the rise of truly intercontinental ballistic missile delivery of nuclear weapons, and the stage was set for what I have termed "nuclear blindness" and defined as "the tunnel vision suffered by successive governments, brought on by the mistaken belief that it is only the size of the bang that matters." Throughout this period, both the former Soviet Union and Iraq conceived, albeit in different ways, their new biological weapons programs. It took until 1989-1991 for government technical experts in the West to persuade the world and their own governments that these programs were real and of enormous potential importance to the security of the West, if not the whole world.

Too many times in the past we have failed to anticipate future developments; refused to think the unthinkable and expect the unexpected. Too many times we have been out maneuvered by those who take the time to think and plan and do

Address for correspondence: Christopher J. Davis, O&H House, Ogbourne St. George, Marlborough, Wiltshire, SN8 1TF, United Kingdom; fax: 44-(0)-1672-841418; e-mail: christopher.davis@virgin.net.

In 1973 and 1974, the Soviet Politburo formed and funded the organization known most recently as Biopreparat (Chief Directorate for Biological Preparations), designed to carry out offensive biological weapons R&D and production concealed behind legal and civil biotechnology research. At no time did civilian biotechnology work ever comprise more than 15% of the activity at any of the 52 sites under the aegis of Biopreparat. Ultimately it was controlled by the Ministry of Defense, the Military Industrial Commission, and other state organs, all the way up to the Central Committee and what became eventually the Office of the President. Its head, a general, retained special access to the Central Committee from its inception, and through its links with the Academies of Science and Medical Sciences, Ministry of Health, and the Anti-Plague Institutes, recruited a generation of scientists who elsewhere in the world underpinned the expanding pharmaceutical and biotechnology industries and academic life-sciences research. The whole system probably employed at its height at least 50,000 people, many of whom were scientists and technicians with very high security clearance that identified them as part of a biological weapons program more closely held and more secret than its nuclear weapons counterpart. The system was always able to draw on the best from any source but was, to a certain extent, self-sufficient. Not all of the 52 establishments were occupied with microbiology or weapons—some were workshops, garages, and cover operations; others supported the program directly with fermenter design and construction or building of weapons test chambers; while yet others carried out advanced research, which would then be given to other institutes for development. Often there was internal competition, with one project being given to a number of facilities to see who would come up with the best idea. In its first 15 years alone, Biopreparat probably cost at least 1.5 billion rubles to create and run—a large sum for life-sciences R&D but relatively modest compared with the cost of nuclear weapons R&D and, therefore, in terms of strategic weapons, extremely cost-effective.

The main purpose of the enormous Biopreparat capability was to hide biological weapons research, development, and production formerly carried out solely in Ministry of Defense establishments behind a facade of nominally civilian biotechnology and pharmaceutical enterprises. The two systems, the former Ministry of Defense complex of biological weapons facilities and the new Biopreparat facilities, continued to operate side-by-side. The Ministry of Defense facilities themselves probably employed another 15,000 workers and had a separate budget, so that the potential within the system as a whole, which is how it should be considered, was large and dwarfed the by-then long-abandoned U.S. offensive program. Its capacity for production of agent was measured not in tons but in hundreds of tons for each of at least nine separate sites, primarily plague, tularemia, glanders, anthrax, smallpox, and Venezuelan equine encephalomyelitis.

Another mission of Biopreparat was to apply advances in biotechnology (genetic engineering, in particular) to improving the biological weapons capability of the former Soviet Union. This mission took several forms, supported primarily by the then vice-president of the Academy of Sciences, Yuri Ovcshinnikov, the most influential Soviet biomedical scientist of the 1970s. He saw a way around arms control treaties and weapons conventions by using microbes to produce biologically active substances that would replace classic chemical weapons; their production could then be concealed in the biotechnology or pharmaceutical industry. He also envisaged that the government would use genetic engineering to produce a new generation of biological weapons agents with enhanced capability for expressing toxins and other biologically active substances and to improve overall weapons effectiveness. The outcome of the first of these two programs is not known, but the latter was very successful. Moreover, the new Biopreparat-based program was able to address all aspects of agent production and delivery, not just the most advanced microbiological ones. It built strength in depth, having as its main aims to improve industrial production scale-up techniques, microbial production rates, yields of viable microorganisms, virulence, and resistance of microorganisms to antibiotics; to maximize viability of agent during dissemination and increased survivability of biological aerosols; and to enhance the ability of microorganisms to degrade the target's natural defenses. The leaders of the program foresaw increasing encroachment of international arms control

NOTICE:
This material may be protected by copyright law (Title 17 U.S. Code)

processes into the territory of sovereign states. Thus, they perceived the need for its weapons to become invulnerable to first strike or counterattack. Key technical targets associated with such an approach were the development of dry solid particulate agent formulations, miniaturized production facilities, mobile production and filling facilities, strains resistant to multiple antibiotics, cruise missile dissemination system, and combination organisms.

By addressing every aspect of weapon production, from selection of new strains of organisms to the behavior of biological aerosols under every possible condition of climate and topography, through the genetic engineering of antibiotic resistance and the design of optimum dissemination and delivery systems, the former Soviet Union was able to envisage the achievement of a miniaturized mobile production and weapon-making capability invulnerable to clandestine monitoring, invasive arms inspection, or attack in the event of war (because it was beyond identification); agents precisely matched to particular scenarios and human targets and incapable of being treated; a variety of dissemination systems, including cruise missiles; agents resistant to degradation by heat, light, cold, UV radiation, ionizing radiation, and various antibiotics; and dry formulations of agents capable of remaining viable in long-term storage.

By the time of the breakup of the former Soviet Union, from which the Russian Confederation emerged in 1992, much had been achieved and war mobilization plans were in place for the surge production of huge quantities of the agents mentioned earlier, as well as a number of others, such as Marburg virus. Of overwhelming importance has been the capability to undertake a strategic attack using plague or smallpox. Intercontinental ballistic missiles with MIRVed warheads containing plague were available for launch even before 1985, and SS-11 and SS-18 missiles have been mentioned in this connection. Concepts of use had been developed for each of the biological agents formally accepted into use by the army. For instance, the principal agents designated as tactical or operational for use on the battlefield were tularemia and Venezuelan equine encephalomyelitis, whereas anthrax and Marburg virus were nominated for attacking rear areas. The third category of agents comprised the highly transmissible agents smallpox and plague,

which were categorized as strategic weapons and destined for use against enemy population centers.

What happened after Vladimir Pasechnik (the former general director of Science Production Organisation Farmprhor and director of The All Union Scientific Research Institute of Ultra Pure Biopreparations in Leningrad (St. Petersburg) defected in 1989 constitutes a long and complex story, but in January 1991 the first-ever visit to Biopreparat facilities was undertaken, by a joint U.K./U.S. technical team, under a cloak of secrecy. After the subsequent defection of Kanadjan Alibekov (a former senior deputy director of Biopreparat) in 1992, the United States and the United Kingdom were certain enough that the offensive biological weapons program was continuing that they challenged the new Russian regime openly about it as late as 1993. By then substantial changes had taken place within Biopreparat, and today a concerted effort is under way to help the Russians civilianize these former biological weapons R&D establishments. However, questions remain about the Russian program: What happened to the part of the program in Ministry of Defence facilities that western experts have been unable to visit? What happened to plans detailing every aspect of production and deployment? What happened to the Ovchinnikov bioregulator program? What happened to the thousands of personnel involved in the Biopreparat program? What happened to the R&D centered on anticrop, antiplant, and antivivivox biological weapons? What happened to the stocks of seed cultures of biological weapons agents designed to be used to fuel the mobilized production of weapons? Was there space-based biological weapons capability? Was there any human genetics-related biological weapons research?

Despite the passage of nearly 10 years, the fundamental change in political structure of Russia, the extreme economic upheaval and budget restrictions, the reorientation of Biopreparat's work, and the help and support given by the West to civilianize programs and stop the transfer of technology and scientists into illegal biological weapons programs, the capability of the old Russian Ministry of Defence sites remains largely unknown.

Iraq

Iraq has stated that its biological weapons program dates to at least 1974. It was carried out

in great secrecy, after the Biological and Toxin Weapons Convention had been signed. The program was first conducted in an ostensibly civilian organization called the State Organization for Trade and Industry until this was superseded by the Military Industrial Commission. As with all other major military programs, biological weapons R&D was able to call upon many of its leading scientists who undertook undergraduate or postgraduate training in the west. Much of what happened between the supposed inception of the program in 1974 and the establishment of a group of biologists within the Al-Muthanna chemical weapons complex in 1984 is unknown.

In 1987, the Al-Muthanna research group was transferred to the Al-Satman facility, and work was expanded to include the investigation of fungal and antipant agents; 1988 saw the establishment of the Al-Hakam Factory, an industrial-scale production facility designed to produce anthrax and botulinum toxin for filling into weapons. This project was completed quickly by using equipment from nominally civilian facilities, such as those used to produce vaccines; the factory itself produced biological agent, which was filled into weapons and deployed in late 1990. The program was further expanded in 1990 when viruses were added to the range of agents under development and production capacity was enhanced by the acquisition and integration of civilian biotechnology facilities by the Military Industrial Commission.

According to the Iraqis, the program was terminated in 1991, after the adoption of UN SCR687, and agents, weapons, munitions, and documents were destroyed. However, the United Nations Special Commission (UNSCOM) believes that from 1991 to 1995 Iraq actively preserved biological weapons capability.

The Agents, Weapons, and Means of Delivery

The UNSCOM belief that three biological agents were filled into weapons is supported by Iraqi statements concerning the filling of munitions and their deployment ready for delivery. For one of these agents, *Botulinum* toxin, UNSCOM also possesses objective evidence; the other two were probably anthrax and *Clostridium perfringens* spores. Approximately

380,000 liters of *Botulinum* toxin were manufactured, along with 84,250 liters of anthrax spores and 3,400 liters of *C. perfringens* spores. In addition, 2,200 liters of aflatoxin were produced. All these figures represent preconcentration totals and may be underestimates. Ricin toxin and the antipant agents wheat bunt and corn smut were also produced. Camel pox is known to have been under development as well. This disparate list of biological agents, which at first seems to contain substances not previously conceived as potential offensive biological weapons agents, on closer inspection reveals a rationale based on the possession of a multipotent arsenal having lethal, incapacitating, oncogenic, ethnic, economic, terror, and variable time-onset capabilities. In addition, these agents are capable of being used to attack people through the lungs and the skin, as well as with carriers such as triethylamine, CN or CS, or as a toxic coating in fragmentation weapons.

Agents were filled into various weapons for dissemination. By the end of 1990, according to Iraqi statements, 25 SCUD/Al-Husseini missiles were readied for use with biological weapons warheads (each carrying 145 liters of agent) and deployed for action. At least 160 R400 retarded aerial bombs, carrying the distinctive black-stripe identification around them, may also have been filled with 90-liter charges of *Botulinum* toxin and ready for use. UNSCOM has evidence to corroborate the Iraqi claim. The Iraqis also intended to fill R400 bombs with anthrax and aflatoxin. Originally designed and filled with chemical agents, 155-mm shells were also tested with a ricin toxin fill. At least three fuel drop tanks were completely modified and fitted with Venturi mechanisms to facilitate aerosol release, for dispersal of 2,200-liter loads of anthrax and possibly *Botulinum* toxin, using F1 aircraft as the delivery means.

Postscript

UNSCOM has no confidence that Iraq has abandoned its biological weapons program. The true scale and scope of the Iraqi biological weapons program are, despite all UNSCOM's efforts, still not known.

Dr. Davis is consultant in Pharmaceutical Medicine and Applied Physiology and director of the ORAQ Consultancy Ltd., United Kingdom.



NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

FOREIGN POLICY RESEARCH INSTITUTE

Providing Ideas in Service to Our Nation Since 1955

War and Technology

by Alex Roland

February 2009
Vol. 14, No. 2

Alex Roland is professor of history at Duke University. This essay is based on his presentation at “Teaching the History of Innovation,” a two-day history institute for teachers held October 18-19, 2008. The Institute was hosted by the Ewing Marion Kauffman Foundation in Kansas City, Missouri and webcast. See www.fpri.org/education/innovation for videocasts and texts of lectures.

The History Institute for Teachers is co-chaired by David Eisenhower and Walter A. McDougall. Core support is provided by the Annenberg Foundation and Mr. H.F. Lenfest; funding for the innovation program is provided by the Ewing Marion Kauffman Foundation. The next history weekends are Teaching the Nuclear Age, March 28-29, 2009, at the Atomic Testing Museum in Las Vegas, and America’s Wars, Part II, May 2-3, at the First Division Museum, Wheaton, Illinois.

Military technology often seems to be the dark side of innovation, the Mr. Hyde roaming the back alleys of civilization for opportunities to work his worst on society. Its foundational figure in Western civilization is the Greek Hephaestus (whose counterpart was the Roman “Vulcan”), the only god to have been lame and misshapen. But countless inventors and innovators, from Alfred Nobel to Robert Boyle, thought of weapons positively. They believed that they could banish the scourge of war, or at least restrain its excesses, if they could only invent the ultimate weapon, the instrument so horrible that no one would dare use it.

More than six decades into the nuclear age, there is growing evidence that the hydrogen bomb may prove to be the long-sought war-stopper.[1] But should that be the case, it will run counter to the sorry record of prior human civilization, when each new instrument of

war contributed to the carnage without altering the human nature Thucydides believed to be at the heart of war. Melvin Kranzberg, a co-founder of the Society for the History of Technology and the founding editor of its journal, *Technology and Culture*, was fond of observing that technology is neither good nor bad, nor is it neutral. Technology in essence is a process of manipulating the material world for human purposes. Whether it does good or ill depends not on the technology itself but on what humans choose to do with it.

Military machines and instruments can nonetheless be understood using the same concepts and categories that scholars apply to technology in general. Below I put forward four propositions about military technology, but the principles at work could be applied as easily in any realm of technological endeavor. They sometimes have a special relevance or poignancy when applied to war, but they say more about the nature of technology than they do about the nature of war.

In addition to their heuristic value, these concepts also have pedagogical utility. They can help demystify the arcane and often secretive world of military research and development and also clarify the impact on society of all complex technological systems. They offer students a set of conceptual tools for thinking about change in warfare over time and the role that technological innovation has played in that process.

My propositions are these: (1) technology, more than any other outside force, shapes warfare; and, conversely, war (not warfare) shapes technology. (2) Military technology is, however, *not* deterministic. Rather, (3) technology opens doors. And, finally, (4) these characteristics of military technology are easier to see in the modern period than previously, though they have always been at work.

Technology Shapes Warfare

Technology shapes *warfare*, not war. War is timeless and universal. It has afflicted virtually every state known to human history. Warfare is the conduct of war. It is the clash of arms or the maneuver of armed forces in the field. It entails what military professionals call operations, whether or not the opposing forces actually unleash their organized violence on one another. War is a condition in which a state might find itself; warfare is a physical activity conducted by armed forces in the context of war. Of course, many kinds of group violence, from gang fights to terrorism, might display some or all of the characteristics of warfare without rising to this definition of war, but more often than not these violent conflicts use instruments of war. To understand the technology of warfare is to understand the technology of most public violence.

Wording is also important in articulating exactly what impact technology has on warfare. A number of verbs suggest themselves. Technology defines, governs, or circumscribes warfare. It sets the stage for warfare. It is the instrumentality of warfare.

The most important verb describing the impact of technology on warfare is that it *changes* warfare. Technology has been the primary source of military innovation

throughout history. It drives changes in warfare more than any other factor. Consider a simple thought experiment. Sun Tzu and Alexander the Great are brought back to life and assigned to lead coalition forces in Afghanistan in 2008. These near contemporaries from the fourth century BCE would understand almost everything they would need to know. Alexander actually fought in Afghanistan, and Sun Tzu (if such a person really existed) fought in comparably mountainous terrain in China.[2] Both were masters of strategy and tactics. What came to be called the “principles of war” are simply the tacit knowledge that all successful commanders throughout history have carried around in their bank of experience: an understanding of intelligence, surprise, maneuver, command and control, concentration of force, unity of command, terrain, etc. Even Clausewitz’s seminal contributions to military art and science—chance, violence, the “fog of war,” and “friction”—were concepts that Alexander and Sun Tzu knew by different names.

The only modern tool of command they would not know and could not readily assimilate would be the technology of war. Airplanes, missiles, tanks, drones, satellites, computers, GPS, and all the remaining panoply of the modern high-tech battlefield would be incomprehensible to them. A sergeant from their operations staff could exploit these resources more fully and effectively than either of our great captains. Sun Tzu and Alexander would be incompetent on the modern battlefield.

The point is even more obvious in humankind’s other two fields of battle—the sea and the air—to say nothing of space, perhaps the battlefield of the future. Naval warfare does not occur without ships, which, through most of human history, were the most complex of human technological artifacts. Of course the same is true of planes for air warfare, missiles for strategic warfare, and spacecraft for star wars. In each case, the vehicle defines the warfare. Horatio Nelson, perhaps the greatest naval commander of all time, would have been powerless to understand the strategy and tactics of World War II’s air warfare in the Pacific or submarine warfare in the Atlantic. The cat-and-mouse contest of Soviet and American attack submarines in the Cold War would have been even more incomprehensible to him. He might have gone back in time and intuited the essence of galley warfare, but he could not command in the age of steam, let alone the nuclear age, without a solid grounding in modern science and technology.

The more modern, or postmodern, the warfare becomes, the more the generalization holds true. Technology defines warfare. Air warfare was not even possible before the twentieth century, save for the vulnerable and inefficient reconnaissance balloons that were pioneered in Europe and America in the nineteenth century. In the twenty-first century, air warfare ranges from strategic bombing to close air support of ground troops to dog fights for air superiority to pilotless drones that carry the eyes and ears, and sometimes the ordnance, of operators hundreds, even thousands, of miles away. The U.S. boasts a missile defense installation that can stop the unstoppable, an intercontinental ballistic missile. Space-faring nations flirt with anti-satellite weapons launched from earth and even the prospect of space-based weapons to fight one another and threaten the earth below. Air warfare differs from naval warfare, not because the strategy and tactics of conflict in those realms differs, but because planes differ from ships. And, of course,

both differ from tanks and rockets and satellites. Each technology shapes, defines, circumscribes, and governs a new kind of warfare.

Nor is it just the evolution of weaponry that changes warfare. It is the distribution of the weaponry. Throughout history, states have usually fought one another in weapons symmetry.[3] In the first Gulf War, for example, Saddam Hussein attempted to defeat a conventional, industrialized, mechanized American army with a conventional, industrialized, mechanized Iraqi army. The quality and quantity of the American technology prevailed. In the second Gulf War, however, the insurgents resorted to asymmetrical warfare, fighting the high-tech American arsenal with low-tech instruments of assassination, sabotage, and terror. Only when the United States adjusted its technology to meet the new threat did the enemy tactics lose their edge. Of course training, morale, numbers, will, and politics also contributed to the outcome in Iraq, but the nature of the technology set the stage for the struggle.

Technology Does *Not* Determine Warfare

However much technology may *change* warfare, it never *determines* warfare—neither how it will be conducted nor how it will turn out. Technology presides in warfare, but it does not rule.

The whole notion of “technological determinism” is a red herring.[4] Humans can always resist the historical forces surrounding them. To believe in determinism is to believe in inevitability. This begs the question, “Why”? What historical force or law pushes events to some inescapable outcome? In hindsight, events may appear predetermined or inevitable, but nothing in human activity can be predicted with certainty.

Think about the instances in history when technology appeared to determine the nature and even the result of warfare. Chariots were perhaps the most dominant instrument of warfare before nuclear weapons. Indeed, historian William H. McNeill has called them the superweapon of their day.[5] When they appeared in the Levant in the eighteenth century BCE, they swept all before them. From Egypt to Mesopotamia, states either adopted chariots or ceased to compete in interstate war. The chariot craze bred an international chariot aristocracy, the Maryannu, who sold their services to the highest bidder.[6] States built up enormous chariots corps with attendant supply and maintenance trains, culminating in the battle of Kadesh in 1275 BCE, when the contending Egyptian and Hittite forces committed an estimated 5,000 chariots to a cataclysmic but ultimately indecisive day of battle. Western warfare through most of the second millennium BCE was chariot warfare. The chariot defined, drove, governed, circumscribed ground warfare.

And then it was gone. Within a century after the Armageddon at Kadesh, the chariot disappeared as the dominant technology of Levantine warfare. Just as there is no sure evidence of where the chariot came from and why it ruled, so is its fall from dominance a mystery. Robert Drews notes that it lost power in “The Catastrophe,” the wave of wars, raids, and forced migrations that swept the eastern Mediterranean around 1200 BCE.[7]

William McNeill believes that the introduction of iron weapons at just this time gave infantry new power to stand up to chariots.[8] Another possible explanation is state bankruptcy brought on by the arms race in chariots and the horses to pull them. Still another is a change in infantry tactics, perhaps coupled with McNeill's iron weapons. In any case, the apparent determinism of the chariot evaporated.

Countless other examples through history of seemingly irresistible weapons leading to inevitable triumph have similarly risen and fallen in their turn, from gunpowder through the "*Dreadnought* revolution" and strategic bombing to the recent enthusiasm for the "revolution in military affairs," a technological superiority that was to have given the U.S. unassailable military prowess.

The Open Door

A better conceptual model for the technology of war is "the open door." This metaphor was introduced by medieval historian Lynn White, Jr., in his classic study *Medieval Technology and Social Change* (Oxford, 1962). Seeking to demonstrate that medieval society spawned its share of technological innovation, White presented a series of interrelated case studies. One revisited and refined the discredited claim by Heinrich Brunner that the appearance of the heavily armed and armored mounted knight on the battlefields of eighth-century Europe had bred feudalism. Brunner had imagined that Charles Martel first conceived the scheme of a feudal array and the social, political, and economic system to sustain it at or immediately following the battle of Poitiers in 732, when his posse of mounted warriors drove off Muslim raiders spilling into southern France from the Iberian Peninsula. But White showed that Martel had begun confiscating church property for distribution to a new class of mounted warriors before the battle of Poitiers. What, then, asked White, might have inspired Martel, if not his victory over the lightly armed and armored Muslim mounted warriors? White's answer was the stirrup, an Asian innovation just making its first appearance in Europe about the time of Poitiers. White imagined that this technology allowed the heavily armed and armored mounted knight to lean into his lance and overwhelm mounted and unmounted warriors alike with irresistible force.

White's argument was widely and roundly attacked, especially by Marxist historians. Most of his critics accused him of technological determinism, of arguing that the stirrup produced feudalism. But White had gone out of his way to avoid any such claim. He called the stirrup a "catalyst." It did not create feudalism out of whole cloth. Rather, when it was added to the complex soup of medieval society, feudalism precipitated out. Other societies with different ingredients and different chemistries would produce different residues. Technology does not determine outcomes, said White, it opens doors. People must decide if they want to pass through. The availability of the stirrup in Europe did not mean that Martel would adopt it to make the heavily armed and armored mounted knight the mainstay of an emergent military system.

The "open door" is a powerful conceptual tool for thinking about all technology, especially military technology. It adds what most accounts of technological innovation

lack: human agency. Humans must decide if they are going to, or can, take up a given military innovation. And they must adapt it to their circumstances. Technology is a possibility, not an imperative. The varieties of technology—think, for example, of the different models of automobiles in the world—testify to the countless contexts in which people will apply the same fundamental innovation with differing results.[9] In the years between the world wars, for example, the U.S. and Britain, geographically isolated from continental Europe, developed strategic bombers with which to project their military power, while the major continental powers concentrated on fighter aircraft to contend with each other for air superiority over the battlefields in their back yards.

What White did *not* discern was that human agency intervenes not once, but twice, at the moment of innovation. When discovering who chooses to walk through a door opened by technological innovation, it is equally important to understand who opened it for them. People invent and innovate. People open the door. Often there is a relationship between those who open the door and those who pass through it. In modern military experience we often think of that relationship as the military-industrial complex. This byproduct of the Cold War reminds us that the door can be swinging, even revolving. People who are anxious to pass through may hire agents to open the door for them. Likewise, those who figure out how to open doors may entice others to pass through. None of this makes the technology itself deterministic, but it can make the social system attending the door self-replicating and self-sustaining. In time, the participants may pass into realms that they would not otherwise have chosen to enter and that do not serve their interests.[10] The reasons for their decisions, even the bad ones, however, lie not in the technology but in the personal, political, economic—in sum, contextual—forces already at work. The catalyst precipitates the consequences dictated by the ingredients and the chemistry.

This raises my fourth point.

Modern Military Technology Is Different

Modern military technology is not different in kind, but in degree. World War II was the first war in history in which the weapons in use at the end of the war differed significantly from those employed at the outset. The atomic bomb is the most obvious example, but the list of military technologies introduced between 1939 and 1945 includes as well jet aircraft, guided missiles, microwave radar, and the proximity fuse, to name just a few. Some military leaders concluded from this experience that industrial production had won the world wars but military innovation would win the next war. Especially in the U.S., the military establishment began to institutionalize research and development, adopting from industry a kind of planned obsolescence that would keep American armed forces a generation ahead of their potential foes. They created what President Dwight Eisenhower called in his farewell address a “military-industrial complex,” a perpetual arms race, not necessarily with any particular enemy, but with the status quo.

The introduction of systematic, institutionalized innovation makes modern military technology seem radically different from all that went before.[11] That difference is

simultaneously real and illusory. The reality stems from the accelerated pace of technological change in the modern world and an unprecedented mastery of energy and materials ranging across a dimensional scale from nanotechnology to floating cities like the modern aircraft carrier. The illusion arises from our growing inability to think of war in non-material terms. Modern commanders can hardly imagine how their predecessors thought about science and technology. A career officer in today's armed forces expects the arsenal at his or her disposal to change constantly over the course of a career. Before the second half of the twentieth century, however, commanders fully expected to retire with the same instruments they took up in their apprenticeship. Personal arms might even pass from father to son. Some innovation intruded on this static picture in the late nineteenth and early twentieth century, but nothing like the sustained hothouse environment of today's military arsenals.

Even the terms "science" and "technology" are modern, both coined in the nineteenth century. Before these conceptual categories took hold of the modern consciousness, premodern commanders thought of their armies and navies in terms of men (human capital) and material (arms and armor, forts and roads, food and ammunition). Improvements in any of these areas were made not by scientists and engineers, but by craftsmen with little formal schooling. "Engineers" in the premodern world were individuals who built and operated "engines" of war, i.e., ballistae and catapults. There was high-quality steel long before its composition was revealed by crystallography in the 1920s, but it was produced by artisans who passed their techniques from generation to generation through apprenticeship, not by industrialists whose staffs of scientists prescribed formulas that would produce steel of requisite characteristics. A handful of premodern geniuses, such as Michelangelo and Leonardo da Vinci, mastered art and engineering to imagine weapons that were centuries ahead of their time. Mechanics operated machines of war. Sailors were always mechanics operating the most complex machines of their age, be it the galley of classical Mediterranean warfare or the fully-rigged, side-gunned sailing ship of the line in the early modern world. Architects designed and erected fortifications, probably the most influential military technology before gunpowder. These marvels of what we would now call civil engineering—the Great Wall of China, the walls of Constantinople, the Roman limes—shaped countless conflicts throughout history and sometimes ensured that conflicts never happened. Other premodern builders oversaw civil works with equally important military implications, such as the Roman road network, or Caesar's bridge across the Rhine, or the earthen ramp at Masada.

The material resources that these premodern materialists delivered to their armies and navies were not seen as the result of abstract enterprises like science and technology. The producers of these wonders were simply practicing what Lewis Mumford called "technics," from the Greek *techne*.^[12] It is related to *episteme* in something like the relationship between modern technology and science. One was about doing, the other about knowing. One was learned in apprenticeship, the other was gained by the study of knowledge accumulated in a canon. But just as modern technology is more complex and independent than "applied science," so is technics more subtle than simply craft knowledge. It is perhaps more helpful to think of premodern producers of military

instruments as “improvers,” the generic term that Robert Friedel applies to all those people in the last millennium of Western history who have manipulated the material world in search of better ways to do whatever it is that people choose to do.[13]

In short, the tools of war have been evolving slowly throughout the course of human history, but only in the modern world has there been an institutionalized and rationalized mechanism for continuously and systematically innovating military technology. Some tantalizing hints from the ancient and classical world place our modern world in bold relief. One anonymous author from classical Greece offered the opinion that the only real utility for third-order equations was to compute the trajectory of ballistae.[14] In this instance, at least, military technology really was applied science. Dionysius I, tyrant of Syracuse in the early fourth century BCE, recruited knowers and doers from around the Mediterranean to work in his arsenal to develop new machines of war, perhaps the first instance of a research and development laboratory.[15] A Syrian engineer by the name of Kallinikos delivered “Greek fire” to the Byzantines in hopes that they would use it to defeat the Muslims.[16] The Byzantines made of this the only truly secret weapon of the ancient and medieval world. “Secret weapons” as we now think of them were an invention of early modern Europe.[17]

Kallinikos provides a fitting ending for this account. When he is mentioned at all in history books, he usually appears as a kind of “deus ex machine,” an unknown and unknowable historical actor who delivers a war-changing weapon that in turns changes the course of history. Such representations give rise to the belief that “technological determinism” is at work. But in all such instances, it is well to think of a door. Who opened it and who passed through? Who was Kallinikos? How did he come by this formula, and why did he offer it to the Byzantines? Why did the Byzantines take up his new weapon? And why did they treat it as a state secret? “Technological determinism” is a distraction, a rhetorical device that diverts attention from the real historical questions that surround the relationship between war and technology. That relationship is defining, subtle, and evolving. Now, more than ever, it drives innovation in warfare.

Notes

1. ▲ John Lewis Gaddis, *The Long Peace: Inquiries into the History of the Cold War* (New York: Oxford University Press, 1987).
2. ▲ One the historicity of Sun Tzu, see Sunzi, Ralph Sawyer, and Mei-chün Sawyer, *The Art of War* (Boulder: Westview Press, 1994).
3. ▲ Robert O’Connell, *Of Arms and Men: A History of War, Weapons, and Aggression* (New York: Oxford University Press, 1989), 8-9 et passim.
4. ▲ Alex Roland, “Was the Nuclear Arms Race Deterministic?” paper prepared for the 50th Anniversary Workshop of the Society for the History of Technology, Washington, DC, 18 Oct. 2007.
5. ▲ William H. McNeill, *The Rise of the West: A History of the Human Community* (Chicago: University of Chicago Press, 1991), 104–106
6. ▲ See Arthur Cotterell, *Chariot: The Astounding Rise and Fall of the World’s First War Machine* (London: Pimlico, 2004), 67–68, 86; R. T. O’Calligan, “New

- Light on the Maryannu as ‘Chariot Warrior,‑” *Jahrbuch für kleinasiatische Forschung* 1 (1950-51): 309–24
7. ▲ Robert Drews, *The End of the Bronze Age: Changes in Warfare and the Catastrophe CA. 1200 B.C.* (Princeton University Press, 1993), 209–14, and *passim*.
 8. ▲ McNeill, *Rise of the West*, 117–18
 9. ▲ George Basalla, *The Evolution of Technology* (New York: Cambridge University Press, 1988).
 10. ▲ Donald A. MacKenzie, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge, MA: MIT Press, 1990).
 11. ▲ And not just military technology. As Alfred North Whitehead observed many years ago, “the greatest invention of the nineteenth century was the invention of the means of invention.” Alfred North Whitehead, *Science and the Modern World* (New York: Free Press, [1925] 1953), 96
 12. ▲ Lewis Mumford, *Technics and Civilization* (1932).
 13. ▲ Robert Friedel, *A Culture of Improvement: Technology and the Western Millennium* (Cambridge, MA: MIT Press, 2007)
 14. ▲ J. G. Landels, *Engineering in the Ancient World* (Berkeley: University of California Press, 99–132).
 15. ▲ Brian Caven, *Dionysius: War-lord of Sicily* (New Haven: Yale University Press, 1990), 90-97
 16. ▲ J. Haldon and M. Byrne, “A Possible Solution to the Problem of Greek Fire,” *Byzantinische Zeitschrift* 70 (1977): 91–99.
 17. ▲ Pamela Long and Alex Roland, “Military Secrecy in Antiquity and Early Medieval Europe: A Critical Reassessment,” *History of Technology* 11 (1994): 259–90.



Stuxnet's Secret Twin

The real program to sabotage Iran's nuclear facilities was far more sophisticated than anyone realized.

- BY [Ralph Langner](#)
- *Foreign Policy*, NOVEMBER 21, 2013

Three years after it was discovered, Stuxnet, the first publicly disclosed cyberweapon, continues to baffle military strategists, computer security experts, political decision-makers, and the general public. A comfortable narrative has formed around the weapon: how it attacked the Iranian nuclear facility at Natanz, how it was designed to be undiscoverable, how it escaped from Natanz against its creators' wishes. Major elements of that story are either incorrect or incomplete.

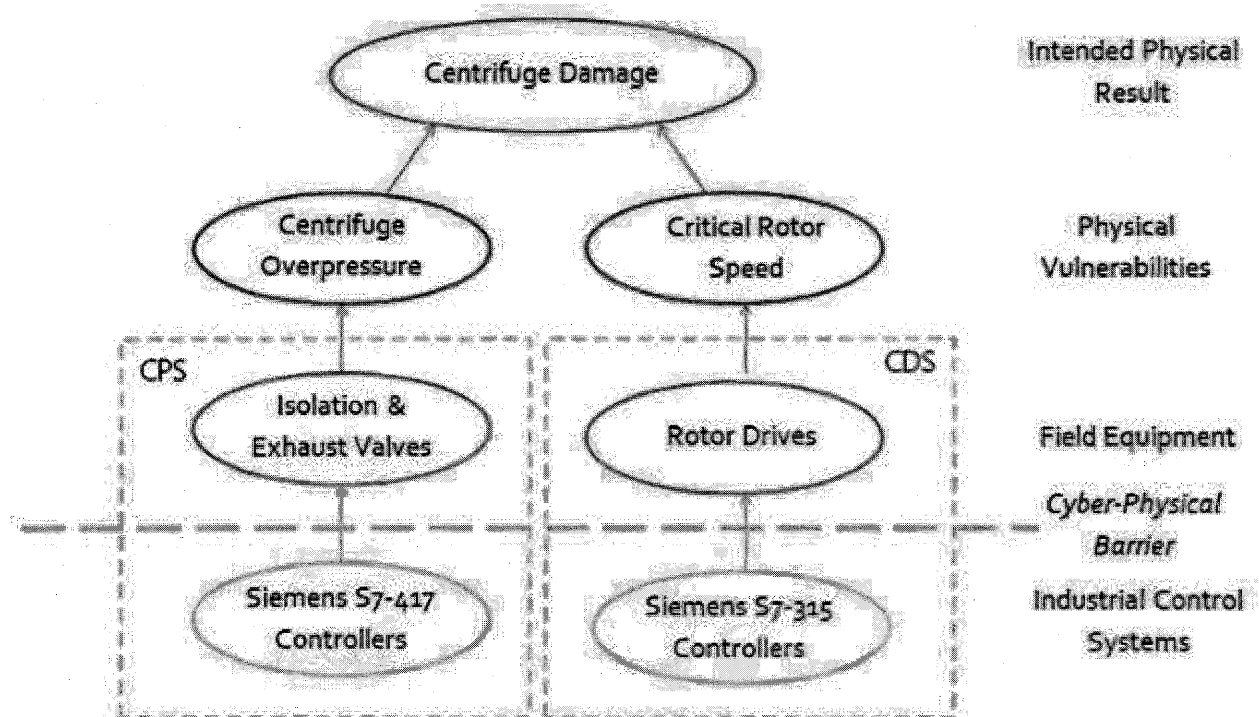
That's because Stuxnet is not really one weapon, but two. The vast majority of the attention has been paid to Stuxnet's smaller and simpler attack routine -- the one that changes the speeds of the rotors in a centrifuge, which is used to enrich uranium. But the second and "forgotten" routine is about an order of magnitude more complex and stealthy. It qualifies as a nightmare for those who understand industrial control system security. And strangely, this more sophisticated attack came *first*. The simpler, more familiar routine followed only years later -- and was discovered in comparatively short order.

With Iran's nuclear program back at the center of world debate, it's helpful to understand with more clarity the attempts to digitally sabotage that program. Stuxnet's actual impact on the Iranian nuclear program is unclear, if only for the fact that no information is available on how many controllers were actually infected. Nevertheless, forensic analysis can tell us what the attackers *intended* to achieve, and how. I've spent the last three years conducting that analysis -- not just of the computer code, but of the physical characteristics of the plant environment that was attacked and of the process that this nuclear plant operates. What I've found is that the full picture, which includes the first and lesser-known Stuxnet variant, invites a re-evaluation of the attack. It turns out that it was far more dangerous than the cyberweapon that is now lodged in the public's imagination.

In 2007, an unidentified person submitted a sample of code to the computer security site [VirusTotal](#). It later turned out to be the first variant of Stuxnet -- at least, the first one that we're aware of. But that was only realized five years later, with the knowledge of the second Stuxnet variant. Without that later and much simpler version, the original Stuxnet might still today sleep in the archives of anti-virus researchers, unidentified as one of the most aggressive cyberweapons in history. Today we now know that the code contained a

payload for severely interfering with the system designed to protect the centrifuges at the Natanz uranium-enrichment plant.

Stuxnet's later, and better-known, attack tried to cause centrifuge rotors to spin too fast and at speeds that would cause them to break. The "original" payload used a different tactic. It attempted to overpressurize Natanz's centrifuges by sabotaging the system meant to keep the cascades of centrifuges safe. "Protection systems" are used anywhere where abnormal process conditions can result in equipment damage or threaten the health of operators and the environment. At Natanz, we see a unique protection system in place to enable sustained uranium enrichment using obsolete and unreliable equipment: the IR-1 centrifuge. This protection system is a critical component of the Iranian nuclear program; without it, the IR-1s would be pretty much useless.

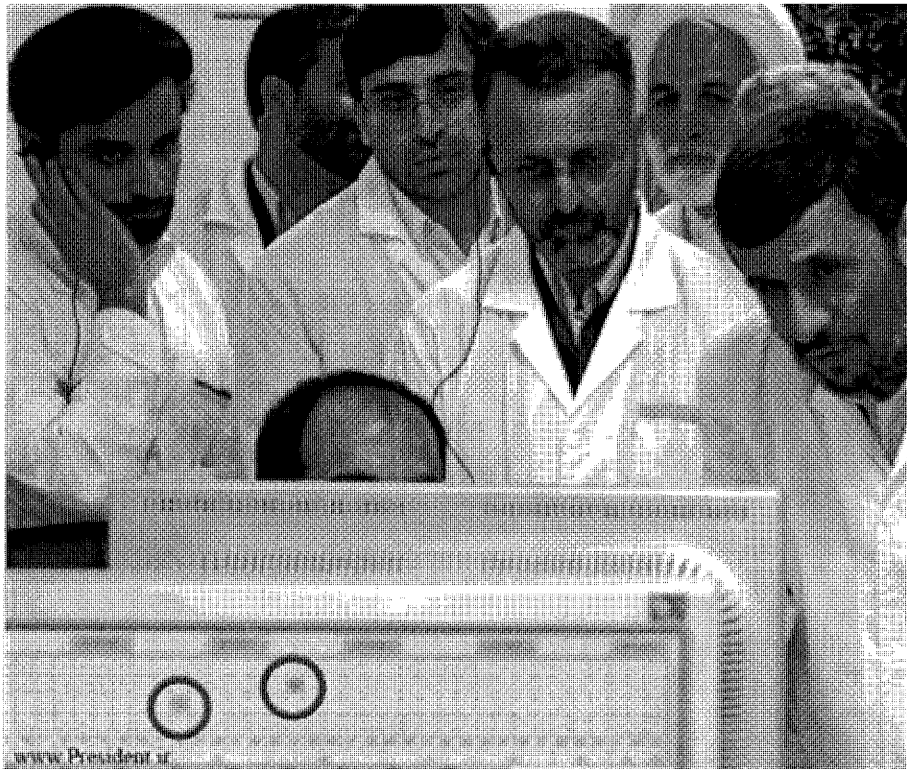


The IR-1 centrifuge is the backbone of Iran's uranium-enrichment effort. It goes back to a European design from the late 1960s and early 1970s that was stolen and slightly improved by Pakistani nuclear trafficker A.Q. Khan. The IR-1 is an all-metal design that *can* work reliably. That is, if parts are manufactured with precision and critical components such as high-quality frequency converters and constant torque drives are available. But the Iranians never managed to get a high degree of reliability from the obsolete design. So they had to lower the operating pressure of the centrifuges at Natanz. Lower operating pressure means less mechanical stress on the delicate centrifuge rotors, thereby reducing the numbers of centrifuges that have to be put offline because of rotor

damage. But less pressure means less throughput -- and thus less efficiency. At best, the IR-1 was half as efficient as its ultimate predecessor.

As unreliable and inefficient as the IR-1 is, it offered a significant benefit: Iran managed to produce the antiquated design at industrial scale. Iran compensated reliability and efficiency with volume, accepting a constant breakup of centrifuges during operation because they could be manufactured faster than they crashed. But to make it all work, the Iranians needed a bit of a hack. Ordinarily, the operation of fragile centrifuges is a sensitive industrial process that doesn't tolerate even minor equipment hiccups. Iran built a cascade protection system that allows the enrichment process to keep going, even when centrifuges are breaking left and right.

At the centrifuge level, the cascade protection system uses sets of three shut-off valves, installed for every centrifuge. By closing the valves, centrifuges that run into trouble -- indicated by vibration -- can be isolated from the rest of the system. Isolated centrifuges are then run down and can be replaced by maintenance engineers while the process keeps running.



Then-President Mahmoud Ahmadinejad looks at SCADA screens in the control room at Natanz in 2008. The screen facing the photographer shows that two centrifuges are

isolated, indicating a defect, but that doesn't prevent the respective cascade from continuing operation (red highlighting added).

But the isolation valves can turn into as much of a problem as a solution. When operating basically unreliable centrifuges, one will see shut-offs frequently, and maintenance workers may not have a chance to replace damaged centrifuges before the next one in the same enrichment stage gets isolated. Once multiple centrifuges are shut off within the same stage, operating pressure -- the most sensitive parameter in uranium enrichment using centrifuges -- will increase, which can and will lead to all kinds of problems.

The Iranians found a creative solution for this problem -- basically another workaround on top of the first workaround. For every enrichment stage, an exhaust valve is installed that allows pressure to be relieved if too many centrifuges within that stage get isolated, causing pressure to increase. For every enrichment stage, pressure is monitored by a sensor. If the pressure exceeds a certain threshold, the exhaust valve is opened, and overpressure is released.

The system might have kept Natanz's centrifuges spinning, but it also opened them up to a cyberattack that is so far-out, it leads one to wonder whether its creators might have been on drugs.

Natanz's cascade protection system relies on Siemens S7-417 industrial controllers to operate the valves and pressure sensors of up to six cascades, or groups of 164 centrifuges each. A controller can be thought of as a small embedded computer system that is directly connected to physical equipment, such as valves. Stuxnet was designed to infect these controllers and take complete control of them in a way that previous users had never imagined -- and that had never even been discussed at industrial control system conferences.

A controller infected with the first Stuxnet variant actually becomes decoupled from physical reality. Legitimate control logic only "sees" what Stuxnet wants it to see. Before the attack sequence executes (which is approximately once per month), the malicious code is kind enough to show operators in the control room the physical reality of the plant floor. But that changes during attack execution.

One of the first things this Stuxnet variant does is take steps to hide its tracks, using a trick straight out of Hollywood. Stuxnet records the cascade protection system's sensor values for a period of 21 seconds. Then it replays those 21 seconds in a constant loop during the execution of the attack. In the control room, all appears to be normal, both to human operators and any software-implemented alarm routines.

Then Stuxnet begins its malicious work. It closes the isolation valves for the first two and last two enrichment stages. That blocks the outflow of gas from each affected cascade and, in turn, raises the pressure on the rest of the centrifuges. Gas centrifuges for uranium enrichment are extremely sensitive to increases of pressure above near vacuum. An increase in pressure will result in more uranium hexafluoride getting into the centrifuge,

putting higher mechanical stress on the rotor. Rotor wall pressure is a function of velocity (rotor speed) and operating pressure; more gas being pressed against the rotor wall means more mechanical force against the thin tube. Ultimately, pressure may cause the gaseous uranium hexafluoride to solidify, thereby fatally damaging centrifuges.

The attack continues until the attackers decide that enough is enough, based on monitoring centrifuge status. Most likely, they would use vibration sensors, which let them abort a mission before the matter hits the fan. If catastrophic destruction is intended, one simply has to sit and wait. But in the Natanz case, causing a solidification of process gas would have resulted in simultaneous destruction of hundreds of centrifuges per infected controller. While at first glance this might sound like a goal worthwhile achieving, it would also have blown the attackers' cover; the cause of the destruction would have been detected fairly easily by Iranian engineers in postmortem analysis. The implementation of the attack with its extremely close monitoring of pressures and centrifuge status suggests that the attackers instead took great care to *avoid* catastrophic damage. The intent of the overpressure attack was more likely to increase rotor stress, thereby causing rotors to break early -- but not necessarily during the attack run.

Nevertheless, the attackers faced the risk that the attack would not work at all because the attack code is so overengineered that even the slightest oversight or any configuration change would have resulted in zero impact or, worse, in a program crash that would have been detected by Iranian engineers quickly.

The results of the overpressure attack are unknown. Whatever they were, the attackers decided to try something different in 2009.

This new Stuxnet variant was almost entirely different from the old one. For one thing, it was much simpler and much less stealthy than its predecessor. It also attacked a completely different component of the Natanz facility: the centrifuge drive system that controls rotor speeds.

This new Stuxnet spread differently too. The malware's earlier version had to be physically installed on a victim machine, most likely a portable engineering system, or it had to be passed on a USB drive carrying an infected configuration file for Siemens controllers. In other words, it needed to be disseminated deliberately by an agent of the attackers.

The new version self-replicated, spreading within trusted networks and via USB drive to all sorts of computers, not just to those that had the Siemens configuration software for controllers installed. This suggests that the attackers had lost the capability to transport the malware to its destination by directly infecting the systems of authorized personnel, or that the centrifuge drive system was installed and configured by other parties to which direct access was not possible.

What's more, Stuxnet suddenly became equipped with an array of previously undiscovered weaknesses in Microsoft Windows software -- so-called "zero day" flaws

that can fetch hundreds of thousands of dollars on the open market. The new Stuxnet also came equipped with stolen digital certificates, which allowed the malicious software to pose as legitimate driver software and thus not be rejected by newer versions of the Windows operating system.

All this indicates that a new organization began shaping Stuxnet -- one with a stash of valuable zero days and stolen certificates. In contrast, the development of the overpressure attack can be viewed as the work of an in-group of top-notch industrial control system security experts and coders who lived in an exotic ecosystem quite remote from standard IT security. The overspeed attacks point to the circle widening and acquiring a new center of gravity. If Stuxnet is American-built -- and, according to published reports, it most certainly is -- then there is only one logical location for this center of gravity: Fort Meade, Maryland, the home of the National Security Agency.

But the use of the multiple zero days came with a price. The new Stuxnet variant was much easier to identify as malicious software than its predecessor was, because it suddenly displayed very strange and very sophisticated behavior. In comparison, the initial version looked pretty much like a legitimate software project for Siemens industrial controllers used at Natanz; the only strange thing was that a copyright notice and license terms were missing. The newer version, equipped with a wealth of exploits that hackers can only dream about, signaled to even the least vigilant anti-virus researcher that this was something big, warranting a closer look.

Just like its predecessor, the new attack operated periodically, about once per month, but the trigger condition was much simpler. While in the overpressure attack various process parameters were monitored to check for conditions that might occur only once in a blue moon, the new attack was much more straightforward.

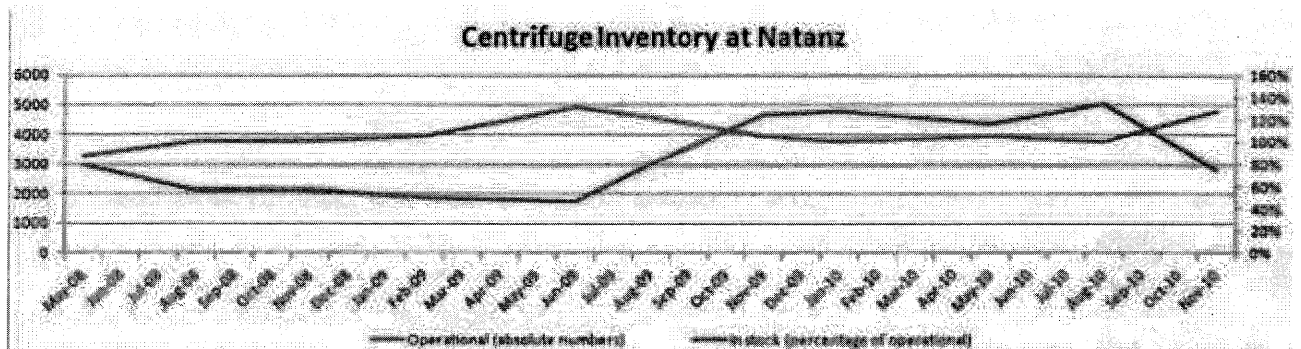
The new attack worked by changing rotor speeds. With rotor wall pressure being a function of process pressure and rotor speed, the easy road to trouble is to overspeed the rotors, thereby increasing rotor wall pressure. And this is what Stuxnet did. The normal operating speed of the IR-1 centrifuge is 63,000 revolutions per minute (rpm). Stuxnet increased that speed by a good one-third to 84,600 rpm for 15 minutes. The next consecutive run brought all centrifuges in the cascade basically to a stop (120 rpm), only to speed them up again, taking a total of 50 minutes. The IR-1 is a supercritical design, meaning that the rotor has to pass through so-called critical speeds before reaching normal operating speed. Every time a rotor passes through these critical speeds, also called harmonics, it can break.

If a single rotor did crack during an attack sequence, the cascade protection system would kick in to isolate and run down the respective centrifuge. But if multiple rotors were to crash -- a likely possible outcome -- Iranian operators would be left with the question of why all of a sudden so many centrifuges broke at once. Not that they didn't have enough new ones in stock for replacement, but unexplained problems like this are among any control system engineer's most frustrating experiences, usually referred to as *chasing a demon* in the machine.

At some point the attacks should have been recognizable by plant floor staff just by the old eardrum. Bringing 164 centrifuges or multiples thereof from 63,000 rpm to 120 rpm and getting them up to speed again would have been noticeable -- if experienced staff had been cautious enough to remove protective headsets in the cascade hall. It's another sign that the makers of this second Stuxnet variant had decided to accept the risk that the attack would be detected by operators.

Much has been written about the failure of Stuxnet to destroy a substantial number of centrifuges or to significantly reduce Iran's enriched-uranium production. While that is undisputable, it doesn't appear that either was the attackers' intention. If catastrophic damage had been caused by Stuxnet, that would have been by accident rather than on purpose. The attackers were in a position where they could have broken the victim's neck, but they chose continuous periodical choking instead. Stuxnet is a low-yield weapon with the overall intention of reducing the lifetime of Iran's centrifuges and making the Iranians' fancy control systems appear beyond their understanding.

Reasons for such tactics are not difficult to identify. When Stuxnet was first deployed, Iran had already mastered the production of IR-1 centrifuges at industrial scale. During the summer of 2010, when the Stuxnet attack was in full swing, Iran operated about 4,000 centrifuges, but kept another 5,000 in stock, ready to be commissioned. A one-time destruction of the Iranians' operational equipment would not have jeopardized that strategy, just like the catastrophic destruction of 4,000 centrifuges by an earthquake back in 1981 did not stop Pakistan on its way to getting the bomb. By my estimates, Stuxnet set back the Iranian nuclear program by two years; a simultaneous catastrophic destruction of all operating centrifuges wouldn't have caused nearly as big a delay.



Centrifuge inventory at Natanz between 2008 and 2010. Iran constantly kept a stockpile of at least 50 percent spare centrifuges, invalidating the idea that a simultaneous catastrophic destruction of all operating centrifuges would have meant the end of its nuclear ambitions.

The low-yield approach also offered added value. It drove Iranian engineers crazy, up to the point where they might have ultimately ended up in total frustration about their capabilities to get a stolen plant design from the 1970s running and to get value from

their overkill digital protection system. When comparing the Pakistani and Iranian uranium-enrichment programs, one cannot fail to notice a major performance difference. Pakistan basically managed to go from zero to successful low-enriched uranium production within just two years during shaky economic times, without the latest in digital control technology. The same effort took Iran over 10 years, despite the jump-start from Pakistan's A.Q. Khan network and abundant money from sales of crude oil. If Iran's engineers didn't look incompetent before, they certainly did during the time when Stuxnet was infiltrating their systems.

Legend has it that in the summer of 2010, while inflicting its damage on Natanz, Stuxnet "escaped" from the nuclear facility due to a software bug that came with a version update. While that is a good story, it cannot be true. Stuxnet propagated only between computers that were attached to the same local network or that exchanged files through USB drives. In other words, Stuxnet must have spread largely by human hands. But in these days of remote access by modem or via Internet virtual private networks, human hands can extend across continents.

Contractors serving at Natanz worked for other clients as well. And those contractors most likely carried their Stuxnet-infected laptop computers to their secondary clients and connected their laptops to the clients' "local" networks. Let's say they spread it to a cement plant. That cement plant then had other contractors, who in turn connected their mobile computers to the infected "local" network. Those computers carried the malware farther -- to another cement plant, maybe in another country. At some link in the chain, infected contractors or employees remotely accessed their machines, allowing the virus to travel over continents. All of a sudden, Stuxnet has made its way around the globe -- not because of the fact that billions of systems are connected to the Internet, but because of the trusted network connections that tunnel through the Internet these days. For example, remote maintenance access often includes the capability to access shared folders online, giving Stuxnet a chance to traverse through a secure digital tunnel. My colleagues and I saw exactly that when we helped Stuxnet-infected clients in industries completely unrelated to the nuclear field back in 2010.

Given that Stuxnet reported Internet protocol addresses and hostnames of infected systems back to its command-and-control servers, it appears that the attackers were clearly anticipating (and accepting) a spread to noncombatant systems and were quite eager to monitor that spread closely. This monitoring would eventually deliver information on contractors working at Natanz, their other clients, and maybe even clandestine nuclear facilities in Iran.

Stuxnet also provided a useful blueprint to future attackers by highlighting the royal road to infiltration of hard targets. Rather than trying to infiltrate directly by crawling through 15 firewalls, three data diodes, and an intrusion detection system, the attackers acted indirectly by infecting soft targets with legitimate access to ground zero: contractors. However seriously these contractors took their cybersecurity, it certainly was not on par with the protections at the Natanz fuel-enrichment facility. Getting the malware on the contractors' mobile devices and USB sticks proved good enough, as sooner or later they

physically carried those on-site and connected them to Natanz's most critical systems, unchallenged by any guards.

Any follow-up attacker will explore this infiltration method when thinking about hitting hard targets. The sober reality is that at a global scale, pretty much every single industrial or military facility that uses industrial control systems at some scale is dependent on its network of contractors, many of which are very good at narrowly defined engineering tasks, but lousy at cybersecurity. While experts in industrial control system security had discussed the insider threat for many years, insiders who unwittingly helped deploy a cyberweapon had been completely off the radar. Until Stuxnet.

And while Stuxnet was clearly the work of a nation-state -- requiring vast resources and considerable intelligence -- future attacks on industrial control and other so-called "cyber-physical" systems may not be. Stuxnet was particularly costly because of the attackers' self-imposed constraints. Damage was to be disguised as reliability problems. I estimate that well over 50 percent of Stuxnet's development cost went into efforts to hide the attack, with the bulk of that cost dedicated to the overpressure attack which represents the ultimate in disguise - at the cost of having to build a fully-functional mockup IR-1 centrifuge cascade operating with real uranium hexafluoride. Stuxnet-inspired attackers will not necessarily place the same emphasis on disguise; they may *want* victims to know that they are under cyberattack and perhaps even want to publicly claim credit for it.

And unlike the Stuxnet attackers, these adversaries are also much more likely to go after civilian critical infrastructure. Not only are these systems more accessible, but they're standardized. Each system for running a power plant or a chemical factory is largely configured like the next. In fact, all modern plants operate with standard industrial control system architectures and products from just a handful of vendors per industry, using similar or even identical configurations. In other words, if you get control of one industrial control system, you can infiltrate dozens or even hundreds of the same breed more.

Looking at the two major versions of Stuxnet in context leaves a final clue -- a suggestion that during the operation, something big was going on behind the scenes. Operation Olympic Games -- the multiyear online espionage and sabotage campaign against the Iranian nuclear program -- obviously involved much more than developing and deploying a piece of malware, however sophisticated that malware was. It was a campaign rather than an attack, and it appears that the priorities of that campaign shifted significantly during its execution.

When my colleagues and I first analyzed both attacks in 2010, we first assumed that they were executed simultaneously, maybe with the idea to disable the cascade protection system during the rotor-speed attack. That turned out to be wrong; no coordination between the two attacks can be found in the code. Then we assumed that the attack against the centrifuge drive system was the simple and basic predecessor after which the

big one was launched, the attack against the cascade protection system. The cascade protection system attack is a display of absolute cyberpower. It appeared logical to assume a development from simple to complex. Several years later, it turned out that the opposite was the case. Why would the attackers go back to basics?

The dramatic differences between both versions point to changing priorities that most likely were accompanied by a change in stakeholders. Technical analysis shows that the risk of discovery no longer was the attackers' primary concern when starting to experiment with new ways to mess up operations at Natanz. The shift of attention may have been fueled by a simple insight: Nuclear proliferators come and go, but cyberwarfare is here to stay. Operation Olympic Games started as an experiment with an unpredictable outcome. Along the road, one result became clear: *Digital weapons work*. And different from their analog counterparts, they don't put military forces in harm's way, they produce less collateral damage, they can be deployed stealthily, and they are dirt cheap. The contents of this Pandora's box have implications much beyond Iran; they have made analog warfare look low-tech, brutal, and *so 20th century*.

In other words, blowing the cover of this online sabotage campaign came with benefits. Uncovering Stuxnet was the end of the operation, but not necessarily the end of its utility. Unlike traditional Pentagon hardware, one cannot display USB drives at a military parade. The Stuxnet revelation showed the world what cyberweapons could do in the hands of a superpower. It also saved America from embarrassment. If another country -- maybe even an adversary -- had been first in demonstrating proficiency in the digital domain, it would have been nothing short of another Sputnik moment in U.S. history. So there were plenty of good reasons not to sacrifice mission success for fear of detection.

We're not sure whether Stuxnet was disclosed intentionally. As with so many human endeavors, it may simply have been an unintended side effect that turned out to be critical. One thing we do know: It changed global military strategy in the 21st century.

EPA/ABEDIN TAHERKENAREH

Session 6 – War After Western Civilization

The wars of the 21st century

HERFRIED MÜNKLER*

In a passage of his work *On War* to which commentators have not given the attention it deserves¹, the Prussian military theorist Carl von Clausewitz described war as “a true chameleon”, forever changing and adapting its appearance to the varying socio-political conditions under which it is waged. Clausewitz elucidated his metaphor by distinguishing three elements of warfare: the intrinsic violence of its components, the creativity of the strategists and the rationality of the political decision-makers. The first of these, the “intrinsic violence of its components, the hatred and enmity which should be regarded as blind instinct”, he ascribes to the populace; he sees the second, the “play of probabilities and chance which makes it [war] a free activity of the soul”, as being a matter for the generals; and lastly, the “subordinate nature of a political tool, whereby it belongs purely to reason”, as making war an instrument for the government.² In each of these domains, social developments, shifting political relationships, technological progress and finally cultural change are continuously bringing about new configurations. In consequence, war, too, is forever assuming new and different forms. From Clausewitz’s perspective, the factor that brings about the most far-reaching and momentous changes in the forms taken by war is the interdependence between elemental violence, strategic creativity and political rationality.

Asymmetry as the salient feature of the new wars

Strategic creativity and the theory of speed

In the light of Clausewitz’s definition of war, the special creativity of Mao-Tse-tung as a theorist of guerrilla warfare lies in his discovery that a slow approach, a deceleration of the course of events, provides an opportunity for successful armed resistance against an enemy who is superior in terms

* Professor, Political Theory at Humboldt University, Berlin, Germany. He has recently published a book on new wars: *Die neuen Kriege*, Rowohlt, Reinbek bei Hamburg, 2002.

both of military technology and of military organization, a discovery which was to raise small-scale war – previously conceived of simply as a concomitant strategy of large-scale war – to the level of a political-military strategy in its own right. A technologically and organizationally superior military apparatus tends to accelerate the course of war because that is the best way for it to bring its superiority to bear. Examples are Murat's cavalry swiftly pursuing and destroying the enemy defeated by Napoleon on the battlefield; Guderian's tanks exploiting small breakthroughs to prise open deep gaps in the enemy front; and Schwartzkopf's fighter bombers and cruise missiles in the Second Gulf War, paralysing Iraq's command and supply structures before the war on the ground had even begun. The consummate strategic skill of Helmuth von Moltke the Elder in waging the wars of German unification of 1866 and 1870-71 reflected not least the fact that he was better than his adversaries in deploying resources to accelerate events. Similarly, the dramatic superiority the US military apparatus has achieved over all potential enemies in the last two decades is largely due to its capacity to exploit the various opportunities for accelerating the pace at the different combat levels.

It might be argued then — and Paul Virilio, the French theorist of speed, and his adherents are of this view³ — that the development of war constantly follows the imperatives of acceleration and that, in any conflict, victory will go to whoever has the greater potential for acceleration and the ability to use it effectively. However, Clausewitz's chameleon metaphor is a reminder that the history of war does not follow such one-way development models, generally based on technological progress, but is subject to the interplay of far more complex factors. There is a price to pay for acceleration; it entails above all an ever-increasing expenditure on logistics, a correspondingly decreasing proportion of fighting forces in the total number of troops, spiralling costs to equip troops with modern weapons and, finally, a more and more vulnerable and problem-prone military apparatus.⁴

1 An exception to this rule is Andreas Herberg-Rothe, *Das Rätsel Clausewitz: Politische Theorie im Widerstreit*, Munich, 2001, esp. pp. 98 ff.

2 Carl von Clausewitz, *Vom Kriege*, 19th ed., Werner Hahlweg, Bonn, 1980, pp. 212 ff.

3 Of Virilio's extensive works, see especially his essay *La vitesse de libération*, Paris, 1995.

4 The availability of the atom bomb apart, it is above all the inflated logistical structure of modern armies on which Martin van Creveld, in his much discussed book *The Transformation of War*, Simon & Schuster Inc., New York, 1991 (published in German under the title, *Die Zukunft des Krieges*, Munich, 1998), bases his argument that the wars of the future will no longer be waged with conventional armies.

Mao's creativity lay in his refusal to join in the race for a greater acceleration of hostilities, as his peasant army would not have been able to win a war of that kind. Instead, he rejected the principle of acceleration and, turning weakness into strength, made slowness his watchword, defining guerrilla warfare as the "long war of endurance".⁵ Guerrilla strategy also consists in using every possible means to make the enemy really pay the price of acceleration, to such an extent that the war ultimately becomes unaffordable. Raymond Aron encapsulated this situation in the formula that partisans win the war if they don't lose it and those who are fighting against partisans lose the war if they don't win it.⁶ The two sides each have a different time frame. In Vietnam, the Americans learned to their cost how effective this approach can be. Asymmetrical warfare, the salient feature of the new wars in recent decades, is based to a large extent on the different velocities at which the parties wage war on each other: asymmetries of strength are based on a capacity for acceleration which outstrips that of the enemy, whereas asymmetries of weakness are based on a readiness and ability to slow down the pace of the war. This strategy generally involves a considerable increase in the casualties suffered by one's own side. Symmetrical warfare, on the other hand, as exemplified by the wars of the eighteenth, nineteenth and even the twentieth centuries, may be defined as warfare conducted by the parties concerned at the same pace. In symmetrical warfare, it was generally only minimal advantages in terms of acceleration which made the difference between victory and defeat.

High and low-tech weapons

The wars of the twenty-first century — as will be seen from the strategic significance of deceleration in the age of acceleration — will hardly be a linear extension of the trends of the twentieth century. Greater material resources and a more advanced technological development alone will not automatically tip the scales between victory and defeat. The enormous superiority of the United States in military technology is no guarantee that the USA will emerge victorious from all the wars it seems ever more ready to wage. Yet the economically highly advanced societies of the West, based on the rule of law, political participation and a "post-heroic" mentality (i.e. for which "heroic war" and

⁵ For more detail, see Herfried Münkler, *Über den Krieg: Stationen der Kriegsgeschichte im Spiegel ihrer theoretischen Reflexion*, Weilerswist, 2002, pp. 173 ff.

⁶ Raymond Aron, *Der permanente Krieg*, Frankfurt/M., 1953, p. 48.

the sacrifice of life is no longer an ideal) will have no choice but to pursue the technological development of their military apparatus if they wish to remain capable of a military response.

The western democracies are simply unable to wage Mao Tse-tung's "long war of endurance". As they are programmed for interchange rather than sacrifice — and this is what distinguishes "post-heroic" societies from those of the "heroic" age — they will do their utmost to avoid or minimize their own losses in combat, and that is possible only with superior military technology. Examples of this include the Gulf War of 1991, in which the Iraqi forces lost around 100,000 men as compared with about 140 on the side of the US-led Coalition, and, most striking of all, Kosovo, which has gone down in military history as the first war in which the victors did not lose a single man in combat. Accordingly, the arms races of the twenty-first century will no longer be symmetrical like those of the nineteenth and twentieth centuries, when Germany and England vied with each other in the construction of warships or the USA and the USSR in nuclear delivery systems. On the contrary, it will be an asymmetrical competition between high-tech and low-tech weapons. Since 11 September we are aware that mere box cutter's knives, if used to hijack airliners so as to crash them into buildings and cities, can serve to shake a superpower to its foundations. In that case, however, it was not deceleration alone which enabled the terrorist operatives to attack the USA but a combination of speed and slowness. The infrastructures of the side attacked were exploited by a clandestine group, which was able to go about preparing the attacks quietly and calmly, and then turned aircraft into rockets and jet fuel into explosive. Mohammed Atta and his accomplices attacked the USA by using its own speed — from the concentration and intensity of air transport to a media system which broadcast the catastrophe of 11 September 2001 to the whole world in real time — as a weapon against it.

Elemental violence

Strategic creativity cannot, of course, unfold independently of the other two elements of Clausewitz's trinity, namely the genuine violence of war and the political rationality of the top decision-makers. Thus the principle of a systematic deceleration of violence — as in a guerrilla war — can be applied successfully only where an overwhelming majority of the population see no other means of resolving social, economic and political problems than to wage a war that will entail heavy losses and large-scale destruction. Only then will the

people provide the guerrilla groups with logistical support, refrain from collaborating with the enemy and continue to allow more and more of their young men (and women) to be recruited for the war. Otherwise the guerrilla fighters cannot move freely within the population like fish in the water, but are out of their native element and fall easy prey to the enemy. For a long time, this precondition limited the applicability of the asymmetric strategy of guerrilla warfare. It has been known in the form described above since the early nineteenth century, for it could in principle be used only on the defensive and only if the population was prepared to make heavy sacrifices.

The really threatening aspect of the latest forms of international terrorism is that they have overcome the constraints on the use of asymmetric warfare which have proved so effective for so long — to use Clausewitz's terminology, the limited extent of hatred and enmity and the resultant limitations to the use of war as a political tool — by discovering that the enemy's civilian infrastructure can serve as the functional equivalent of one's own civilian population and its readiness for sacrifice.⁷

Moreover, current trends also suggest that in the twenty-first century large sections of the population may well see their sole chance for the future in waging wars and emerging successful. Growing environmental risks, such as water shortages, increasing desertification and rising sea levels; a greater global inequality in the distribution of consumer goods, in educational opportunities and in living conditions; the imbalance in demographic rates and the related waves of migration; the instability of the international financial markets and the dwindling ability of States to control their own currency and economy; and, finally, in some parts of the world, the rapid disintegration of States — all these are sufficient grounds for assuming that many people will see violent change rather than peaceful development as a better chance to assure their future. Thus the use of force for a better future will become the key element of their political reasoning and they will be ready not only to fight for vital resources but also to begin asymmetrical wars with superior adversaries.

Vulnerability of the developed world

Precisely because of their advanced socio-economic development, these superior adversaries are themselves highly vulnerable and, however great their military superiority, they cannot eliminate this vulnerability. The aim of the

⁷ See Herfried Münkler, *Die neuen Kriege*, Reinbek bei Hamburg, 2002, p. 175.

US in its various projects to establish a missile defence system is to make itself invulnerable. Such missile defence systems are of course no longer directed against the Soviet Union but against enemies who, though small and relatively weak, pose a serious threat through their possession of nuclear warheads and a few delivery systems. Moreover, the hopes placed in those projects were dispelled by the attacks of 11 September. In principle, war has become not only politically but also economically unattractive for the developed countries. The costs outweigh the returns. In "post-heroic" societies, the highest value is the preservation of human life and with it the multiplication and intensification of individual sensations of well-being.

Since the end of the Second World War at the latest, western societies have therefore justified every type of armament on the grounds of defence, the purpose of such a build-up is not to prepare for war but to prevent it. If the socio-political world consisted only of such societies, Kant's concept of eternal peace would long have become a reality.⁸ However, that would require all societies to be following a course of development modelled on the western secularization of politics, social individualization and, finally, the pluralization of values. Yet it is precisely against this model that the various fundamentalist movements are fighting. Far from merely defending remnants of backward traditions, they are on the contrary resisting modernization along western lines. The dilemma that has already determined socio-political development in the 1980s and 1990s will also be crucial for the twenty-first century. The fact that a world in which society has developed through interchange and cooperation is based on assumptions which can be taken for granted only if there is an extensive levelling out of the particularities stemming from religion, culture and civilization. Thus apart from a fight to establish new rules for the distribution of economic assets, educational opportunities and the necessities of life, the defence of cultural identity could also become a recurring reason for going to war. Above all, a development theory that looks forward optimistically to peace regularly tends to overlook the fact that, not least through the socio-economic development of recent decades, new opportunities have arisen for profitable economies of developing countries based on war and violence.

⁸ Klaus-Jürgen Gantzel, for example, has defended the view that, once capitalist society has been adopted on a world scale, war will disappear as a means of acquiring goods and services. Klaus-Jürgen Gantzel, "Kriegsursachen: Tendenzen und Perspektiven", *Ethik und Sozialwissenschaften*, Vol. 3, No. 8, 1997, pp. 257-266.

The privatization and commercialization of war

Historical perspectives of the profits of war

But how has war once again become a distinctly lucrative affair? It must be recalled that waging war has not always been a loss-making business. On the contrary, at various times in European history, when the circumstances were right, the raising of private armies could be perfectly profitable. Otherwise it would be impossible to explain the emergence of mercenary forces such as the Italian *condottieri*, the Swiss *Reisläufer* or the German *Landsknechte*. All of them presumably must have seen war as a means of earning a living. As the axiom has it, *bellum se ipse alet* — war feeds on war. Italy in the fourteenth and fifteenth centuries was a particularly fertile ground for this development. The substantial financial resources accumulated in the Italian trading cities made them a lucrative target for armed aggression. At the same time, the urban upper classes were disinclined to wage war themselves. As there was plenty of surplus labour in rural areas to do the military work, nothing was easier than to establish a fixed-term employment contract, a so-called *condotta*. The urban upper classes got the rural lower classes to do their fighting for them. The latter were not slow to realize what potential power and opportunities for enrichment had come their way. Fighting wars paid well. Within a few years, many who had started out with little or nothing were living in comfort and a whole string of petty noblemen turned *condottieri* had risen to the rank of dukes and princes.

One of the characteristic features of the commercialized wars fought by the warlords of the late Middle Ages and early modern history was that those waging them sought to avoid major battles and indeed, wherever possible, even decisive ones. Fighting such battles would have undermined their interest in long-term employment and, more importantly, it would have put them at risk of life and limb, an action hardly in keeping with the attitudes of those who live from war but do not really want to die by it. The *condottieri* armies operated by trying to cut each other's lines of supply and so force the adversary to capitulate without a fight. That was a considerably more attractive proposition than mutual slaughter, and the ransoms which could be earned by capturing enemy officers and soldiers represented a highly desirable bonus. If the ransom was paid, the enemy could be released and the war could start all over again.

As a rule, those who suffered from this type of warfare were the cities and nobles who employed the mercenaries. They seldom saw their objectives achieved and were constantly having to raise funds in order to finance their wars.

For this reason they burdened the populace of their lands with special levies and war taxes. This might be described as the *civilized form* of waging war against the population, since so long as it worked, i.e. if the warlords and their soldiers were regularly paid, the population was spared in the areas in which they operated. Things soon changed if the pay was not forthcoming. Then the warlords switched to the *uncivilized form* of warfare against the population, i.e. they pillaged and plundered, setting fire to farmsteads and villages, killing the men and raping the women, in order to get all concerned to see that it was better to pay up regularly than to be subjected to this extreme form of debt collection.

The continuous rise in the cost of the military apparatus during the sixteenth and seventeenth centuries made war too expensive for the private sector, and those warlords of early modern history gradually disappeared from the scene. Albrecht von Wallenstein, the last great warlord, enjoyed considerable success at first, but was doomed to failure for political reasons.

Inter-State wars and the decrease of privatisation

There were three main reasons for the steadily rising cost of waging war: the development of artillery, the use of which was decisive in battle; the transformation of foot-soldiers into a disciplined and tactically trained infantry, who took up position in long lines to engage the enemy and were increasingly equipped with firearms; and, finally, the growth in size of the army, which had to be able to combine the deployment of infantry, cavalry and artillery in order to achieve victory in the field. Any party failing to join in the technological and organizational developments of the “military revolution” of early modern history⁹ soon fell behind and disappeared from the ranks of those waging war in accordance with symmetrical principles. However, as infantry, artillery and the enlarged army all cost money, it was not long before only the State could raise the necessary funds. A full array of artillery with guns of various calibres was now beyond the resources of private military patrons. The size of the army, the need for exercises to harmonize the deployment of the three different arms and, particularly, the requirement for constant training of the infantry over long periods, all made the provision of troops ever more costly and the waging of war a less and less attractive business proposition for the private sector. War and preparations for war were disconnected from the logic of capital amortization and transferred to the direct authority of the State.

⁹ See Geoffrey Parker, *The Military Revolution: Military Innovation and the Rise of the West, 1500-1800*, Cambridge University Press, Cambridge, 1988.

The first consequence of bringing war under State control was that, as a rule, the hostilities became shorter, both sides being interested in a quick and decisive outcome. The means of arriving at such a result was the battle and thus there arose a type of warfare designed with this end in mind, i.e. to bring about battles in order to end the war and conclude a peace. While this led to a dramatic intensification of violence on the battlefields of Europe, it simultaneously set clear limits to the use of violence in terms of time and space. War of this kind was a war of soldiers against soldiers and the civilian population was largely spared from violence and destruction, unless they had the misfortune to live in the path of an advancing army or to find themselves on the battlefield. The sharp distinction drawn between combatant and non-combatant in modern international law is based largely on this development or, at any rate, would otherwise hardly have come to be recognized and applied.

It was therefore due not least to the development of arms technology and military organization that war and peace were each given a distinct legal status, the transition from one to the other itself being marked by a legal act, i.e. a declaration of war and the conclusion of a peace treaty; that war between States and civil war came to be regarded as separate and clearly distinguishable forms of war, the former being hedged about by conventions whereas the latter was not; and that, finally, in wars between States, a distinction was made between combatants and non-combatants under the relevant provisions of the Hague Convention respecting Laws and Customs of War on Land of 1899/1907 and the Geneva Convention of 1864, and belligerents were required to do everything in their power to spare non-combatants from the effects of hostilities.

The return of privatisation in the new wars

In the new wars, the opposite applies in nearly every respect. Most of these wars are not fought by well-equipped armies but by the hastily recruited militias of tribal chiefs or heads of clans, plus the armed followers of warlords and the like. Above all, the weapons used in the new wars are cheap — small arms, automatic rifles, anti-personnel mines and machine guns mounted on pick-up trucks. Heavy weapons are only rarely used and, when they are, consist mostly of remnants from the stockpiles of the Cold War. That wars of this type can be fought — and even fought successfully — is mainly due to the fact that they are not decided on the battlefield between two armies but drag on interminably in violence directed against the civilian population. Whereas in symmetrical conflict conditions the mere preparation for a war — to say nothing of waging one — has

become ever more expensive, the strategists of the new wars have succeeded in making direct warfare so cheap that it is once again a promising business.

Obviously this does not mean that the full social cost of a war is also low. On the contrary, the long-term consequences of an internal war are immense — the destruction of the infrastructure, the devastation of the countryside, the roads and fields infested with mines, the growing up of a generation of children who have no experience of anything but war and violence.¹⁰ However, these costs do not have to be borne by the protagonists of the war. To adapt an old phrase, it could be said that the warlords and militia leaders have managed in an absolutely outrageous way to privatize the profits of the wars they wage and to nationalize the costs. That this is possible has much to do with the failure of nation-building in large swathes of the Third World. In the so-called *failed States*, there are no functioning institutions capable of putting a stop to the nationalization of costs or at least keeping them within bounds. The country's civilian population and natural resources fall prey to those who, with the help of their armed henchmen, exercise control over them. Thus the violence propagated by the warlords gouges ever deeper into society until, in the end, the only possibility of rescue is through the intervention of outside powers. Yet it remains an open question whether these powers can bring peace to the country or whether they themselves will be drawn into the hostilities and the conflict, as a result of their intervention and a possible counter-intervention, will become transnational. Events in Angola, Congo, Somalia, Afghanistan and the Caucasus region are emphatic reminders of this danger.

The growing number of new wars that have been observed over the last two decades or so are mainly characterized by the fact that in them the distinction between gainful activity and the open use of force, a distinction which developed from the nationalization of war and is prerequisite for every stable economy based on peace, has been eroded to vanishing point. In the new wars, for those who have the weapons and are ready to use them, force has become a source of income, whether to procure the means of subsistence or, frequently also, to get rich. Thus in the new wars the old axiom is making a comeback: war feeds on war, and so must be fed by war. Accordingly, these new wars are typified by the emergence of warlords who control a territory by force of arms in order to

¹⁰ See Anne Jung (ed.), *Ungeheuer ist nur das Normale: Zur Ökonomie der „neuen“ Kriege*, medico international, Frankfurt/M., 2002; Mats Berdal and David M. Malone (eds), *Greed and Grievance: Economic Agendas in Civil Wars*, Lynne Rienner Publishers, Boulder/London, 2000; François Jean and Jean-Christophe Rufin (eds), *Economie des guerres civiles*, Paris, 1996.

exploit its natural resources — from oil and minerals to precious metals and diamonds — or to issue licences for their exploitation. At the same time, there is not only a proliferation of mercenaries, the well-paid hired labour for these wars, but also an increased use of child soldiers, who have proved to be a cheap and effective means of warfare. The indifference of these children to danger, their brutality and cruelty to their adversaries, the problems that having to fight them poses particularly for international peace-keeping forces, the fact that a supply of drugs and food is enough to ensure their obedience, all of these have made the child soldier one of the warlords' favourite tools. And, from a different perspective, the poverty and distress prevailing in large parts of the Third World have assumed such proportions that to obtain regular meals or, where that is not possible, perhaps to plunder what they need, many children are ready to enter the service of a warlord. According to UN estimates, there are some 300,000 child soldiers worldwide, defined as children aged between eight and fourteen who have permanently joined the ranks of a warring party and bear arms and use force on its behalf.

It is not only the disintegration of the State in many parts of the so-called Third World which has made war on a private basis and for one's own account an attractive proposition again, but also and especially the ease with which civil war economies are able to tap into the flows of capital and goods in the world market. Apart from oil and strategic raw materials such as ores and minerals, gold and diamonds, the warlords use above all illegal or fraudulently certified goods to finance their wars and frequently to accumulate enormous fortunes. Trafficking in drugs and increasingly in young women has also proved extremely lucrative because of the high level of demand in the affluent countries. The economic entities of the OECD countries are not entirely blameless for the renewed profitability of war.

Two factors play a crucial part in the emergence of the new wars: the ability to finance them from the flows of goods and capital generated by globalization and, more important still, the fact that they have become cheap to wage. The war that East and West spent over forty years preparing for in order to prevent it taking place was an enormously costly confrontation. To some extent, the very costs of that unremitting arms race can be said to have caused the collapse of one of the sides, the USSR. While peace and conflict research institutions were still busy reconstructing and measuring the symmetries of the East-West arms race, the planners and strategists of the new wars had already succeeded in breaking away not only from the spiralling arms race but also from the compulsion to prepare for and wage symmetrical wars. This process,

which has hitherto received too little attention, is paving the way for the privatization and commercialization of war described above that could, in the long run, prove even more momentous and fateful than the East-West conflict.

These new wars are not likely to remain forever confined to the regions now affected by them, i.e. parts of Central and South America, sub-Saharan Africa, and central and southern Asia, but will spread via various channels to the affluent regions of the northern hemisphere. These are areas which the South cannot attack with traditional military means. And this is where our brief introductory remarks in connection with Clausewitz come in. War is a chameleon which adapts to the current socio-political configuration, its only constant feature being elemental violence. September 11 has given some idea of what new forms of war may take and to what extent there may eventually be a gradual demilitarization of war.

The demilitarization of war

The demilitarization of war means that the wars of the twenty-first century will be fought only partly by soldiers and, for the most part, will no longer be directed against military objectives. A return to the forms of war which the nationalization of warfare brought to an end during the sixteenth and seventeenth centuries and replaced by a disciplined military organization can already be observed. Civilian targets are now taking the place of military objectives in many areas, starting with towns and villages overrun and despoiled by militia leaders and warlords and extending to the symbols of political and economic might that were targeted by terrorist commandos on 11 September. Even the means used to carry out these attacks are less and less of a genuinely military nature. For instance, in the wars of Africa and central Asia a civilian vehicle, the Toyota pick-up, has come to symbolize the emergence of militia groups and warlords. Likewise, the terrorist attacks of 11 September were made possible only by transforming civilian facilities into assault weapons.

The attacks of 11 September and especially the series of terrorist bombings in Israel have highlighted a specific new threat: operatives who turn their own bodies into weapons and thus link the successful use of force to their own certain death. Attacks of this kind are possible only by renouncing altogether any means of escape. In other words, suicide-bombers compensate for their military inferiority by giving up any chance of survival.¹¹ For a whole series of good

¹¹ An exhaustive account of old and new forms of so-called suicide attacks is to be found in Christoph Reuter, *Mein Leben ist eine Waffe: Selbstmordattentäter – Psychogramm eines Phänomens*, Munich, 2001.

reasons this may be considered morally reprehensible, but it can hardly be disputed that a new form of “heroism” has developed which, for the “post-heroic” societies of the West, is highly dangerous in terms not only of the instruments used but also of the underlying symbolism. Apart from giving bloody proof of the vulnerability of the societies attacked, these new forms of terrorist action convey a further message for them, namely that, because they are oriented towards the preservation of life, they will ultimately be defeated by those who are ready to sacrifice themselves. The act of suicide is an expression of contempt for societies which, out of principles of social self-organization, have repudiated such sacrifice of life or make use of it only metaphorically.¹² The strategists of terror have recognized that “post-heroic” societies, with their lifestyle and self-assurance, are particularly vulnerable to attack by individuals with values of martyrdom. This is a further example of the strategic creativity which for Clausewitz is the essential characteristic of the chameleon of war.

From asymmetrical strategies...

From the strategic use of deceleration against a military apparatus which relies on stepping up hostilities to the rediscovery of suicide as a threat to interchange-based societies, the latest changes in the conduct of war are nearly always characterized by asymmetric strategies. It is therefore predictable that the wars of the twenty-first century will be predominantly asymmetric, contrary to the so-called classic wars of European history since the seventeenth century, which were almost entirely symmetrical in character. For the reciprocal use of force to be symmetrical, numerous conditions must be met, foremost among them the acknowledgement by those concerned that they are *on a par* with each other. However, this acknowledgement, which may come about by the adversaries’ mutual inclusion in a system of values thus considered binding on them both (chivalry) or by their common subjection to legal rules (international law, laws of war), depends on assumptions of equality which need to be largely satisfied: broadly similar weaponry, no strategic disparities in information, and a socially analogous form of recruitment and training of combatants. On this basis, a limitation of the use of force is possible, e.g. force to be used only between such equals who are able to identify each other as combatants.

¹² See also Herfried Münkler and Karsten Fischer, “‘Nothing to kill or die for...’ – Überlegungen zu einer politischen Theorie des Opfers”, in: *Leviathan* 28, 2000, Vol. 3, pp. 343-362, and Herfried Münkler, “Terrorismus als Kommunikationsstrategie: Die Botschaft des 11 September”, *Internationale Politik* 56, 2001, Vol. 12, pp. 11-18.

Those who fall outside this equation will be spared from the deliberate use of force, though only on condition that they, for their part, refrain from the use of force. In this way force can be confined to specific places and areas: the duelling ground, the battlefield, the front. Hence, symmetrical wars are generally characterized by a limited use of force. In asymmetrical wars, on the other hand, there is a tendency for the violence to spread and permeate all domains of social life.¹³ This is because in asymmetrical warfare the weaker side uses the community as a cover and a logistical base to conduct attacks against a superior military apparatus. The starting point of this process is marked by guerrilla warfare and its end, at present, by international terrorism.

...to transnational wars

The main feature of the symmetrical war in modern European history was that it was an international war. Once war became a monopoly of the State and was consequently fought only between States, the equality and mutual recognition needed for symmetrical warfare were institutionally guaranteed. It was only in the course of the Second World War, with the war of annihilation in the East and the strategic bombing of urban residential areas, that the limitations established on that basis to the use of force were finally breached. Until then, the State had drawn the boundaries, distinguishing between internal and external affairs, friend and foe, war and peace, military and police, loyalty and treachery, and so on. For many years, the relevant literature has recognized and used the term *internal* or *civil war* as the antonym of international war or war between States. Even so, the antonym depended on the reference system of statehood, in that it derived its meaning from the boundaries drawn by the State. The term *civil war* is the symmetrical opposite of the term *international war*; the asymmetrical antonym is *transnational war*, i.e. one in which the boundaries drawn by the States no longer play a role. This type of war crosses national borders without being waged as a war between States, such as the wars in and around Angola, Zaire/Congo, Somalia and Afghanistan. It is characterized by a constant switching of friends and foes and by a breakdown of the institutional authorities (such as the military and the police) responsible for

¹³ This distinction is discussed in detail by Mary Kaldor in *Neue und alte Kriege: Organisierte Gewalt im Zeitalter der Globalisierung*, Frankfurt/M., 2000.

ordering and having recourse to the use of force. In this context, acts of war and criminality become indistinguishable and the war drags on with no prospect of a peace accord to end it. Such wars, which had already multiplied in the 1980s and 1990s, look set — along with guerrilla-terrorist wars — to determine the course of violence in the twenty-first century in many parts of the world.

Conclusion

Is there any way to halt or at least to slow down the developments outlined above? Probably a return to the stability of statehood at the world level will be the only effective means of curbing the privatization of war, the growing asymmetry of the strategies of force and the demilitarization of war, i.e. the assertion of autonomy by elements previously incorporated in politico-military strategies. Statehood, after all, is subject to the criteria of political rationality, which are irreconcilable with such developments.¹⁴ However, in view of the trends subsumed under the term *globalization*, any such renewed nationalization of politics at the world level seems doubtful. Above all, it would have the desired success only if elites capable of resisting corruption were to come to power in these States. In view of the developments currently to be observed, this, too, appears a rather unlikely prospect. Thus the wars of the twenty-first century will, in the majority of cases, not be waged with massive firepower and tremendous military capabilities. They will tend to go on smouldering with no clear beginning or end, while the dividing line between the warring parties on the one hand and international organized crime on the other will become more and more blurred. For this reason, some people are already disputing the fact that such situations do indeed constitute war.¹⁵ They forget that before war became a State monopoly in Europe, there was even then a close alliance between mercenaries and bandits. It looks as though, during the twenty-first century, the chameleon of war will increasingly change its appearance to resemble in many respects the wars waged from the fourteenth to the seventeenth centuries.

¹⁴ This question is discussed in greater detail in Herfried Münkler, "Die Kriege der Zukunft und die Zukunft der Staaten" in Wolfgang Knöbl/Gunnar Schmidt (eds), *Die Gegenwart des Krieges: Staatliche Gewalt in der Moderne*, Frankfurt/M., 2000, pp. 52-71.

¹⁵ For example, Erhard Eppler, *Vom Gewaltmonopol zum Gewaltmarkt: Die Privatisierung und Kommerzialisierung der Gewalt*, Frankfurt/M., 2002.

Résumé

Les guerres du XXI^e siècle

Herfried Münkler

Cet article identifie et fait ressortir les traits saillants des « nouvelles guerres » du XXI^e siècle et analyse trois phénomènes qui leurs sont propres : l'asymétrie, la « démilitarisation » ainsi que la privatisation et la commercialisation de la guerre.

L'asymétrie entre les parties aux conflits est le premier élément servant à distinguer les guerres actuelles de celles du siècle dernier. L'auteur explique ce point de vue en liant la théorie de la vitesse à la guerre : d'un côté les belligérants bénéficiant d'une technologie plus performante l'utilisent comme moyen pour accélérer la guerre et obtenir ainsi une victoire rapide ; de l'autre côté les guérilleros, en ralentissant la guerre, leur font payer cette accélération au prix fort. C'est entre autres pour cette raison que les sociétés les plus avancées sur le plan technologique n'ont pas nécessairement l'ascendant sur leur ennemi dans les conflits asymétriques.

Deuxièmement, les « nouvelles guerres » vont se « démilitariser » parce qu'elles ne seront plus combattues uniquement par des soldats et qu'elles ne seront plus dirigées principalement contre des cibles militaires. Ces changements reflètent à nouveau les effets d'une stratégie asymétrique. Cette tendance est accentuée et liée à la confusion sur les règles humanitaires applicables dans les conflits déstructurés ou transnationaux.

Le troisième élément est l'augmentation de la privatisation et de la commercialisation des conflits. Les événements du 11 septembre 2001 ont montré que parmi les parties aux conflits peuvent figurer des groupes criminels ou terroristes transnationaux. Cela pourrait aboutir à des conflits privés entre certains États et ces types d'acteurs internationaux. Le phénomène des seigneurs de la guerre tirant profit du conflit et ayant de ce fait un intérêt dans la continuation de celui-ci a déjà ressurgi lors des « nouvelles guerres ». L'article retrace aussi l'aspect historique de cette commercialisation de la guerre.

L'auteur conclut que ces tendances vont probablement continuer à affecter la majorité des guerres dans un avenir proche, à moins qu'un retour à la stabilité des États puisse freiner la privatisation et la démilitarisation de la guerre. La globalisation pourrait également équilibrer la distribution du pouvoir et de la richesse, en diminuant les causes sous-jacentes des conflits asymétriques.

The Evolution of Law of War

Sheng Hongsheng*

Idea of Law of War

It is generally concluded that since time immemorial war may have existed, but the nature of war is nonetheless an issue that continues to perplex humankind. In ancient times, each state had its own theories and perceptions of how the nature of war should be defined. For example, in ancient Greece, 'waging war against barbarians is considered just'.¹ It was Saint Augustine (354–430 A.D.), a representative Christian figure in ancient Rome, and Saint Thomas Aquinas (1225–1274 A.D.), renowned Medieval theologian, who first put forward just war doctrine. To their mind there were so-called 'three principles' that govern the prerequisites for the initiation of war, namely, the authoritative nature of the initiator of war, just reasons for waging war, and the legitimate intentions of war.² Hugo Grotius later expanded upon the concept of just war, and it eventually became one of the sources of modern law of war. Inspired by natural law and elaborated upon by theologians, just war doctrine consistently held prestige within French legal writings, and when the 16th century ended, it was codified by Spanish jurists such as Francisco de Vitoria and Francisco Suárez. Just war doctrine regards war as neither a fortuitous event nor a simplistic fact, but a procedure through which to resolve international conflict. A just war must accord with four conditions: (1) in the name of justice—that is, a war can only be waged by a legitimate public authority, consequently private wars in the Middle Ages were condemned; (2) a just cause—that is, a reason predicated upon what is just, but also justifies the damages caused by war; (3) necessity—there are no other means by which justice can be achieved; and (4) appropriate conduct—that is, war must be conducted in a manner appropriate for the renewed restoration of order and peace.³

¹ W. G. Vitsthum et al., eds., *Völkerrecht*, trans. Wu Yue et al. (Beijing: Falu chubanshe, 2002), p. 56.

² Gu Dexin, *Zhanzheng Xintun (New Theory of War)* (Beijing: Shijie zhishi chubanshe, 1990), p. 48; Peter Haggenmacher, 'Just War and Regular War in Sixteenth Century Spanish Doctrine', *International Review of the Red Cross*, No. 290, September/October 1992, pp. 434–45.

³ Charles Rousseau, *Le Droit des Conflits Armés*, trans. Zhang Ning et al. (Beijing: Zhongguo duiwai fanyi chubanshe, 1987), p. 12.

* Corresponding author. Email: adamsheng@163.com

When discussing the laws of armed conflict, most Chinese international lawyers focus on the question of a just war versus an unjust war, but their Western counterparts do not. They do not judge the nature of war; their discussions are limited to the necessity for and manner of application of the laws of armed conflict. The view of the International Committee of the Red Cross (ICRC), a key initiator, promoter, and the guardian of international humanitarian law, on this question is that the purpose of international humanitarian law in protecting and assisting war victims is to constrain the damages and suffering caused by warfare. International humanitarian law, therefore, seeks to resolve the practical problems of armed conflict rather than the reasons for or legality of resorting to armed force. The provisions of international humanitarian law are applicable to all wars, regardless of causes or whether or not the actions taken by either or both sides are just. The standards by which aspects of armed conflict are assessed in international humanitarian law stem exclusively from humanitarianism—the so-called ‘law in time of war’. Nowadays, at times of international armed conflict, it is often difficult to determine which states are in violation of the United Nations Charter. Application of international humanitarian law does not involve condemnation of the transgressing Power at times of war, because the ensuing controversy, where all parties concerned would claim to be the victims of aggression, would put the law into a state of paralysis. Moreover, the purpose of international humanitarian law is to protect the victims of war and their basic rights, regardless of which side they are on. This is why it is necessary to differentiate ‘*jus in bello*’ (law in war), ‘*jus ad bellum*’ (law on the use of force) and ‘*jus contra bellum*’ (law on the prevention of war).⁴ To determine the nature of war has already exceeded the scope of the laws of armed conflict, in accordance with the ICRC. It is considered that the fundamental objective of the laws of armed conflict is to provide war victims with a minimum level of protection. International humanitarian law should be applicable to all belligerent parties, regardless of the origins of the conflict. The essence of this principle of universal application is in ancient international law. For example, ‘In history and international relations of ancient India, one will not find differences between believers and non-believers. Even when both are involved in a life-and-death struggle, no one will raise the question of whether it is a “just war” or an

⁴ International Committee of the Red Cross, *International Humanitarian Law: Answers to Your Questions* (Geneva: International Committee of the Red Cross, 2000), p. 6. The Chair of the International Committee of the Red Cross holds that there is no concept of a ‘Just War’ or an ‘Unjust War’ within International Humanitarian Law. The provisions of the Laws of Armed Conflict are concerned only with the protection of war victims (such as civilians). See Jakob Kellenberg, ‘International Humanitarian Law at the Beginning of the 21st Century’, 26th Round Table in San Remo on Current Problems of International Humanitarian Law, *The Two Additional Protocols to Geneva Conventions: 25 Years Later Challenges and Protects*, <http://www.icrc.org/Web/eng/siteeng0.nsf/html/5E2C8V?OpenDocument>.

“unjust war”. Dharma clearly states that during any time and under any circumstances, all belligerent parties must adhere to publicly acknowledged law of war. The universal application of such laws is an important contribution to international law made by ancient India.⁵ Any difference in the respective treatment of belligerent parties, therefore, would negate the existing moral premise upon which international humanitarian law is based and, as such, is incompatible with the basic tenet of the laws of armed conflict.

Since its inception, the laws of armed conflict has been struggling to survive and develop within the narrow space between ‘the right of a State to resort to war’ and ‘the efforts by the laws of armed conflict to provide a minimum level of protection to war victims’—in other words, between the principles of ‘military necessity’ and ‘humanitarianism’. When commenting on the essence of war, Prussian military philosopher Carl von Clausewitz held that, ‘War is a kind of violent action, and the use of violence is without limits’.⁶ Restricted as it is by international politics, the laws of armed conflict have limited room to manoeuvre. From a theoretical perspective, international law has abolished war, but a majority of wars or armed conflicts actually occurring do not adhere to the two circumstances under which the use of force is lawful, stipulated by the United Nations Charter. Today, the international community faces a dilemma between a State’s right to resort to war and the idea of the law of war. Resorting to war remains a State’s inherent right until it is officially renounced, and the ultimate ideal of war, according to a general law concerning the use of armed force, is to eliminate an enemy while preserving oneself. The idea behind the laws of armed conflict, on the other hand, is to limit the brutality of war by providing impartial treatment and a minimum level of protection to all war victims, regardless of what side they are on. If international law could be considered as ‘weak law’, the laws of armed conflict as a branch of it are even more subtle and vulnerable. One clear instance is that in international relations, when acts that seriously contravene standard conduct under international law (e.g., aggression) occur, the situation will leave a State no remedial course of action other than individual or collective self-defense, rather than restoring its legal status prior to aggression the assistance by ‘international police’ or ‘international armed force’. The unique legal structure of the laws of armed conflict, however, differs from that of other laws by virtue of encompassing ‘assumptions’ (it states the conditions necessary for application of its norms); ‘procedures’ (basic requisites for standardized acts); but no ‘punishment’ (there are no provisions for standard legal

⁵ Li Jiashan, *Guojifaxueshi Xinlun (New Comment on History of International Law Theory)* (Beijing: Falu chubanshe, 1987), p. 43.

⁶ Carl von Clausewitz, *Vom Kriege (On War)* (abridged version in Chinese) (Beijing: Zhanshi chubanshe, 1978), p. 5.

consequences of violating the law aforementioned).⁷ The legal consequences of violating prohibitions set forth in the laws of armed conflict are instead left to international criminal law and international criminal justice. For instance, 'The Hague Conventions of 1899 and 1907 as well as the Geneva Convention of July 27, 1929 regarding the Treatment of Prisoners of War, neither of the above contain provisions that penalize individuals for acts in violation of the treaties. Only Article 30 of the Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, July 27, 1929, bears a brief provision in this regard'.⁸

As the means of regulating social conduct under extreme and unique circumstances, the entire philosophical foundation and legislative purposes upon which the laws of armed conflict are based—lie in the distinction between combatants and non-combatants (like civilians) or those no longer able to engage in warfare (such as prisoners of war, the wounded and the sick), thereby minimizing the gravity of violence to what is sufficient, to achieve the aim of war (regardless of the causes of war), and to the extent that enemy forces are weakened. As earlier elaborated upon, in general, the basic principles of the laws of armed conflict include: humanity, necessity, proportionality, distinction, prohibition of causing unnecessary suffering, and independence of *ius in bello* from *ius ad bellum*. The intrinsic flaws of the laws of armed conflict are obvious—it does not prohibit the use of violence; it is unable to protect all those affected by armed conflict; it is unable to distinguish how the law should be applied according to the aims of armed conflict; it is unable to prevent the defeat of one side by another during war—assuming that belligerent parties have rational reasons for going to war. To some extent, these demerits undermine the effectiveness and usefulness of the laws of armed conflict.

Recalling the history of human society, international battlefields, in some sense, were the cradle of international law.⁹ In other words, international law originated in warfare. The ancient Greek philosopher Plato believed that war is a normal condition between nations and that peace is a degenerate state.¹⁰ Marcus Tullius Cicero, great thinker of ancient Rome,

⁷ 'Regardless of the stipulations of the sanctions, legal norms generally have punishments, because punishments are the coercive measures that ensure the enforcement of legal norms. They are symbols of legal norms'. See *Chinese Encyclopedia (Law)* (Beijing: Zhongguo dabaikē chubanshe, 1984), p. 101.

⁸ 'On the request of a belligerent, an enquiry shall be instituted, in a manner to be decided between the interested parties, concerning any alleged violation of the Convention; when such violation has been established the belligerents shall put an end to and repress it as promptly as possible'. (Art. 30, the 1929 Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armies in the Field.) See Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility under International Law', *International Law Review of the Red Cross*, Sep. 30, 1999, No. 835, pp. 531–53.

⁹ Li Jiashan, *New Comment on History of International Law Theory*, p. 13.

¹⁰ *Ibid.*, p. 15.

denied the existence of the law of war¹¹, while other legal theorists were of the opinion that, 'where there is society there is law'.¹² It is generally believed that 'even wars have limits',¹³ otherwise, in the realm of societal relations as unique as that of war, a state of legal vacuity and lawlessness is bound to emerge. Over the course of history, this school of thought has gradually become more widely accepted. War is a unique societal phenomenon wherein the belligerent parties involved make every effort to utilize violence without constraint. This type of 'endless expansion of evil' does not, in actuality, help to achieve the aim of war. Consequently, wars have given rise to various rules that limit the extent of violence in warfare and which have, in turn, created the logic upon which law of war emerged.

Vestiges of Law of War in Ancient Times and Emergence of Modern Law of War

International law originated in war, and vestiges of conduct norms relating to war can be found in almost every ancient civilization. In 546 A.D., during the Chunqiu Period of ancient China, feudal lords of various States organized the first conference ever on *mibing* (elimination of war) and *qvbmg* (disarmament)¹⁴ in the State of Song. During the Warring States period, rules existed amongst all the seven feudal States (also known as the 'Seven Powers') regarding declaration of war, conclusion of a peace, favorable treatment for prisoners of war¹⁵, but also specific rules applicable in war, such as 'non-pursuing a fleeing enemy army', 'no use of ruses', 'non-attacking on a mourning State', 'prohibition on surprise attack', 'no severely injuring' and 'no capture of the elderly'.¹⁶ Although these rules applied to each of the warring States, not all of them were strictly adhered. Sima Qian wrote about the Battle in Changping (250 B.C.) where the King of Qin defeated the State of Zhao. The Zhao nationals surrendered to the

¹¹ Cicero pleaded, '*...silent enim leges inter arma* (Laws are silent among those who use weapons)', cited in Cicero, *Pro Milone*, 4. 11. See Marco Sassoli and Antoine A. Bouvier, eds., *How Does Law Protect in War? Cases, Documents and Teaching Materials on Contemporary Practice in International Humanitarian Law* (Geneva: International Committee of the Red Cross, 1999), p. 68.

¹² Liang Xi, ed., *Guojifa (International Law)* (Wuhan: Wuhan daxue chubanshe, 2000), p. 6; A. Verdross et al., *Völkerrecht (International Law)*, trans. Li Haopei (Beijing: Shangwu yinshuguan, 1981), pp. 9–10; Li Jiashan, *New Comment on History of International Law Theory*, p. 9; Zhao Lihai, *Guojifa jiben lilun (Basic Theory of International Law)* (Beijing: Beijing daxue chubanshe, 1990), p. 116.

¹³ S. Fleming, ed., *People On War* (Geneva: International Committee of the Red Cross, 1999), Front Cover.

¹⁴ *Chunqiu Zuozhuan jizhu (Collected Interpretation of History of the Chunqiu Period, Volume 3)* (Shanghai: Shanghai Renmin Chubanshe, 1977), p. 1077.

¹⁵ Liang Xi, ed., *International Law*, p. 35.

¹⁶ Wang Tieya, ed., *Guojifa (International Law)* (Beijing: Falu chubanshe, 1981), p. 509; Sun Yurong, *Gudai Zhongguo guojifa yanjiu (Study on International Law in Ancient China)* (Beijing: China University of Political Science and Law Press, 1999), pp. 182–90.

Qin army, yet the King of Qin immediately ordered the live burial of as many as 400,000 prisoners of war, violating the promise to spare their lives.¹⁷ Even so, it must be remembered that these rules and regulations were very primitive and hardly set in stone. Moreover, the feudal States at that time were not the same as States in the general sense because they did not constitute a true international community. When taking into consideration the particular situation of ancient China, it is open to question whether or not the rules observed can be construed as 'international law',¹⁸ but they can most certainly be seen as historical antecedents to the ancient law of war. In other areas of the world, in common with ancient China, there were a variety of rules regarding warfare. For example, in Egypt, Babylon, India, ancient Greece and Rome, there are documented records of rules of engagement in warfare, such as prohibition on using concealed, barbed or poisoned weapons, no attacking on fleeing, surrendered enemy forces, or those who put down their weapons, and prohibition on poisoning drinking water.¹⁹

In ancient times, war was generally regarded as a natural state. During the Middle Ages, it was the main feature of dispute over rights, and the use of armed forces during conflict was considered a kind of evidence which could be submitted to a court. If it were to be said that only the legality of war was considered in ancient times, on the same premise, (as related to the procedure through which war was declared) by the Middle Ages more and more attention was paid to demonstrating the justness of war. Feudal, separatist regimes caused extensive wars that, during the 11th and 12th centuries, engendered great, ongoing efforts to limit the expansion of warfare or even more to abolish war. A rule emerged step by step that war could be declared and conducted only by the head of State (the sole person entitled to do so), but that there must also exist legitimate and justified reasons for war. The reasons stipulated in the law were presumably for reasons of reclamation of rightful property and safeguarding the homeland. It was demanded that declaration of war was necessary and acknowledged. All men were eligible as combatants (fighters). During the initial wars of this era, there were no limitations on the means of warfare. It was with the onset of the use of guns and cannon that attempts to impose restrictions on them were made. In 1139, for example, at the Second Lateran Council, an attempt to prohibit the use of long-range catapults and crossbows was made, but this effort soon failed. The people of the Middle Ages also constantly sought to impose restrictions upon the time when war could be conducted, which resulted in the so-called 'peace of God' or 'holiday of God'. The institution of neutrality also slowly developed throughout the Middle Ages.

¹⁷ See Sima Qian, *Shiji: Qin benji (Historical Chronicles: Qin Almanac)*. Nonetheless, even today it is sceptical about the allegation that 400,000 prisoners of war were buried alive by the Qin State after Battle in Changping. Its veracity remains to be proven.

¹⁸ Liang Xi, ed., *International Law*, p. 35.

¹⁹ Wang Tieya, ed., *International Law*, pp. 509–10.

Between the 14th and 15th centuries, treaties containing provisions on neutrality were signed, the earliest being the one between the Duke of Lorraine and the Duke of Bavaria in 1322. Initially, neutrality meant one party's refusal to aid either of the belligerents; it was not until later that it included the stipulation stating an obligation that the neutral subject and the belligerents should not attack or cause breach of neutrality. Such neutrality means that a Neutral does not provide military assistance to either of the warring parties, but neither prohibits the army of either side from passage or stationing in neutral territory, nor prohibit from providing either side with diplomatic and other forms of non-military aid. As regards to law of war at sea institution related to prize and privateering also achieved a certain level of development. The focus of naval war was on destroying enemy trade and transport connections, and widespread plundering at sea gave impetus to the development of 'prize law'. Lawful prize is defined along very broad lines. In addition to the prize of enemy ships and enemy goods, it includes prized ships and goods of neutral States. Prize law also prohibits the trading of goods with the enemy or transporting goods on behalf of (or to) the enemy. These stringent regulations were seldom met, but nevertheless exerted limitation upon maritime libertine conduct. In the 14th century, the 'prize law' was incorporated into 'Maritime Code'. This code also includes a provision on returning what is captured; it states that prized ships must be returned to their rightful owners, and also that it is only at the point where prized ships have been taken to the port of the captor's native State or to the port of another 'dependable location' that the reality of the prize is acknowledged. The institution of privateering also gradually entered the ambit of internal law. In the 13th to 14th centuries, decrees were issued outlining a method of granting permits to privateers that would render them legitimate seafarers. The order, issued by Alfonso III (King of Aragon) in 1228, is considered to be the first-ever decree to address privateering. A century or so later when similar laws were in force in other States came the establishment of prize courts (for example, France's edict of 1373, and the English Parliamentary decree of 1414).²⁰ Warfare in the Middle Ages was extraordinarily brutal, yet rules governing warfare were still at a developmental stage. The works of many scholars give systematic descriptions of the rules of war, notably *De Jure Belli ac Pacis* (On the Law of War and Peace) by international law pioneer Hugo Grotius, in which he comprehensively summarizes the traditional concepts and rules of warfare, as well as those of the Middle Ages.

There was a popular preoccupation with rules relating to war itself in the era preceding the entry of gunpowder based arms into warfare. As human society entered the era of firearms, advanced science and technology

²⁰ Firdman and Bastin, *History of International Law* (Moscow: International Relations Press, 1990), trans. Huang Daoxiu et al. (Beijing: Falv Chubanshe, 1992), p. 62.

accelerated the advent of new means of war (and weaponry), their increased killing potential considerably intensified the brutality of war. In order to limit the means and methods of warfare, the conduct of all warring entities was standardized, thereby decreasing the destructive gravity of war. After much hard work on the part of the international community, a series of international instruments were formulated at the end of the 19th century, one of them was the Declaration of Paris in 1856, and the compilation of the laws of armed conflict entered a new phase. After the Declaration of St. Petersburg in 1869, The Hague Peace Conferences of 1899 and 1907 and their respective conventions, the Geneva Protocol of 1925, the Geneva Convention of 1929, the four Geneva Conventions of 1949 and the two Additional Protocols of 1977, the laws of armed conflict had progressively developed into the most accomplished branch within the system of international law. As commented upon by a Western scholar, 'In the last years of the 19th century, legal empiricism and the perception of national sovereignty had occupied the dominant position within the idea of international law. This led to the widespread codification of law of war, which is the earliest area of international law where codification was achieved'.²¹

French Enlightenment philosopher Jean Jacques Rousseau is one of the earliest Western scholars to analyse the nature of war from a theoretical standpoint. In his book, *The Social Contract* of 1762, he points out that: 'War is not between individuals, it is between States. In war, the accidental enmity among individuals that occurs is not because they are acting in their capacity of an individual, but in that of a citizen. They are not acting as elements of their State but rather as its protectors'.²² Also, as affirmed by an Italian scholar, 'Already in the Ordinance for the Government of the Army, published in 1386 by King Richard II of England, limits were established to the conduct of hostilities and—on pain of death—acts of violence against women and unarmed priests, the burning of houses and the desecration of churches were prohibited. Provisions of the same nature were included in the codes issued by Ferdinand of Hungary in 1526, by Emperor Maximilian II in 1570 (humanitarian rules are found in Articles 8 and 9) and by King Gustavus II Adolphus of Sweden in 1621. Article 100 of the Articles of War decreed by Gustavus II Adolphus established that no man should 'tyrannise over any Churchman, or aged people, Men or Women, Maydes or Children'.²³ Providing needed humanitarian aid to war victims and

²¹ Rupert Ticehurst, 'The Martens Clause and the Laws of Armed Conflict' in Li Zhaojie, ed., *Guoji rendaozhuyifa wenxuan (Selected Articles From The International Review of The Red Cross)* (Beijing: Falu chubanshe, 1999), p. 32.

²² Charles Rousseau, *Le Droit des Conflits Armés*, trans. Zhang Ning et al. (Beijing: Zhongguo duiwai fanyi chuban gongsi, 1987), p. 2.

²³ Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', *International Review of the Red Cross*, No. 835, pp. 531–53.

compiling widely accepted humanitarian norms is a phenomenon whose origins are in remote antiquity. There are specific references to the concept in various philosophical writings emanating from different cultures and in the speeches of enlightened monarchs. Contemporary notions of humanitarianism originate in 'A Memory of Solferino', written by Henri Dunant in 1862, and the founding of the International Committee of the Red Cross.²⁴ Much of the development of the laws of armed conflict occurred under the influence of humanitarian ideals, manifest in the series of international conferences that took place at the end of the 19th century, which acted as impetus towards the compilation of the laws of armed conflict. Notable provisions of the 1856 Declaration of Paris are: privateering is, and remains, abolished; the neutral flag covers enemy goods, with an exception of war contraband; neutral goods, with the exception of war contraband, are not liable to prize under the enemy flag; in order to be binding, blockades must be effective, that is, they must be maintained by forces sufficient to prevent access to the enemy coast. The two Hague Peace Conferences had a great impact upon the formation of laws of war. The First Hague Peace Conference of May 18, 1899 convened at The Hague in the Netherlands with 26 participating States, was so-called because its official purpose was 'to limit arms and to safeguard peace'. The Second Hague Peace Conference of 1906–1907 called by Russia and the US drew on the experience and lessons of the 1899 conference, and convening States concentrated on the issue of laws of war. The treaty that was adopted during this conference replaced the decision reached at the 1899 Peace Conference. The Second Hague Peace Conference passed the Declaration on Prohibiting the Launching of Projectiles and Explosives from Balloons, Final Act, as well as 13 other conventions. The areas and issues covered by these aforementioned instruments are extensive but fall under two categories: (1) adjustment of laws concerning peaceful settlement of international disputes (the first and second conventions); and (2) procedures for the opening of hostilities (the third convention), laws and customs of war on land (the fourth and fifth conventions) and rules concerning the conducting of naval war (the sixth to thirteenth conventions). The Third Hague convention stipulates that, 'unless there is advanced and explicit warning', hostilities cannot commence. The Contracting Powers recognize that warning can take the form of either a reasoned declaration of war or an ultimatum. Under this reservation clause, the humanitarian stipulations of each of the remaining conventions suffered huge limitations. In cases of laws and customs of war on land, the most important clause within the annex to the Fourth Hague Convention is the distinction between 'combatant' and 'non-combatant'. Civilians have the

²⁴ Jacques Stroun, 'International Criminal Jurisdiction, International Humanitarian Law and the Humanitarian Action' in Li Zhaojie, ed., *Selected Articles From The International Review of The Red Cross*, p. 134.

right to take up arms in resistance (but on the condition that 'arms must be openly carried and the laws and customs of war must be respected'). In addition, any person captured by the enemy enjoys the rights applicable to prisoners of war. The legal status of a prisoner of war is stipulated in great detail (Article 4): 'Prisoners of war are in the power of the hostile government and not in the power of the individual or the corps that captured them. They must be treated humanely. All their personal belongings, with the exception of arms, horses and military papers, remain their property.' Prisoners of war utilized in labor, 'must be done so according to their rank and aptitude', and 'officers are excepted'. This kind of labor should not be 'excessive' and should be paid a wage. The wage, 'should be paid according to the same standards as those of soldiers in the national army'. Officers taken prisoner shall receive, 'the same pay as officers of corresponding rank within the state where they are detained'. The soldiers and others accompanying and serving the army in an official capacity, mobile medical units, hospital and medical personnel shall all receive humane treatment and protection from all belligerent states. When captured, they shall be treated according to the standards of international law.²⁵ Provisions regarding the methods of war include: prohibition on the destruction or confiscation of enemy property, 'unless the destruction or confiscation in question is required by the necessities of war, and where there are no alternatives'; prohibition on killing those that have surrendered and declaring that 'no quarter will be given' (Article 23). The conventions request that commanding officers take all possible measures to protect cultural and historical landmarks as well as medical facilities (Article 27). The pillage of a town or place, even when taken by assault, is prohibited (Article 28). During occupation, the authority of the legitimate power having in fact passed into the hands of the occupant, the latter shall take all the measures in his power to restore, and ensure, as far as possible, public order and safety, while respecting, unless absolutely prevented, the laws in force in the State (Article 43); private property cannot be confiscated (Article 46); pillage is formally forbidden (Article 47), but 'if, in the territory occupied, the occupant collects the taxes, dues, and tolls imposed for the benefit of the State, he shall do so, as far as is possible, in accordance with the rules of assessment and incidence in force, and shall in consequence be bound to defray the expenses of the administration of the occupied territory to the same extent as the legitimate Government was so bound' (Article 48); no contribution shall be collected except under a written order, (Article 51); nothing more than the enemy's ready money, reserve funds, bonds and other property used for war can be confiscated (Article 53); The occupying State shall be regarded only as administrator and usufructuary of public buildings, real estate, forests, and agricultural estates belonging to the hostile State,

²⁵ Firdman and Bastin, *History of International Law*, p. 129.

and situated in the occupied State. It must safeguard the capital of these properties, and administer them in accordance with the rules of usufruct (Article 55). With regard to laws governing neutrality, the Fifth Hague Convention of 1907 stipulates that, the territory of neutral Powers is inviolable (Article 1). Belligerents are likewise forbidden to: (a) Erect on the territory of a neutral Power a wireless telegraphy station or other apparatus for the purpose of communicating with belligerent forces on land or sea; (b) Use any installation of this kind established by them before the war on the territory of a neutral Power for purely military purposes, and which has not been opened for the service of public messages (Article 3); Every measure of restriction or prohibition taken by a neutral Power in regard to the matters referred to in Articles 7 and 8 must be impartially applied by it to both belligerents, and 'A neutral Power must see to the same obligation being observed by companies or private individuals owning telegraph or telephone cables or wireless telegraphy apparatus' (Article 9). In addition, the use of force by a neutral Power in resisting attempts to violate its neutrality shall not be regarded as a hostile act. In terms of naval war, there are eight Hague conventions with related provisions. The main problems addressed include: the conversion of merchant ships into warships; the status of enemy merchant ships at the outbreak of hostilities; the laying of submarine mines and the bombardment by naval force; application of the Geneva Convention principles to naval war; the issue of injured and sick persons; the exercising of the right to prize; the creation of an international prize court and the rights and obligations of neutral Powers during naval war; etc. These conventions carry a series of supplemental provisions pertaining to the peculiarities of naval war, but they are essentially in line with the basic principles underlying the conventions regarding war on land. During the London Naval Conference of 1908–1909, the Hague Conventions were supplemented and written into the Declaration concerning the Laws of Naval War adopted at the Conference. This declaration settled the issues of blockades, legal norms regarding wartime embargoes and the procedure for neutral Powers in aiding belligerents. The declaration also touched upon other more concrete issues, such as the changing of flags on ships, the difference in nature between the cargo of neutral Powers and that of the enemy, convoys and indemnities, etc. These rules were very successful and remained effective through the entire 20th century. Although many States signed the Declaration of London, it was not ratified by a number sufficient to make it go into effect officially.²⁶

The key points within this entire stage of development of the laws of war indicate that, on the one hand, States refused to give up their right to resort

²⁶ International Committee of the Red Cross, *International Law Concerning the Conduct of Hostilities: Collection of Hague Conventions and Some Other International Instruments* (Geneva: International Committee of the Red Cross, 1989), p. 69.

to war, and on the other hand, new inventions and achievements in science and technology in the wake of the industrial revolution were rapidly applied to military affairs. The brutality arising from war increased in tandem with the increased killing potential of weapons. At the same time, under the influence of natural law which states that 'the validity of international law is rooted in human rationality and conscience', humanism and idealistic notions of international politics gained the upper hand, spurring the development of the laws of war of antiquity into a consummate modern system.

Law of War after World War II

The two World Wars were catastrophic events in human society. All the belligerent Powers that engaged in these wars violated the laws of armed conflict. At the end of World War II, and in order to adapt to changing circumstances, representatives of a large number of States converged in Geneva, on the recommendation of the ICRC, for a diplomatic conference wherein to deliberate on the formulation and regulation of new laws on wartime conduct. On August 12, 1949, 63 States signed the new Geneva Conventions which went into force on October 1, 1950. They are collectively called the Four Geneva Conventions of 1949 Relative to the Protection of War Victims, including the Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field of August 12, 1949 (First Geneva Convention); the Geneva Convention for the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea of August 12, 1949 (Second Geneva Convention); the Geneva Convention relative to the Treatment of Prisoners of War of August 12, 1949 (Third Geneva Convention) and the Geneva Convention relative to the Protection of Civilian Persons in Time of War of August 12, 1949 (Fourth Geneva Convention). The first convention (the Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field) was first signed in 1864. In 1906 and 1929, the convention was twice revised and supplemented. In 1949, it was revised for the third time, and now contains 64 articles and two annexes in total. Its main contents cover: affirmation that under any circumstances, the sick and wounded of any belligerent party shall receive impartial and humane treatment; prohibition on the killing, torturing and subjection to biological experimentation of the sick and wounded, or knowingly not giving medical aid or care; the principle that sick and injured persons as well as medical units, medical buildings, medical supplies and medical personnel (marked by the emblem of a red cross on a white background or a red crescent) are inviolable. The second convention (the Geneva Convention for the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea) contains 63 articles

and an annex. It is a revision of and supplement to the Tenth Hague Convention of 1907. The characteristics of naval war are incorporated into this convention; it stipulates that the wounded, the sick, medical ships, medical units and their personnel, equipment and vessels must be protected. The convention applies to naval troops only; ground troops should adhere to the rules and regulations of the First Geneva Convention of 1949. The third convention (the Geneva Convention relative to the Treatment of Prisoners of War) contains 143 articles and 5 annexes. It is a revision of and supplement to the eponymous convention of 1929. Finally, the fourth convention (the Geneva Convention relative to the Protection of Civilian Persons in Time of War) is a new convention that comprizes a total of 159 articles and 3 annexes. Its provisions stipulate that: civilians in the power of a belligerent party to the conflict shall receive protection and humane treatment; unfortified cities, towns and villages shall not be destroyed; peaceful inhabitants shall not be killed, coerced, abused or exiled; the person, family, dignity, property, religious beliefs, and social customs of civilians must be respected. The principles and rules included within the Four Geneva Conventions of 1949 are as follows: (1) each convention is applicable not only to all openly declared wars but also to all armed conflicts; in the event of either or both sides' denying the existence of a state of war, the conventions nonetheless apply. (2) Each convention between the contracting Powers is binding, and when one or more States on one side of a conflict that are not signatories to the treaty accept and use the convention, the convention is then binding among those States and the States that are party to the convention, thus eliminating the restriction imposed by the so-called general participation clause that was previously stipulated within treaties regarding law of war. (3) Each convention is not only applicable to armed conflict that is international in nature, but also to armed conflict that is not international in nature (such as civil wars) and guarantees that the victims of war receive a minimum protection. The main contents of the conventions state that: 'First, those who do not actually engage in hostilities, including the troops who put down their arms, those who are sick, wounded, or have lost their ability to fight due to internment or some other reason, must be treated humanely under all circumstances and cannot be discriminated against on the basis of race, skin color, religious beliefs, gender, upbringing, wealth or other such factors. Under no circumstances, therefore, may the following actions be taken towards the aforementioned people: (a) violence to life and person, in particular murder of all kinds, mutilation, cruel treatment and torture; (b) taking of hostages; (c) outrages upon personal dignity, in particular humiliating and degrading treatment; (d) the passing of sentences and the carrying out of executions without previous judgment pronounced by a regularly constituted court, affording all the judicial guarantees which are recognized as indispensable by civilized people.

Second, the sick and wounded shall be collected and cared for (Article 3 common to the Four Geneva Conventions of 1949).

After the signing of the Four Geneva Conventions of 1949, international armed conflicts (especially the occurrence of national liberation movements) exposed faults and failings within them. In order to supplement the Four Geneva Conventions of 1949, the ICRC adopted a resolution at the 21st International Conference of the Red Cross in 1969, at which the plan to add two Protocols²⁷ was drafted. From 1974 to 1977, the Swiss government convened four 'diplomatic conferences regarding the renewal and development of international humanitarian law applicable in armed conflict'. On June 10, 1977, representatives of all the relevant States gathered together in Geneva and signed the Protocols Additional to the Four Geneva Conventions of 1949 (its full name being 'Protocols Additional to the Four Geneva Conventions of August 12, 1949'). They are Protocol Additional to the Four Geneva Conventions of August 12, 1949 relating to the Protection of Victims of International Armed Conflicts (Protocol I) and the Protocol Additional to the Four Geneva Conventions of August 12, 1949 relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), effective as of 7 December 1978. The first Protocol contains 102 articles and 2 annexes, and is supplementary to the Four Geneva Conventions of 1949 regarding protection for civilians, war victims and combatants, etc. Its main contents include: (1) expansion of the scope of situations to which the convention is applicable to include armed conflicts over colonial domination, alien occupation and racist regimes (Article 1.4), and establishing the legality of such actions (in the spirit of the Charter of the United Nations and the 1970 Declaration on Principles of International Law). If combatants in these situations fall into enemy hands, they should have the same prisoner-of-war status as is applicable in an international armed conflict. (2) Relaxation of the standards regarding the reclassification of guerrilla fighters as legal combatants (Article 43). In order to promote the protection of the civilian population from the effects of hostilities, combatants are obliged to distinguish themselves from the civilian population while they are engaged in an attack or in a military operation preparatory to an attack. Recognizing, however, that there are situations in armed conflicts where, owing to the nature of the hostilities an armed combatant cannot so distinguish himself, he shall retain his status as a combatant, provided that, in such situations, he carries his arms openly: (a) during each military engagement, and (b) during such time as he is visible to the adversary while he is engaged in a military deployment preceding the launching of an attack in which he is to participate. (Article 44.3). Protocol I also stipulates that: 'A combatant who falls into

²⁷ Roberts and Guelff, eds, *Documents on the Laws of War* (Oxford: Oxford University Press, 2000), p. 419.

the power of an adverse Party while failing to meet the requirements set forth in the second sentence of paragraph 3 shall forfeit his right to be a prisoner of war, but he shall, nevertheless, be given protections equivalent in all respects to those accorded to prisoners of war by the Third Convention and by this Protocol. This protection includes protections equivalent to those accorded to prisoners of war by the Third Convention in the case where such a person is tried and punished for any offences he has committed' (Article 44.4). 3) It places further restrictions on the methods and means of warfare (Articles 35 to 38), the most important provision of which restates the principle that, 'in any armed conflict, the right of the Parties to the conflict to choose methods or means of warfare is not unlimited', wherein the phrase 'armed conflict' replaces that of 'war' as appearing in the annex of the Fourth Hague Convention; Prohibition on the use of weapons, projectiles and methods of warfare of a nature that will cause superfluous harm or unnecessary suffering; prohibition on the use of methods or means of warfare that will cause widespread, long-term and severe damage to the natural environment. Also, as regards the research, development, acquisition or adoption of any new weapon, means or method of warfare, a High Contracting Party is obliged to determine whether or not they—that is, the weapon, means or method—are prohibited by this Protocol or by the laws of war. (4) Increasing the number of provisions regarding the protection of unoccupied territories and civilians, and strengthening protection of civilians. The second Protocol contains 28 articles and includes the basic substance of the first Protocol, but is simple and rather brief; it does not contain that pertaining to 'the methods and means of warfare'. It is noteworthy that Protocol II was the first international treaty to make law of war providing for the protection of war victims applicable to armed conflict not of an international character (such as civil wars). On one hand, it developed and supplemented Article 3 common to the Four Geneva Conventions of 1949, with regard to the scope of application; but on the other, it also states that, 'This Protocol shall not apply to situations of internal disturbances and tensions, such as riots, isolated and sporadic acts of violence and other acts of a similar nature, as not being armed conflicts' (Article 1). In order to prevent the invocation of this Protocol by other States so as to intervene in the internal affairs of a High Contracting Party, the Second Protocol stipulates that, 'None of the rules in this Protocol shall be invoked as justification for directly or indirectly intervening, for any reason, in the internal or external affairs of the High Contracting Party, within whose territory the armed conflict occurs' (Article 3). Also noteworthy is the growing importance of the Martens Clause as new weaponry and methods of warfare is under rapid development. In terms of legislation, its formulation utilizes a process of 'inclusive' (rather than a process of 'enumeration' or a process

of 'elimination'), making it clear that whether or not international treaties have yet to create a rule, 'Populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the law of humanity and the requirements of the public conscience'. This principle was written into the Fourth Hague Convention of 1907 and Protocol I. In addition, it was decided during the drafting of the first Protocol at the 1977 diplomatic conference that the status of the Martens Clause be raised to a substantive norm, rather than a paragraph in the preamble of the 1973 draft. This act indicates the immense significance of the Martens Clause.²⁸ Upon closer examination, it is apparent that the Martens Clause is tremendously important in its utilization of customary law, which emphasizes the importance of customary law in the regulation of armed conflicts.²⁹ Certain Western scholars believe that even the most complete law cannot encompass all possible situations. The more specific situations for application of a law that are enumerated, the greater the risks are of oversights. Whilst it is impossible to invoke specific provisions in the Four Geneva Conventions of 1949 to solve problems, the Martens Clause would function as a 'safety valve'.³⁰

At the end of the Second World War, the United Nations Charter stipulated each member State's responsibility to settle international disputes peacefully, thereby theoretically eliminating the legality of using force. There are nevertheless two exceptional situations; the first is self-defense. Article 51 of the United Nations Charter provides that: 'Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until such time as the Security Council takes the measures necessary to maintain international peace and security'. The word 'inherent' is used to indicate that 'self-defense' is a State's innate and unalienable right. Another situation is that United Nations Security Council authorizes or undertakes enforcement action according to the United Nations Charter. Article 42 of the United Nations Charter stipulates that: 'Should the Security Council consider that measures provided for in Article 41 are inadequate or have proved to be inadequate, it may take such action by air, sea or land forces as may be necessary to maintain or restore international peace and security'. On the other hand, the development of international relations has given rise to three specific situations' elevation to the status of 'international armed conflict'—that is, armed conflict involving fighting against colonialism, foreign occupation and racist regimes. The expansion of the area to which the

²⁸ Rupert Ticehurst, 'The Martens Clause and the Laws of Armed Conflict' in Li Zhaojie, ed., *Selected Articles From The International Review of The Red Cross*, p. 28.

²⁹ *Ibid.*, p. 30.

³⁰ Stanislav E. Nahlik, 'A Brief Outline of International Humanitarian Law', *International Review of the Red Cross*, July/August 1984, Offprint, pp. 36–7.

laws of armed conflict are applicable has deep and profound political and historical underpinnings. After the Second World War, national liberation movements abounded. States which had been colonized by the West undertook, one after another, various measures (including armed conflict) to achieve independence. At the same time, former suzerains were reluctant to relinquish their vested interests, and did all they could to maintain their former privileges. Fighting between the two factions became more and more intense. Burgeoning numbers of independent States requested protection of their legal rights under international law, which led to the gradual emergence of a situation where the laws of armed conflict were made applicable to civil wars. In actuality, the three situations described in Protocol I to which the laws of armed conflict are also applicable represent and rooted in factual occurrences. For example, 'anti-colonialism' is in reference to Angola, which at the time was a former colony of Portugal; 'foreign occupation' is in reference to the relations between Israel and Pakistan; and 'racist regimes' is in reference to apartheid in South Africa. Developed States of the West and third-world States had very different perceptions as to the handling of these problems. At the time the ICRC conferences were held, fighting between each of the factions was extremely intense. In essence, the application of the laws of armed conflict in civil wars and other similar conflicts is to avoid a situation where there exists a state of legal vacuity and lawlessness. One of the unique characteristics of human society lies in its organization. Even in an extreme situation of anarchy such as war, there still exists a need for regulations of conduct. According to traditional law of war, treaties are only binding between signatory States. This is the so-called 'general participation clause'. However, as it is difficult for either belligerents of internal conflicts to attend international conferences and express their positions towards the laws of armed conflict. The laws of armed conflict should be applicable to all parties of a domestic conflict, regardless of whether or not each of the parties has signed or accessed at an international convention. In this respect, the provisions of the laws of war are similar to the effects of international customary law. If it is possible to establish a rule as applicable under international customary law, it will also apply to all States that have not ratified either the First or Second Protocol and to armed conflicts not of an international nature.³¹

International law provides protection for war victims of both international and non-international armed conflicts, but they are plainly differentiated by their unequal levels, that for war victims of international conflicts being comprehensive and all-encompassing, while protection of war victims of non-international conflicts is lacking. During non-international armed conflicts, there is no concept of a 'protected persons'. The laws of armed

³¹ ICRC, 'Study on Customary Rules of International Humanitarian Law', <http://www.icrc.org/Web/eng/siteeng0.nsf/iwpList74/912774C5011D3922C1256BC80052362E>.

conflict only provide a minimum protection for war victims.³² Protocol II expands the scope of protection offered to war victims under law of war to include non-international armed conflict (such as civil wars), but, at the same time, restrictions are imposed upon internal warfare, such as ‘this Protocol shall not apply to situations of internal disturbances and tensions, such as riots, isolated and sporadic acts of violence and other acts of similar nature’ (Article 1.2). Also, Protocol II is applicable only to ‘armed conflicts within the territory of a High Contracting Party between its armed forces and dissident armed forces or other organized armed groups which, under responsible command, exercise such control over a part of its territory as to enable them to carry out sustained and concerted military operations, and to implement this Protocol’.³³ Some Western scholars hold that the partial territory referred above can generally be understood to mean 30% or more of a State’s territory. Nonetheless, new provisions contained in the Statute of the International Criminal Court adopted in 1998 during Diplomatic Conference in Rome (also known as the Rome Statute of the International Criminal Court) raised the standards for application of the laws of armed conflict to those of civil war. Article 8, paragraph 2 of the Rome Statute of the International Criminal Court regarding its jurisdiction over war crimes stipulates that: ‘Paragraph 2 (e) applies to armed conflicts that take place in the territory of a State when there is protracted armed conflict between governmental authorities and organized armed groups or between such groups’.³⁴ This stipulation further complicates the situation because the term ‘protracted’ describes a key time factor that is difficult to analyse quantitatively. In addition, because the number of States that accessed to Protocol II and the Rome Statute of the International Criminal Court are limited, they are far less universal than the Four Geneva Conventions of 1949 and Protocol I.³⁵ It is consequently still unclear as to whether or not this portion of international regulations has become a part of international customary law.³⁶ Certain scholars might indicate that any violation of provisions within

³² See Article 3 common to the *Geneva Conventions* of 1949.

³³ *Protocol Additional to the Geneva Conventions of August 12, 1949, and relating to the Protection of Victims of Non-International Armed Conflicts* (Protocol II), Article 1, Paragraph 1.

³⁴ See Article 8, paragraph 2 (e), the Statute of the International Criminal Court.

³⁵ German scholar Fleck believes that in recent years, States that did not access to Protocol I participated in military operations involving multinational forces, spurring the acceptance by each State of the standards established by Protocol I as customary law. See Dieter Fleck, ed., *The Handbook of International Humanitarian Law in Armed Conflict* (Oxford: Oxford University Press, 1995), p. vi.

³⁶ According to data from the ICRC, by 9 September, 2002, 190 States were contracting parties to the Geneva Conventions of 1949, 160 States were contracting parties to Protocol I and 153 States were contracting parties to Protocol II. Jakob Kellenberger, ‘International Humanitarian Law at the Beginning of the 21st Century’, 26th Round Table in San Remo on Current Problems of International Humanitarian Law, *The Two Additional Protocols to Geneva Conventions: 25 Years Later Challenges and Protects*, see <http://www.icrc.org/Web/eng/siteeng0.nsf/html/5E2C8V?Open-Document>.

Protocol II is now within the jurisdiction of the International Tribunal of Rwanda. However, because this practice is sparse on an international level, it is unlikely that the cumulative number of occurrences and repeated practice would elevate the status of Protocol II to that of international customary law. In direct contrast, 'the basic principles of the 1948 Convention on the Prevention and Punishment of the Crime of Genocide have been recognized by the International Court of Justice as international customary law'.³⁷ Moreover, 'Without a doubt, after 1946, crimes against humanity had already developed into a part of International Customary Law, a point that was clearly and undoubtedly confirmed during the judgment of the Tadic trial by the International Criminal Tribunal for the Former Yugoslavia. The Statute of the International Criminal Tribunal for Rwanda established crimes against humanity as a separate category of offence, thereby severing its relationship to war crimes. Article 1 of the November 26, 1968 Convention on the Non-Applicability of Statutory Limitations to War Crimes and Crimes Against Humanity reiterates Article 6 of the Nuremberg Charter regarding crimes against humanity, but adds the phrase 'whether committed in time of war or in time of peace'.³⁸

Theoretically speaking, violations of 'provisions of a prohibitive nature' of the laws of armed conflict led to respective culpability. The development of international regulations concerning the penalty for war crimes has, however, long been extremely deficient. The development of international criminal law—the general terms for the substantive and procedural laws that regulate international crimes and penalty—after the Second World War greatly strengthened the efficacy of the laws of armed conflict. This law primarily includes principles, rules, and institutions that focus on the prevention and punishment of international crimes at international plain and on mutual assistance and cooperation to achieve international criminal justice. It has recently been commonly acknowledged within the Western law circle, however, that international crime has become subdivided into the main categories of 'war crimes', 'crime against humanity', 'crime of genocide', and of the other most heinous offences,³⁹ while all other types of international crimes are left to domestic jurisdiction. In the situation where warfare is considered to be a State's intrinsic right, an action will be considered a war crime only if it violates widely recognized laws and customs of war. This type of action, such as killing and looting of soldiers that have lost their ability to fight due to injury or sickness or who have already put down their arms,

³⁷ Reservations to the Convention on the Prevention and Punishment of the Crime of Genocide, Advisory Opinion of May 18, 1951, I.C.J. Reports, 1951, p. 23. See Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', pp. 531–53.

³⁸ Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', pp. 531–53.

³⁹ See Article 5 of the Statute of the International Criminal Court.

not only violates international law but typically also violates the municipal criminal law of the transgressor's home State. Prior to the Second World War, laws concerning the legal responsibilities of those in violation of law of war were scattered amongst various treaties. Article 56 of Annex to the Hague Convention of 1907 stipulates that anyone engaging in willful damage to priceless historical monuments or works of art shall be punished according to law; Articles 28 to 30 of the 1929 Geneva Convention that ameliorated the treatment of the sick and wounded also includes provisions on the punishment for misuse of the Red Cross emblem and for any conduct that violates the convention (such as abuse of the sick and wounded and injury to medical personnel). Article 3 of the Washington Treaty of 1922 stipulates that any person in the service of any Power that violates the humane rules of existing law concerning the attack, seizure and destruction of merchant ships, whether or not such person is under orders of a superior, shall be deemed to have violated the laws of war, and shall be liable to trial and punishment as if for an act of piracy, and may be brought to trial before the civil or military authorities of any Power within the jurisdiction of which he may be found.

If it were to be said that the two Hague Peace Conferences took the formerly sparse and diffuse principles and rules of warfare that existed as a result of conventional use and turned them into the foundations for law of war, then it was the Four Geneva Conventions of 1949 and the two Additional Protocols of 1977 that were drafted after the Second World War that served to further codify and finalize law of war. Law of war developed from principles and rules limiting the methods and means of war, to protecting war victims into the most complete branch within the international legal system. Moreover, owing to changes in international politics, the range of situations to which the principles and rules of the law of war are applicable expanded to include internal armed conflict, and as a result prevented the state of lawlessness that arises from civil war. The practices of the international criminal justice system in punishing war crimes also 'caused important changes in the international responsibilities of States and their public officials'.⁴⁰

Law of War Safeguarded and Supplemented by International Criminal Justice

Over the course of history, violations of law of war (the laws and customs of warfare) have been common occurrences, yet the international community was for a very long period of time incapable of punishing transgressors according to the law. In the aftermath of the First World War, the victorious States called for the trial and punishment of the German Emperor Kaiser

⁴⁰ Liang Xi, ed., *International Law*, p. 25.

Wilhelm II and other war criminals. The Treaty of Versailles of 1919 stipulated that Wilhelm II and other members of the German armed forces were in severe violation of law of war and must be turned over for trial either at an international court or at a mixed tribunal.⁴¹ The contracting States, however, did not want to set this kind of precedent, and used the Dutch government's refusal to extradite Wilhelm II as an excuse to leave the situation unresolved. Germany, in the meantime, used 'national pride' as a reason not to handover other war criminals. On 8 August, 1945, the Soviet Union, the United States, the United Kingdom and France, acting under the provisions of the Moscow Declaration of 1944, agreed to establish an International Military Tribunal, and in London signed the Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis (the 'London Agreement') and also the additional Charter of the International Military Tribunal. The London Agreement stipulated the jurisdiction, organization and mission of the International Military Tribunal, while the Charter of the International Military Tribunal stipulated the organization, trial and judgment procedures of the court. Article 6 of the Charter lists the categories of war crimes: (a) 'crimes against peace', namely, planning, preparation, initiation or waging of a war or aggression, or a war in violation of international treaties, agreements or assurances, or participation in a common plan or conspiracy for the accomplishment of any of the foregoing; (b) 'war crimes', namely, violations of the laws or customs of war. Such violations include but are not limited to the murder, ill-treatment or deportation of forced labor or any other purpose of civilian population of or in occupied territory; the murder or ill-treatment of prisoners of war or persons of the sea, killing of hostages, plunder of public or private property, wanton destruction of cities, towns or villages, or devastation not justified by military necessity; (c) 'crimes against humanity', namely, murder, extermination, enslavement, deportation and other inhumane acts committed against any civilian population before or during the war, or persecutions on political, racial or religious grounds in execution of or in connection with any crime within the jurisdiction of the Tribunal, whether or not in violation of the domestic law of the State where perpetrated. Leaders, organizers, instigators and accomplices participating in the formulation or execution of a common plan or conspiracy to commit any of the foregoing crimes are responsible for all acts performed by any persons in execution of such plan. International trials for major fascist German war criminals were organized in Nuremberg, Germany in accordance with the London Agreement and the Charter of the International Military Tribunal. The Nuremberg Trials by the International Military Tribunal were held from November 10, 1945 to October 1, 1946 and sentenced a total of 12 accused to death by hanging, seven accused to imprisonment, and also declared the Nazi Party

⁴¹ See The Versailles Treaty of June 28, 1919, Articles 227–9.

Directorate, the Gestapo (Nazi secret police) and the SA (private army of the Nazi Party) criminal organizations. Empowered by the Allied Nations, on 19 January, 1946 the supreme command of the allied forces issued the Charter of the International Military Tribunal for the Far-East, the content of which was essentially the same as that of the Charter of the International Military Tribunal for Europe. It states that the International Military Tribunal for the Far-East comprises the representatives of 11 States—China, the Soviet Union, the United States, the United Kingdom, France, Holland, India, Canada, New Zealand, the Philippines, and Australia. The tribunal put major Japanese war criminals on trial in Tokyo from April 1946 to November 1948, sentencing seven of them to death by hanging and 18 to imprisonment. The Charters of the International Military Tribunals and the Trials of Nuremberg and Tokyo had a profound impact on the development of law of war. The principles included in the Charter of the International Military Tribunal for Europe were unanimously affirmed in Resolution 95 (I) of December 11, 1946 of the United Nations General Assembly. Certain Italian scholars believe that the General Assembly of the United Nations' affirmation of the provisions of the London Charter is characteristic of customary law.⁴² In 1950, International Law Commission of the United Nations formulated the 'Principles recognized in the Charter of the Nuremberg Tribunal and in the Judgment of the Tribunal' through resolution 177 (II) of the General Assembly, stating that: (1) any person who commits an act which constitutes a crime under international law is personally responsible and as such is liable to punishment; (2) that the act is not in violation of internal law within the host State does not exempt responsibility for it under international law; (3) the status of the defendant does not exempt him from responsibility under international law; (4) that the act was an order by the government or superior does not exempt it from responsibility under international law; (5) any person charged with a crime in violation of international law has a right to a fair trial; (6) the crimes in violation of international law are crimes against peace, war crimes and crimes against humanity; (7) collaboration in the foregoing crimes is a crime under international law. To this, Italian scholar Edoardo Greppi comments: 'Many judgments were passed during the Nuremberg Trials (as well as the less influential Tokyo Trials), and from the perspective of precedence law these judgments have largely influenced the formation of individual criminal responsibility under international law'.⁴³ After the Nuremberg and Tokyo Trials, further trials of war criminals were held in the national courts of other States such as the Soviet Khabarovsk Trials and China's Shenyang, Nanjing and Taiyuan Trials. The General Assembly of the United Nations

⁴² Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', pp. 531–53.

⁴³ *Ibid.*

adopted a resolutions in 1967 and 1968 stipulating that war criminals do not have a right to seek asylum nor do legal prescriptions apply to them (such as periods of prescription in a statute of limitations).

In 1993, the United Nations Security Council adopted S/RES/827 (1993) that established an International Criminal Tribunal for the Former Yugoslavia wherein to prosecute the persons responsible for serious violations of international humanitarian law since 1991 in the territory of the former Yugoslavia. In 1994, the United Nations Security Council passed Resolution 935 (1994) establishing International Criminal Tribunal for Rwanda wherein to prosecute the persons responsible for genocide and other serious violations of international humanitarian law in the territory of Rwanda, and also to prosecute the Rwandan citizens responsible for genocide and other serious violations of international humanitarian law in the territory of States neighboring Rwanda between January 1, 1994 and December 31, 1994. The International Tribunal for Rwanda was also the first court to bring into its of jurisdiction acts in violation of Article 3 common to the Four Geneva Conventions of 1949 and Protocol II.⁴⁴

On the one hand, the establishment of these *ad hoc* international tribunals and their judgments spurred the development of law of war, but on the other hand, they prompted a series of contentious legal theories. For instance, during the Nuremberg Trials, the defendants raised a series of defense stating that their acts had not violated international law in effect at the time of the Second World War, and that the trials held by the International Military Tribunal were in violation of the fundamental principle of legality (*nullum crimen sine lege*). The argument was that in actuality, the trial was *ex post facto*,⁴⁵ and the law should never bear retroactivity. Also, that the

⁴⁴ With regard to 'the crime of violating Article 3 Common to the Geneva Conventions of 1949 and Protocol II', the International Criminal Tribunal for Rwanda believes that five conditions must be satisfied in order to ascertain that such crime has been committed: (1) the criminal acts being charged must have been committed during non-international armed conflict; (2) must fall within the time frame in which the two conventions are binding; (3) must be within the territorial parameters set forth by the two conventions; (4) the accused must be related to a party involved with one of the two conventions; (5) victims of the criminal acts being charged must be under the protection of the two conventions. See Ling Yan, 'Lianheguo Luwangda guoji fating 1999nian de gongzuo' ('The Work of the International Criminal Tribunal for Rwanda in 1999'), *Zhongguo guojifa niankan* (*Chinese Yearbook of International Law* 1999) (Beijing: Falu chubanshe, 2002), p. 399.

⁴⁵ However, with regard to this question, Professor Bassiouni holds that, 'The basis moral dilemma in 1945 was whether the extension of international criminal responsibility to perpetrators of Article 6(c) "crimes against humanity"—which constituted a breach of strict legal positivism—was a greater or lesser breach than to allow such perpetrators go unpunished. To uphold the first would be to give greater weight to a given legal theory over the substantive harmful outcome deriving from an undeniable gross abuse of power, which, even according to positivists, begot neither good nor order. The law of the Charter, however, violated the "principle of legality" as applied under then existing French Civilist and Germanic legal traditions, though not necessarily under the views of some common law exponents.' See M. Cherif Bassiouni, *Crimes Against Humanity in International Criminal Law* (The Hague: Kluwer, 1999), 2nd edn., pp. 80–81, 108.

operation of the International Military Tribunal was concerned only with the actualization of the so-called 'victor's justice', since many actions undertaken by the armed force of the Allied Powers should also be considered unlawful (such as the bombing of the Dresden by the British air force). Justice Pal, the Indian Judge of the Tokyo Tribunal, would have acquitted all the defendants on the ground that there had been no individual criminal responsibility under international law.⁴⁶

The theoretical problems discussed above embody academic significance and also important practical implications. Deliberately evading the situation rather than finding reasons to refute these arguments would make them highly detrimental to the development of the laws of armed conflict and international criminal law. An analysis of those issues by the defendants of the Nuremberg and Tokyo Trials will, by means of well-formed arguments and meticulous explanations, produce answers to them. From a theoretical perspective, neither the Nuremberg Trials nor the Tokyo Trials violated the basic principle of *nullum crimen sine lege*. Although international law at the time lacked obvious and concrete written penal code norms, as stated in the judgment of the Nuremberg International Military Tribunal: 'To assert that it is unjust to punish those who in defiance of treaties and assurances have attacked neighbouring states without warning is obviously untrue, for in such circumstances the attacker must know that he is doing wrong, and so far from it being unjust to punish him, it would be unjust if his wrong were allowed to go unpunished'. The ruling also analyses the Kellogg-Briand Pact signed in Paris in 1928 (also known as the Pact of Paris) that renounces war, and to which Germany, Italy and Japan were all signatories. The ruling states that, 'In the opinion of the Tribunal, the solemn renunciation of war as an instrument of national policy necessarily involves the proposition that such a war is illegal in international law; and that those who plan and wage such a war, with its inevitable and terrible consequences, are committing a crime in so doing'.⁴⁷ In this sense, rather than violating the principle of legality, the Nuremberg International Military Tribunal actually spurred the development of international criminal law. Furthermore, the judgments of the Nuremberg Tribunal and the subsequent Tokyo Tribunal have been

⁴⁶ See Kriangsak Kittichaisaree, *International Criminal Law* (Oxford: Oxford University Press, 2001), p. 20.

⁴⁷ It is worthwhile to notice that the Nuremberg International Military Tribunal listed a series of international treaties as evidentiary support towards convicting the defendants, including The British War Blue Book; The Bullet Decree (Karl Erlass), March 4, 1944; Directive No. 1 for the Conduct of the War; The Hague Conventions; Hossbach Memorandum, November 10, 1937; Kellogg-Briand Pact and Associated Documents; Moscow Conference, October 1943; Munich Pact and Associated Documents; Nazi-Soviet Non-Aggression Pact; Night and Fog Decree (Nacht und Nebel Erlass), December 7, 1941; Program of the Nationalist Socialist German Workers' Party (NSDAP); The Strop Report; The Warsaw Ghetto is No More, April-May 1943; The Treaty of Versailles, June 28, 1919 and Tripartite Pact. See <<http://www.yale.edu/lawweb/avalon/imt/imt.htm&isbol=0>>.

affirmed by the United Nations General Assembly as well as the United Nations International Law Commission. As to the charge of 'crimes against humanity' in the Nuremberg rulings, Professor Bassiouni maintains that such charges in international criminal law can arise from analogy, and that are not in violation of the principle of *nullum crimen sine lege*.⁴⁸

The relationship between international criminal law and the principle of *nullum crimen sine lege* must be approached from the following perspectives in order to arrive at an objective, practical and convincing conclusion. First of all, it is impossible for provisions of statutory law to address all the various societal situations that require legal redress. According to 'general principles of law' and the judicial practice of each State, law can be developed in the course of trials and judgments. For example, both 'marital rape' and 'male rape' are not stipulated in penal code, but came out through judicial practice in Western States. The International Criminal Tribunal for the Former Yugoslavia and the International Criminal Tribunal for Rwanda determined that 'thrusting a piece of wood into the sexual organ of a woman as she lies dying is rape' (*Akayesu Case*), 'forced penetration of the mouth which is a humiliation and degrading attack on human dignity can be considered rape' (*Furundzija Case*) and forcing two men to perform fellatio on one another could constitute rape (*Celebici Case*).⁴⁹ This ruling does not violate the principle of *nullum crimen sine lege*. Secondly logically deducting a concrete law to apply to a particular case from an abstract legal principle is permitted in internal law of most States. Moreover, the diversity that has emerged within the existing form of the law is manifest in the existence of statutory law as well as customs. It is submitted that the key to determining and judging the legality of certain charges and penalties lies in whether or not they can satisfy political legitimacy. What should be noted is that, should any dispute as to whether or not the international trials held at the end of World War II adhere to the principle of *nullum crimen sine lege* arise, it is the development of post-war international criminal law that has already sorted out the problem. This argument could find supporting evidence from International Covenant on Civil and Political Rights, in which Article 15 Paragraph 1 states that: 'No one shall be held guilty of any criminal offence on account of any act or omission which did not constitute a criminal offence, under national or international law, at the time when it was committed.'

⁴⁸ M. Cherif Bassiouni, *Crimes Against Humanity in International Criminal Law* (The Hague: Kluwer, 1999), 2nd edn., p. 145. Some Chinese scholars also maintain that the principle of *nullum crimen sine lege* is not applicable (or not entirely applicable) in international criminal law. See Huang Zhaojiong and Tang Xuelian, 'The Nuremberg, Tokyo Trials and International Criminal Law', <http://www.ciol.com.cn/lawthesis.asp>; Liu Yaying, 'Guoji xingfa yu guoji fazui' ('International Criminal Law and International Crimes') (Beijing: Qunzhong chubanshe, 1986), pp. 48–50; Huang Fang, 'Guoji fazui de guonei lifa daolun' ('An Introduction to International Crimes in Domestic Legislation'), *Faxue pinglun* (*Law Review*), No.2, 2000.

⁴⁹ Kriangsak Kittichaisaree, *International Criminal Law*, pp. 112–3.

Nor shall a heavier penalty be imposed than the one that was applicable at the time when the criminal offence was committed. If, subsequent to committing the offence, provision is made by law for the imposition of the lighter penalty, the offender shall benefit thereby'. The second clause, however, states that: 'Nothing in this article shall prejudice the trial and punishment of any person for any act or omission which, at the time when it was committed, was criminal according to the general principles of law recognized by the community of nations'.⁵⁰ On a separate issue, 'the Draft Code of Crimes against the Peace and Security of Mankind, drawn up by the International Law Commission of the United Nations, had major impact on the development of the concept of individual criminal responsibility. In the drafts of 1951 and 1954, Article 1 states that, "offences against the peace and security of mankind are crimes under international law, for which the responsible individuals shall be punished". In the text of the 1996 draft, Article 1 states that, "crimes against the peace and security of mankind are crimes under international law and punishable as such, whether or not they are punishable under national law". Moreover, according to Article 2, "A crime against the peace and security of mankind entails individual responsibility".⁵¹ What should be explained is that although these draft codes do not bear binding force, on a certain level they do reflect the future direction and trend of international criminal legislation.

The legality of the International Criminal Tribunal for the Former Yugoslavia, owing to its establishment by the United Nations Security Council, is questionable and is repeatedly challenged by defendants. During the 1995 Tadic Trial, the defendant maintained that the Tribunal was illegally created and as such had no right to conclude judgment upon him. When, at his hearing in July 2001, former Serbian and Yugoslavian President Slobodan Milosevic was asked to plead guilty, he also pointed out the illegality of the court's existence because it had not been established by the General Assembly of the United Nations.⁵² Aside from the International Criminal Tribunal for the Former Yugoslavia and the International Criminal Tribunal for Rwanda, establishment of an International Tribunal for East Timor is also proposed.

Contrary to what many scholars believe, the range of issues to which international criminal law can be applied is shrinking rather than expanding. The four crimes that fall under the jurisdiction of the International Criminal

⁵⁰ Article 15, International Covenant on Civil and Political Rights, in Wang Tiewa and Tian Ruxuan, eds., *Guojifa ziliao xuanbian (Selected Materials of International Law)* (Beijing: Falu chubanshe, 1986), pp. 172-3.

⁵¹ Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', pp. 531-53.

⁵² 'Milosevic hearing transcript', Transcript of Slobodan Milosevic's first appearance at the UN War Crimes Tribunal for the former Yugoslavia, http://news.bbc.co.uk/hi/english/world/Europe/newsid_1419000/1419971.stm; and 'Court Rejects Milosevic Challenge', http://news.bbc.co.uk/hi/english/world/rope/newsid_1517000/1517905.stm.

Court comprize the crime of genocide, crimes against humanity, war crimes and crimes of aggression. Only those most heinous crimes that endanger the welfare of the international community fall under the adjustment of international criminal law, whilst these less severe offences are left to each respective State's domestic jurisdiction. *International Criminal Law* by a Thai scholar also focuses exclusively on the crime of genocide, crimes against humanity, war crimes and crimes of aggression.⁵³

Moreover, it is worthwhile to notice that an important development should be mentioned here in relation to the practice of codifying international law: there is a growing connection between humanitarian law and human rights law. Indeed, some recently adopted provisions of humanitarian law appear clearly influenced by human rights rules and standards of protection. The Rome Statute refers to concepts like 'personal dignity', the prohibition of 'humiliating and degrading treatment', 'judicial guarantees', the prohibition of 'persecution' (as 'intentional and severe deprivation of fundamental rights contrary to international law by reason of the identity of the group or collectivity'), discrimination and apartheid. These concepts have all been established in the main instruments adopted by the UN for the protection of the rights of the individual. However, the principle of humanity is at the core of international humanitarian law and forms the basis of all the developments.⁵⁴

Looking at the system of armed conflict law in its entirety, certain Western scholars believe that aside from the 'law of Geneva' (or Red Cross Law) and the 'law of The Hague' the so-called 'law of New York' also need to be taken into account. According to two Dutch scholars, everything that the United Nations has achieved towards the development of the laws of war since its establishment may also be referred to as 'law of New York'. The purpose of these laws is five-fold: (1) to facilitate the establishment of the principles relating to the punishment of war crimes; (2) to appeal for the prohibition of or limitations upon the use of nuclear weapons as well as certain conventional weapons; (3) to urge the respect of human rights during armed conflict; (4) to protect the person and rights of women and children as well as the right of reporters to report during warfare; and (5) to raise the status of national liberation movements to that of international armed conflict, and also to advocate that combatants and guerrilla forces also enjoy prisoner-of-war status. The Dutch scholars also believe that the 'law of the Hague', the 'law of Geneva' and the 'law of New York' are like three distinct tributaries slowly merging. While they remain distinct, their relationship with international criminal law has been growing closer

⁵³ Kriangsak Kittichaisaree, *International Criminal Law*.

⁵⁴ Edoardo Greppi, 'The Evolution of Individual Criminal Responsibility Under International Law', pp. 531–53.

since the 1990s.⁵⁵ Yet, this characterization of law of war is the belief of a minority only, and has yet to be widely accepted. There is no actual consensus among Western academic circles on this issue. For example, some British scholars (like Mr Dino Kritsiotis of Nottingham University) believes that the formulation of the 'law of New York' was based on a political agenda, and its success is as a result of its unconventional nature and public appeal. Certain prerequisites must exist in order for a new legal system to be created, according to a leading scholar of international law. Since the end of World War II, relations between States have rapidly developed, creating a number of new areas within the international community that are in need of legal redress. At the same time, the number of international organizations has also increased sharply, and new branches of international law have also appeared in rapid succession. With regard to the standards and qualifications necessary for the establishment of a new field within international law, Professor Liang Xi, a well-known Chinese publicist, holds that, 'each specialized field must have developed into an independent branch of international law with its mature system'.⁵⁶ In the 1980's, an international lawyer in the Soviet Union once suggested two criteria through which to classify branches of international law: the first one is the issue that law is attempting to cope with (i.e., special areas of international law and the purpose behind the international legal norms used to regulate international relations in these particular areas); the second is the means of regulation. For the latter, international lawyers always use three criteria that must be contained unique principles that make up any new branch of international law: States' interest in regulating a particular type of societal relations (as clearly expressed during the act of compiling), the important international implications of each societal relationship, and sufficient norms of international law.⁵⁷ Whether or not a certain group of international regulations satisfy qualifications necessary to become a new branch of law, or the threshold to become a new subdivision of international law, the aforementioned criteria can seemingly determine.

For a long time, the principles and rules of law of war remained only as superficial responses to moral necessities that lacked coercive strength needed to ensure their enforcement. Positivists who emphasized that 'the validity of international law lies in the will of States', whose perspective has long held dominant position on this issue. The advent of international

⁵⁵ Refer to Dutch Scholars Frits Kalshoven and Liesbeth Zegveld, *Constraints on the Waging of War: An Introduction to International Humanitarian Law* (Geneva: International Committee of the Red Cross, 2001), pp. 19, 29–32.

⁵⁶ Liang Xi, ed., *International Law*, p. 1.

⁵⁷ Duan Ruilin, 'Sulian guojifa xuehui diershibajie nianhui' ('The 28th Annual Conference of Soviet Union Society of International Law'), *Zhongguo guojifa niankan 1986 (Chinese Yearbook of International Law 1986)* (Beijing: Zhongguo duiwai fanyi chubanshe, 1986), p. 395.

criminal law marked a breakthrough from traditional stereotype of law of war, providing the strongest guarantee by virtue of criminal law with coercive enforcement to implement law of war.

New Developments in Law of War

Over the last 10 years, law of war has undergone many new changes. In October 1995, the Vienna Diplomatic Conference issued the Protocol on Blinding Laser Weapons (the Fourth Protocol of the 1980 Convention on Certain Conventional Weapons), prohibiting both the use and transfer of laser weapons, one of whose combat specific functions is to cause permanent blindness. The Protocol also requires States to take all appropriate precautions, including training of armed forces, to avoid causing permanent blindness by the lawful use of other laser systems. The use of land mines is the focus of the Second Protocol Additional to the Geneva Convention of October 10, 1980, on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, the regulations of which became more inclusive with the passing of the Amended Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices in Geneva on May 3, 1996. In addition, from December 3 to 4, 1997, the representatives of 121 States converged in Ottawa to sign the Convention on the Prohibition on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, which completely banned the use of land mines. This convention also includes provisions on demining, providing aid to victims of land mines. Another new development in law of war is that of San Remo Manual on International Law Applicable to Armed Conflicts at Sea. Developed by the International Institute of Humanitarian Law under the direction of the ICRC, the San Remo Manual received world-wide recognition, and its significance was evident in the resolutions of the 26th International Conference of the Red Cross and Red Crescent held in Geneva in 1995. The passing of the Rome Statute of the International Criminal Court on 17 July, 1998 was also an important step in ensuring that all States abide by international humanitarian law. The newly established International Criminal Court was given jurisdiction over all war crimes, regardless of whether committed during international armed conflict or non-international armed conflict. Eventually, the most recent development in law of war relates to the means of war. In December 2001, the scope of the 'Geneva Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to be Excessively Injurious or to Have Indiscriminate Effects of October 10, 1980' was extended. In the past, this Convention had only covered situations of

international armed conflict, but Second Review Conference amended Article 1 to include situations of non-international armed conflict.⁵⁸

The type of warfare in the early 21st century had an important impact on the laws of armed conflict. The terms 'formality of warfare' and 'type of warfare' are often used to characterize broadly the unique aspects of a war or armed conflict, but both phrases simply describe the basic features of war. War and armed conflict are generally characterized from two perspectives: the first one is based upon the relationship between the belligerents. For instance, when armed conflict erupts between the forces of two States, it is called 'international war' or 'international armed conflict'; when armed conflict breaks out between government and anti-government forces, it is called an 'internal war' or 'non-international armed conflict'; when armed conflict erupts between one State and non-governmental armed forces in another State, it is called a 'natives' war' or 'war against terrorism'. This perspective stresses the subject of warfare. The second way to classify a war is according to the kinds of weapons and tactics employed in hostilities. For example, a war fought predominantly by aircraft is called an 'aerial war'; when information systems are used as weapons it is called 'warfare of information'; when space-based weapons are used it is called a 'space war'; and when nuclear weapons are mainly used it is called a 'nuclear war'. This perspective stresses the means of warfare. Of course, there are other perspectives from which to characterize the generalities of any particular war, but the aforementioned two are most prevalent. The former perspective looks at the political relations in war, while the latter focuses on the impact of science and technology on warfare.

From the above classifications it can be seen that the formality of war is influenced by two factors, first, changes in political relations, and secondly developments in military technology. Judging by situations at the turn of the millennium, these two factors are both undergoing great changes. In terms of international politics, the onset of severe terrorism led to a war on terror. Such conflicts between a State or a group of States, and a non-State actor, or armed conflict between two terrorist groups, are all unprecedented forms of war. In terms of military technology, the swift and drastic development of modern high-technology led to rapid and successive emergence of new weapons, such as information systems weapons and space-based weapons. Other newly developed weapons include mini nuclear weapons and genetic weapons, the use of which will lead to the rise of warfare of information, space war and nuclear war on small scale.

The impact of changing forms of war on the laws of armed conflict is inevitable, as legal norms' purpose is to regulate conduct of combatants, the means and methods of war. The laws of armed conflict will change and

⁵⁸ International Committee of the Red Cross, *International Humanitarian Law: Answers to Your Questions* (Geneva: International Committee of the Red Cross, 2002), p. 21.

develop in line with the changes and developments in international political relations and military technology that occurred during the early part of the 21st century.

Although new developments in the means and methods of war have expanded the central components of warfare as well as the scope of application of the principles and rules of law of war, they still face new challenges in the new century. The establishment and operation of the International Criminal Court have allowed the enforcement mechanism of law of war to diversify and flourish, but have also created a latent corrosive force that threatens each State's judicial sovereignty.

Conclusion

The reasons underlying the development of law of war are based on the fact that every State that fights a war is bent on using the maximum amount of force in order to achieve its aims. Wars consequently and invariably cause irreparable harm to human lives and welfare. The self-perpetuating philosophy that 'an eye for an eye' is nothing more than a magnification of the 'blood feuds' of primitive society. If allowed to develop, this approach would lead to sustained and harrowing losses on either party to any particular conflict. Analyzed from a chronological perspective, law of war have developed at pace with world history, having evolved since ancient times. Yet traces of law of war that existed in ancient civilizations became applicable to the means and methods of war only during the age of cold arms. Since modern times, the application of law of war has been the result of humanitarian considerations within ethics of politics, as epitomized by the appeals of Henri Dunant. The passing of the two Hague Peace Conventions of 1899 and 1907 gave rise to strong international public opinion that demanded of nations that they, to a certain extent, take measures to achieve political morality and justice by constraining the means and methods of war and protecting war victims. This stage in the development of law of war has since been enriched and systematized. It was during this period that the establishment and rise of the greatest advocate and promoter of laws of war—the International Committee of the Red Cross—occurred. From the midst of diffuse practices of international justice systems also emerged initial attempts to guarantee the enforcement of law of war by international criminal justice. At end of World War II, the creation and judgments of the two international military tribunals set the foundations for international criminal law, initiating an era where enforcement of law of war can be achieved through threat of punishment by the international criminal judicial system. The Rome Statue of the International Criminal Court in 1998 placed values of major Western powers in a dominant position, challenging other States' judicial sovereignty in the name of protecting international human rights. Since the 1970s,

changes in the international political arena have made the principles and rules of law of war also applicable to internal armed conflict, namely those initially applicable to international war or armed conflict only. This has led to the expansion of the subjects and objects of law of war. It is also noteworthy that since its inception, law of war has been caught between a 'State's right to recourse to war' and 'the idea of law of war to provide a minimum of protection for war victims'. In other words, it has fought to survive and develop within the tiny gap that exists between 'military necessity' and 'humanitarian principles'. On another note, owing to flaws within the logic that comprises legal norms, law of war continues to be overly reliant upon international criminal law. In the foreseeable future, it will not be possible for law of war to cast off the restrictions imposed by international politics.

One noticeable trend is that the area over which Western powers have territorial jurisdiction is, with the help of the enforcement mechanism of law of war, slowly encroaching into foreign territory. According to traditional theories of international law, a State's internal law is effective only within its national boundaries. In recent years, however, certain States have attempted to impose their internal jurisdiction abroad, such as during the dispute between the United States and Europe over the problem of economic relations with Cuba. In 1995, American Senator Jesse Helms, chair of the Foreign Relations Committee in the Senate, along with Representative Dan Burton jointly proposed the famous Helms-Burton Act, which strengthened the long-term economic sanctions on Cuba and prohibited foreign companies from investing in the island. This led to strong resentment and opposition among European States, who called into questioning the legitimacy and validity of American municipal legislation beyond its territory. With regard to the implementation of international humanitarian law, certain Western States appear to have taken things one step further. Belgium, for instance, passed a bill in 1993 listing 20 crimes over which the Belgium courts are entitled to exercise universal jurisdiction, regardless of the nationality of the offender, victim's nationality, or the venue in which the crime was committed.⁵⁹ A Belgian court ruled charges against two Rwandan nuns on account of severe violation of international humanitarian law during Rwandan Civil War in 1994. On June 8, 2001, Sister Kisito and Sister Gertrude went to trial in Brussels for these serious crimes. The court sentenced the two nuns (for participating in the 1994 Rwandan genocide) to fixed-term imprisonment; 36-year-old Sister Kisito was sentenced to 12 years and 42-year-old Sister Gertrude to 15 years in prison.⁶⁰ The traditional theory of international law

⁵⁹ Sonja Boelaert-Suminen, 'Grave Breaches, Universal Jurisdiction and Internal Armed Conflicts: Is Customary Law Moving Towards a Uniform Enforcement Mechanism for All Armed Conflicts?' *Journal of Conflict and Security Law*, Vol. 5, No. 1, June 2000, p. 89.

⁶⁰ 'Nuns Sentenced', <http://www.huash.com/wenjian/2001-06-11/2001-06-11-18tpxw6.asp>.

provides that, whether or not these two nuns had been guilty of violating international humanitarian law, Belgium has no right to press charges against them because they are not citizens of Belgium. Moreover, the crimes and consequences thereof did not occur in Belgian territory and the victims are not Belgian citizens. Whether or not a State enjoys the right to exercise universal jurisdiction over such crimes is still a point of contention in theory and in practice as well. A British international lawyer believes that crimes against humanity and war crimes are invariably grouped together, as they incur the same level of condemnation and require similar measures to enforce punishment and prohibition, 'To this day, one cannot claim that existing international law contains general rules empowering States to punish foreign citizens for committing crimes against humanity in the same way as they do for other crimes, such as those empowering States to punish acts of maritime piracy'.⁶¹ Belgium's actions, when compared with the four principles governing traditional national jurisdiction (i.e., jurisdiction by nationality, territorial jurisdiction, protective jurisdiction and universal jurisdiction), constitute a major breakthrough. In accordance with Belgium's view, this principle of 'non-connection' can seemingly only be applied to a very limited number of areas. There is an option to select appropriate defendants, so-called double standard can apply to the plaintiffs of certain States. A newspaper story of 23 December, 2001 reports that 28 Pakistanis in a Belgian court raised charges against Israeli Prime Minister Ariel Sharon for committing crimes against humanity. They accused Sharon of ordering the Israel-backed Lebanese militia to massacre refugees in two Pakistani refugee camps situated near Beirut. Sharon's attorneys countered the lawsuit, stating that the Belgian court did not have the right to try a case involving a head of state.⁶² Elie Hobeika, former Lebanese militia commander and former minister of Hydraulic Resources, agreed to testify against Sharon, claiming that he held new evidence regarding the massacre. However, on 24 January, 2002, Hobeika was assassinated in Beirut, the capital of Lebanon.⁶³ On 13 February, 2002, Israel recalled its ambassador to Belgium as a gesture of retaliation against a ruling made against Sharon by the Belgian Supreme Court. This ruling states that once Israeli Prime Minister Sharon leaves office, charges of war crimes will be pressed against him.⁶⁴ As an American scholar once said, 'law is the boxing

⁶¹ Sir Robert Jennings and Sir Arthur Watts, *Oppenheim's International Law* (Ninth Edition), trans. Wang Tieya et al., Vol. 1, Book 2 (Beijing: Zhongguo dabaike chubanshe, 1998), pp. 363–4.

⁶² 'Sharon was charged by 28 Parlestines for crimes against humanity in Belgium', <http://www.21dnn.com.cn/593/2001-12-26/395@714979htm>.

⁶³ 'Mystery in Assassination of former Lebanese Minister', http://www.yewb.com/gb/content/2002-01/25/content_308244.htm.

⁶⁴ 'Belgian Court Rules Sharon to Face Charges Once Leaves Office, Israel Calls Back Her Ambassador to Belgium', <http://news.sina.com.cn/w/2003-02-13/084349529s.shtml>.

ring for different powers'.⁶⁵ Some Western States attempt to elevate their own values into an international standard, a situation that is not beneficial to developing States and will produce detrimental effects on the development of international legal order.⁶⁶

It is acknowledged that the course of development of law of war to date has been fraught with conflict and compromise between various sects of political ethicists and legalists. This is the result of the combined strengths of the international community. Present day law of war (in common with international legal systems as a whole) are also a product of the gaps between various sects of legal doctrines, and the influences of international law and human rights law on them cannot be underestimated. The Charter of the Nuremberg International Military Tribunal and the Charter of the International Military Tribunal for the Far-East created at the end of World War II both contain an important principle which states, 'That the act was an order by the government or superior does not relieve the defendant's responsibility under international law'. Even so, the International Military Tribunal for the Former Yugoslavia appeals hearing for the famous *Erdemovic Case* showed flexibility. The Chamber of Appeals maintained that while an order from a superior cannot be used as a ground to exclude the defendant's culpability, 'duress under extreme circumstances caused by a superior order' could be considered an appropriate mitigating circumstance in sentencing.⁶⁷

The logic innate to warfare evokes in belligerent parties the desire to exploit violence to the maximum extent. Limiting the means and methods of war, therefore, is equivalent to limiting the level and range of violence employed in it, thereby keeping unnecessary suffering to a minimum. In this regard, law of war brings morality and compassion to this brutal, abnormal phenomenon of human society as a touch of rationality. Even so, the utility and influence of law of war is, in itself, considerably limited. There is still a large gap between the ideal of putting an end to war and the harsh reality that wars continue to exist, and it is something that humanity cannot change in the foreseeable future. All that can be done is to try to shrink this gap, to build some fences around the tyrannical war god and to demarcate the final boundaries within which the use of force is legal. Only then will the laws of armed conflict serve an indispensable, albeit limited, role.

Taking a glimpse at the arduous evolutionary process of law of war within the vast context of historical evolution of human society at large would evoke the emotions of remorse for the destruction of human cultures by war,

⁶⁵ Xin Chunying, 'Hunyinfa xiugai: qinggan chongtu yu lixing xuanze' ('Amendments to "Marriage Law": Emotional Conflict and Rational Choice'), *Dushu (Reading)*, No. 6, 2001.

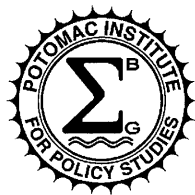
⁶⁶ 'Belgium Suffers a Great Setback in Universal Jurisdiction', <http://news5.hotoa.com.cn/index.cfm/194949.html>.

⁶⁷ See The Appeals Chamber's Judgment, October 7, 1997, <http://www.un.org/icty/erdemovic/appeal/judgment/erd-aj971007e.htm>.

and at the same time gratitude for tenacious human rationality and morality as well. Even as science and technology keep renewing, and new methods and means of warfare continue to emerge, the legacy of sophisticated diction employed in all-encompassing legislation left by antecedents—as exemplified by the Martens Clause—will continue to offer forceful support by a minimum morality and legality in human efforts to constrain the destructive scope of warfare.

NOTICE:
This material may be protected by copyright law (Title 17 U.S.
Code)

CONFLICT IN THE
21ST CENTURY:
THE RISE OF HYBRID WARS



Frank G. Hoffman
Potomac Institute for Policy Studies
Arlington, Virginia
December 2007

ABOUT
POTOMAC INSTITUTE FOR POLICY STUDIES
AND THE
CENTER FOR EMERGING
THREATS AND OPPORTUNITIES

The Potomac Institute for Policy Studies is an independent, not-for-profit public policy research center. The Institute identifies key science, technology and national security issues, and aggressively follows through with focused research and policy advice. From this research and subsequent public discussions, the Potomac Institute has a track record for developing meaningful policy options and assisting their implementation at the intersection of both business and government.

In partnership with the U.S. Marine Corps, the Potomac Institute established the Center for Emerging Threats and Opportunities (CETO) in 2001. The generous sponsorship of the Senate Armed Services Committee was instrumental in this initiative. In short order, CETO assisted the Marine Corps in identifying a number of potential operational challenges as well as a number of potential “game changer” technology solutions. In the last few years, CETO, as a part of the Marine Corps Warfighting Lab (Quantico, Virginia), has been on the front edge of advanced concept design and experimentation in distributed operations, urban reconnaissance and surveillance, counter-sniper and anti-improvised explosive device (IED) tactics and technologies.

© Copyright 2007 Potomac Institute for Policy Studies

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from Potomac Institute for Policy Studies.

Published 2007 by the Potomac Institute for Policy Studies
901 N. Stuart Street, Suite 200, Arlington, VA 22203
Potomac Institute URL: <http://www.potomac institute.org/>
To order the report or to obtain additional information, contact:
Telephone: 703.525.0770 Fax: 703.525.0299

Cover design by: Patrick Worcester

TABLE OF CONTENTS

Table of Contents.....	iii
Foreword	5
Executive Summary	7
Introduction	11
Origins and Development of Hybrid Warfare.....	17
Hezbollah As Prototype	35
Implications.....	43
Conclusion.....	55
About the Author.....	61
Acronyms.....	63
Bibliography	65

[INTENTIONALLY LEFT BLANK]

FOREWORD

There has been much discussion and writing over the past decade about the evolution of modern warfare in the post Cold-war world. Several have claimed that we were in the midst of a “Revolution in Warfare.” Frank Hoffman takes this discussion to a new and much more mature level by recognizing that we are entering a time when multiple types of warfare will be used simultaneously by flexible and sophisticated adversaries who understand that successful conflict takes on a variety of forms that are designed to fit one’s goals at that particular time. Mr. Hoffman calls these “Hybrid Wars.”

Frank Hoffman notes that it is too simplistic to merely classify conflict as “Big and Conventional” versus “Small or Irregular.” Today’s enemies, and tomorrow’s, will employ combinations of warfare types. Non-state actors may mostly employ irregular forms of warfare, but will clearly support, encourage, and participate in conventional conflict if it serves their ends. Similarly, nation-states may well engage in irregular conflict in addition to conventional types of warfare to achieve their goals. Clearly the United States must be prepared for the full spectrum of conflict from all fronts and realize that preparing our forces for only selected types of conflict will be a recipe for defeat.

It is important to note that this work is being evaluated by the U.S. Marine Corps (USMC) in its ongoing long-range strategic planning and experimentation activities. It has been presented to senior officials at the Office of the Secretary of Defense (OSD) policy level, to policy leaders in the Department of Homeland Security (DHS), to the Intelligence Community, to the Defense Advanced Research Projects Agency (DARPA) and the Office of Naval Research (ONR), and to major military educational institutions including the U.S. National War College, the Naval and Army War Colleges, and the British Joint Command and Staff College.

Frank Hoffman’s paper on Hybrid Wars is a masterpiece of enlightened thinking on conflict in our time. It should be required reading for all students and practitioners of modern warfare.



*Michael S. Swetnam
CEO & Chairman*

Potomac Institute for Policy Studies

[INTENTIONALLY LEFT BLANK]

EXECUTIVE SUMMARY

There are a broadening number of challenges facing the United States, as the National Defense Strategy (NDS) noted in 2005. These include traditional, irregular, terrorist and disruptive threats or challengers. This has created a unique planning dilemma for today's military planners, raising a choice between preparing for states with conventional capabilities or the more likely scenario of non-state actors employing asymmetric or irregular tactics. However, these may no longer be separate threats or modes of war. Several strategists have identified an increased merging or blurring of conflict and war forms. The potential for types of conflict that blur the distinction between war and peace, and combatants and non-combatants, appear to be on the rise. Indeed, the NDS itself suggested that the most complex challengers of the future may seek synergies and greater impact by combining multiple modes of war.

As this paper reveals, future contingencies will more likely present unique combinational or *hybrid* threats that are specifically designed to target U.S. vulnerabilities. Instead of separate challengers with fundamentally different approaches (conventional, irregular or terrorist), we can expect to face competitors who will employ *all* forms of war and tactics, perhaps simultaneously. Criminal activity may also be considered part of this problem as well, as it either further destabilizes local government or abets the insurgent or irregular warrior by providing resources, or by undermining the host state and its legitimacy.

It is not just that conventional warfare or interstate conflict is on the decline, there is a fusion of war forms emerging, one that blurs regular and irregular warfare. This emerging understanding is reflected in the recently released national maritime strategy.

Conflicts are increasingly characterized by a hybrid blend of traditional and irregular tactics, decentralized planning and execution, and non-state actors," the strategy states, "using both simple and sophisticated technologies in innovative ways.¹

¹ General James T. Conway, USMC, Admiral Gary Roughead, USN and Admiral Thad W. Allen, USCG, *A Cooperative Strategy For Maritime Security*, Washington, D.C., October 2007.

Hybrid threats incorporate a full range of different modes of warfare including conventional capabilities, irregular tactics and formations, terrorist acts including indiscriminate violence and coercion, and criminal disorder. Hybrid Wars can be conducted by both states and a variety of non-state actors. These multi-modal activities can be conducted by separate units, or even by the same unit, but are generally operationally and tactically directed and coordinated within the main battlespace to achieve synergistic effects in the physical and psychological dimensions of conflict. The effects can be gained at all levels of war.

At the strategic level, many wars have had regular and irregular components. However, in most conflicts, these components occurred in different theaters or in distinctly different formations. In Hybrid Wars, these forces become blurred into the same force in the same battlespace. While they are operationally integrated and tactically fused, the irregular component of the force attempts to become operationally decisive rather than just protract the conflict, provoke overreactions or extend the costs of security for the defender.

We may find it increasingly perplexing to characterize states as essentially traditional forces, or non-state actors as inherently irregular. Future challenges will present a more complex array of alternative structures and strategies, as seen in the summer of 2006 in the battle between Israel and Hezbollah. Hezbollah clearly demonstrated the ability of non-state actors to study and deconstruct the vulnerabilities of Western style militaries, and devise appropriate countermeasures. The lessons learned from this confrontation are already cross-pollinating with other states and non-state actors. With or without state sponsorship, the lethality and capability of organized groups is increasing, while the incentives for states to exploit nontraditional modes of war are on the rise. This will require that we modify our mindsets with respect to the relative frequency and threats of future conflict. It will also require a rethinking of priorities in defense spending, and serious reflection about the role of technology in our strategic culture.

The National Defense Strategy and the 2006 Quadrennial Defense Review (QDR) quite properly recognized that future challengers will avoid our overwhelming military strengths and seek alternative paths. OSD's senior civilian policy makers sought to shift the Department's capability investments to meet these challengers. The Pentagon's strategy and QDR expands the U.S. military's mission set beyond its

preference for fighting conventional forces. We can no longer focus just on *battles* against preferred enemies, vice *campaigns* against thinking opponents.

Hybrid Warfare presents a mode of conflict that severely challenges America's conventional military thinking.² It targets the strategic cultural weaknesses of the American Way of Battle quite effectively. Its chief characteristics—convergence and combinations—occur in several modes. The convergence of various types of conflict will present us with a complex puzzle until the necessary adaptation occurs intellectually and institutionally. This form of conflict challenges longstanding American conceptions about warfighting, and will continue to thwart the West's core interests and world order over the next generation.

The rise of Hybrid Warfare *does not* represent the end of traditional or conventional warfare. But it does present a complicating factor for defense planning in the 21st Century. The implications could be significant, but will have to be carefully thought through. The historical foundation for much of our understanding about war requires fresh and creative approaches if we are going to draw out the correct implications. As Dr. John Arquilla of the Naval Postgraduate School has noted, "While history provides some useful examples to stimulate strategic thought about such problems, coping with networks that can fight in so many different ways—*sparkling myriad, hybrid forms of conflict*—is going to require some innovative thinking."³

This paper lays out some distinct areas in which innovative thinking, rigorous experimentation, and constant adaptation are required. These include changes in our approach to operational art, command and control, leadership development, force structure, and training and education.

We believe that the Marine Corps is particularly well suited for this security environment because of its legacy, its expeditionary culture and its approach to warfighting. The Marine Corps has proven to be an innovative organization, and its fundamental warfighting doctrine and its core competencies provide it with the foundation to effectively counter,

² Credit for the first use of the term can be given to Robert G. Walker, "Spec Fi: The U.S. Marine Corps and Special Operations," unpublished Master's Thesis, Monterey, CA; Naval Post Graduate School, December 1998. Walker described the Marine Expeditionary Unit as "a hybrid force for Hybrid Wars."

³ John Arquilla, "The end of war as we knew it: Insurgency, counterinsurgency and lessons from the forgotten history of early terror networks," *Third World Quarterly*, March 2007, p. 369.

if not thrive, against hybrid challengers. The Marine Corps should exploit its well-founded legacy of warfighting excellence, expeditionary ethos, and institutional agility for this new era.

Because of their perceived success, hybrid challengers will not be a passing fad nor will they remain focused on low tech applications. Future opponents will be dedicated, learn rapidly and adapt quickly to more efficient modes of killing. The ongoing Long War underscores their capacity for incorporating new tactics, techniques and procedures. This diffusion will continue. We can no longer overlook our own vulnerabilities or underestimate the imaginations of our antagonists. In a world of Hybrid Wars, the price for complacency grows steep. This monograph seeks to accelerate our own learning and corresponding institutional adaptation.

INTRODUCTION

The state on state conflicts of the 20th century are being replaced by Hybrid Wars and asymmetric contests in which there is no clear-cut distinction between soldiers and civilians and between organised violence, terror, crime and war.⁴

The tragic events of September 11, 2001 did not change everything, but they did emphatically punctuate the end of one era of war and awaken us to the dawning of a new one. This new era presents policy makers and military planners with its own method of conflict, one that has made conventional thinkers uncomfortable. This kind of war, as Mao suggested long ago, has several constituent components, and overwhelming military power by itself is insufficient to serve our strategic interests. Regardless of unfounded speculation in some corners, this does not eliminate the utility of the timeless Clausewitz or some 15 centuries of recorded military history before Westphalia. Quite the contrary, the Prussian theorist recognized that every age has its own conception of war. While globalization has made war more dangerous, it remains undeniably consistent with Clausewitz's broad theory.⁵ Today's emerging paradigm is reflected by the likes of Osama Bin Laden and our experiences in both Afghanistan and Iraq. These experiences should give pause to strategy makers. Complacency about today's security challenges is unwarranted and highly dangerous.

The so called "unipolar moment" and a spate of unilateral triumphalism went up in smoke on 9/11. Wishful thinking and delusional discussions about the changing nature of human conflict were the principal victims of 9/11, reinforced by subsequent events in Iraq. Rather than Fukuyama's "End of History," our security is challenged by

⁴ Alan Dupont, "Transformation or Stagnation? Rethinking Australia's Defence," *Australian Security in the 21st Century Lecture*, Parliament House, Canberra, November 13, 2002. Accessed at www.mrcltd.org.au/uploaded_documents/ACF30D.doc on November 18, 2007.

⁵ Antulio J. Echevarria, II, "Globalization and the Nature of War," Carlisle, PA: Strategic Studies Institute, Army War College, March 2003. See also Antulio J. Echevarria, II, *Clausewitz and Contemporary War*, New York, NY: Oxford University Press, 2007.

12 | Conflict in the 21st Century

a violent reaction generated as a side product of globalization. This reaction is abetted or exploited by the fervently fanatic and faith-based factions within the Middle East.

The future portends an even more lethal strain of perturbation. Other analysts like Dr. Bruce Hoffman point out that Iraq's insurgents and jihadist foreign fighters will benefit from their education in Iraq and Afghanistan, and will soon return home or to alternative battlespaces with greater motivation, lethal skills and credibility.⁶ Their Darwinian evolution against America's vaunted military has refined their methods and emboldened their plans, while the clash within Islam continues unabated if not accelerated by America's well intentioned, but poorly executed efforts. So our danger mounts, while the West remains unprepared to provide security against a stateless entity that deliberately targets its weaknesses and refuses to play to its conventional military strength. Others, including large states with interests inimical to our own, will learn from this experience.

Western military thinkers have been reluctant and thus slow to address the implications of the increasingly blurred character of modern wars. Many are inclined to look past the uncomfortable and ambiguous nature of today's generational challenge, and long for traditional opponents who will array themselves in properly uniformed formations and fight the wars we prefer to fight. We have been slow to accept the trend lines that go back as far as Beirut in 1983 and recognize that the most frequent form of war is now "amongst the people," and we have been very slow in shaping our institutional tool set.⁷

For more than two decades, most of us overlooked these trends. Only a few talked of post-Westphalian, non-trinitarian or post-Clausewitzian eras.⁸ The American military oriented on its preferred view of its professional scope, at the operational level, and worked to embrace the Information Age. However, much of that effort tried to perfect an

⁶ Statement of Dr. Bruce Hoffman, testimony presented to the HASC Subcommittee on Terrorism, Unconventional Threats and Capabilities on February 16, 2006. Accessed at www.rand.org/pubs/testimonies/CT255/.

⁷ Rupert Smith, *The Utility of Force: The Art of War in the Modern World*, New York, NY: Knopf, 2007.

⁸ The most notable being Martin Van Creveld, *The Transformation of War*, New York, NY: Free Press, 1991.

increasingly infrequent if not outmoded form of war, and made Industrial Age warfare more precise, more predictable and more pristine. We maximized efficiency and the application of ever more modern forms of technology. But we had focused on the wrong set of strategic drivers and indicators. Visions of “unblinking eyes” and information superiority, stand-off attacks and ever faster sensor-to-shooter links drove the defense agenda and the transformation programs of the Department of Defense for more than a decade.⁹

While all of these technologies were beneficial, they were not properly assessed in relation to the ongoing social and political context in which they were to be applied. In effect, we had misidentified the true Revolution in Military Affairs, as Sir Lawrence Freedman has noted. We could not eliminate the “fog of war” with America’s information dominance and magically create a new, long *Pax Americana*. We overlooked what really constituted a threat to our national security interests in key regions of the world, due to an enthusiastic embrace of an idealized and outdated version of warfare, and an under-appreciation of the mobilizing impact of Information Age tools when used to foment disorder and promote hate.

The latest U.S. National Defense Strategy (NDS), published in early 2005, reflects some improved thinking. This white paper explicitly identifies a range of emerging threats and identifies irregular challengers as an increasingly salient problem. It begins to shift American thinking and investments away from merely “fighting and winning the nation’s wars” against its preferred conventional enemies to a range of wars against a wider set of enemies—expanding beyond the traditional to incorporate three other threats or challengers—the irregular, the catastrophic terrorist and the disruptive threat, which seeks to usurp American hegemony and power by displacing it via breakthrough technologies.

The authors of the Pentagon’s 2006 Quadrennial Defense Review (QDR) can also take a bow. The report recognized the shift, concluding, “In the post-September 11 world, irregular warfare has emerged as the dominant form of warfare confronting the United

⁹ See Frederick W. Kagan, *Finding the Target: The Transformation of American Military Policy*, New York, NY: Encounter Books, 2006.

States.”¹⁰ The QDR argued that fighting the Long War against terrorism, providing forces for rotational requirements to protracted conflicts, and increasing the capacity of future partners would be the basis for sizing and shaping the military. This policy shift is significant as it begins to broaden the scope of the combat developers inside each Service to sharpen our focus on this increasingly likely form of warfare.

This paper suggests that we still have a ways to go. Instead of the four distinct challengers presented in a two by two matrix chart (known as the Quad chart in the Pentagon) found in the new NDS, future scenarios will more likely present unique combinations and deliberate synergies that are specifically designed to target Western societies in general and American vulnerabilities in particular. The defense strategy created the impression that our portfolio of capabilities would be measured against four distinct kinds of challengers using different approaches. Our take on the future suggests that future adversaries are smarter than that and will rarely limit themselves to a single tool in their tool kit. Conventional, irregular and catastrophic terrorist challenges will not be distinct styles; they will all be present in some form. The blurring of modes of war, the blurring of who fights, and what technologies are brought to bear, produces a wide range of variety and complexity that we call *Hybrid Warfare*.¹¹

Hybrid Wars can be conducted by both states and a variety of non-state actors. *Hybrid Wars incorporate a range of different modes of warfare, including conventional capabilities, irregular tactics and formations, terrorist acts including indiscriminate violence and coercion, and criminal disorder.* These multi-modal activities can be conducted by separate units, or even by the same unit, but are generally operationally and tactically directed and coordinated within the main battlespace to achieve synergistic effects.

¹⁰ Report of the Quadrennial Defense Review, Washington D.C., March 2006, p. 36.

¹¹ In addition to its first public use by General Mattis at the Defense Forum sponsored by the Naval Institute and Marine Corps Association on September 8, 2005, the concept has been presented by LtGen James N. Mattis USMC and Frank Hoffman, “Future Warfare: The Rise of Hybrid Warfare,” *Naval Institute Proceedings*, November 2005, pp. 30-32; F. G. Hoffman. “Complex Irregular War: The Next Revolution in Military Affairs,” *Orbis*, Summer 2006, pp. 413-430; F. G. Hoffman, “How the Marines are Preparing for Hybrid Wars,” *Armed Forces Journal International*, April 2006; and F. G. Hoffman “Preparing for Hybrid Wars,” *Marine Corps Gazette*, March 2007.

This paper captures the progress and insights of a long-term research project undertaken by the Marine Corps that has already yielded innovative approaches to this emerging challenge. In addition to reviewing the literature and analyses that brought us up to this point in time, the paper provides implications for the defense community at large to consider.

Complicating the problem, the battlespace in tomorrow's Hybrid Wars will take place in complex terrain, most likely the burgeoning cities of the developing world. The hybrid challenger realizes that complex terrain affords defenders a number of advantages that offset our conventional superiority. Recent combat operations suggest a shift towards what can be called *contested zones*.¹² These zones include the dense urban jungles and the congested littorals where the majority of the world's population and economic activity is centered.¹³ Engaging American forces in the "contested zone" with a range of crude yet effective asymmetric approaches is intended to draw out conflicts, protract their duration and costs, and sap American will. This will come as no news to the veterans of Operation Al Fajr in Fallujah.

As seen in Kosovo, Afghanistan, and Iraq, irregular adversaries are adopting tactics and modes of operations to offset our firepower and advantages in intelligence collection, surveillance and reconnaissance. Today, dense urban terrain provides similar safe-havens to the urban guerrilla or terrorist where the density of population, transportation networks, public services and infrastructure, and structures gives him multiple avenues of escape and the ability to hide while planning and rehearsing operations. The density of the urban complex provides sufficient cover and "noise" to mask the adversary's preparation and attack position. We have to take urban encounters seriously, "categorizing urban operations as too difficult and costly must come to an end" as it could become a self-fulfilling prophecy.¹⁴

Clearly future opponents will avoid fighting the American Way of War, where we optimize our Industrial Age mass or Information Age

¹² Robert E. Schmidle and F. G. Hoffman, "Commanding the Contested Zones," *Proceedings*, September 2004.

¹³ Ralph Peters, "Our Soldiers, Their Cities," *Parameters*, Spring 1996, pp. 43-50.

¹⁴ Vincent J. Goulding, Jr., "Back to the Future with Asymmetric Warfare," *Parameters*, Winter 2000-01, pp. 21-30.

prominence and our preferred rule sets of war. The likeliest opponents on future battlefields accept no rules. Their principal approach will be to avoid predictability and seek advantage in unexpected ways and ruthless modes of attack. We can expect to see a lot of tactical plagiarism, with our opponent learning from us, coupled with wild cards or hybrid adaptation where our adversary has learned how to use high technology in unique and unanticipated ways.

We will also face primitive forms of warfare and criminal activity that long ago were proscribed by Western society. Future enemies will seek their own degree of “shock and awe” with crude barbarity (with video) rather than precision weaponry. What we ironically call “irregular” warfare will become increasingly familiar, but with greater velocity and greater lethality than in the past in part due to the diffusion of advanced military technology. In this paper, we have identified the potentially most dangerous and the increasingly most likely form of conflict as Hybrid Warfare. This form of conflict will challenge longstanding American conceptions about war, and its conventionally oriented force structure and investment patterns. This blurring character of conflict will continue to test and thwart the West’s security interests and world order over the next generation.

ORIGINS AND DEVELOPMENT OF HYBRID WARFARE

*In warfare and non-military warfare, which is primarily national and supra-national, there is no territory which cannot be surpassed; there is no means which cannot be used in the war; and there is no territory and method which cannot be used in combination.*¹⁵

The ongoing conflicts in Iraq and Afghanistan reinforce General Krulak's famous forecast about future conflicts. He predicted that future conflicts would be unlike the large-scale mechanized sweeps of Operation Desert Storm, but more like the "Stepchild of Chechnya."¹⁶ The Chechens employed swarming tactics inside their own cities to thwart Russian domination.¹⁷ That model did not resonate with the mainstream national security community at the time, nor did the evidence from a number of other experiences in Beirut, Mogadishu, or Sarajevo. But now the Long War and the Pentagon's belated interest in stability operations, irregular wars, and counterinsurgency have accelerated a debate about the future character of conflict and the allocation of resources within the U.S. national security architecture.

As part of the Center for Emerging Threats and Opportunities' (CETO) ongoing "Changing Character of Conflict" research program, we examined a number of projected models and postulated paradigm changes regarding future conflict. Proposals for "non-trinitarian" wars, 4th Generation Warfare, and Compound Wars were prominent in the literature at the time this project commenced. Others have described current conflicts as "New Wars," noting supposedly unique characteristics, in particular extensive refugee flows, sexual violence, and transnational criminal aspects underpinning recent wars.¹⁸ This section

¹⁵ Qiao Liang and Wang Xiangsui, *Unrestricted Warfare*, Beijing: PLA Literature and Arts Publishing House, 1999.

¹⁶ "[O]ur enemies will not allow us to fight the son of Desert Storm, but they will try to draw us into the stepchild of Chechnya." Robert Holzer, "Krulak Warns of Over-Reliance on Technology," *Defense News*, October 7, 1996, p. 4.

¹⁷ Timothy L. Thomas, "The Battle of Grozny: Deadly Classroom in Urban Combat," *Parameters*, Summer 1999, pp. 87-102; Anatol Lieven, *Chechnya, Tombstone of Russian Power*, New Haven, CT: Yale University Press, 1998, pp. 102-140.

¹⁸ The term is usually attributed to Mary Kaldor, *New and Old Wars: Organized Violence in a Global Era*, Cambridge, UK: Polity, 1999; and more recently, Herfried Munkler, *The New Wars*, Malden, MA: Polity Press, 2005, pp. 5-31.

briefly describes a number of theories examined as part of this project and the underlying rationale for developing and refining CETO's own Hybrid Warfare construct.

Fourth Generation Warfare

We examined the concept of Fourth Generation Warfare (4GW) that appears to be very prophetic.¹⁹ The theory poses significant historical shifts in warfare since 1648. Initially it was dismissed as “elegant irrelevance,” but it is now difficult to dismiss the reality of 4GW.²⁰ Proponents of this theory accurately identified the blurring nature of future conflict, especially the blurring of war and peace, as well as that between combatants and noncombatants. The core of the concept is that the weakening of the state as an organizing and governing mechanism results in the rise of non-state actors willing and able to challenge the legitimacy of the state. The role of political will and internal social disintegration is central to the construct. The 4GW actor uses a range of conventional and unconventional means, including terrorism and information, to undermine the will of the existing state, to de-legitimize it, and to stimulate an internal social breakdown.²¹ The theory is sophisticated but also elusive. Its advocates have been accused of ignoring the history of irregular warfare, a record that is about as long as military history itself.

The theory has numerous advocates, who place emphasis on political will, legitimacy, and culture. Their 4GW adversary exploits societies, adopts an amorphous structure, and utilizes mass mobilization techniques. The novelty of the concept has been challenged. The 4GW notion raised by T. X. Hammes that “superior political will when properly employed can defeat greater economic and military power” was not mysterious to George Washington or to the Continental Congress. Neither was protracted conflict, social and political networks, diasporas, and ideological fervor lost on Michael Collins and the Irish rebels in the period after World War I.

¹⁹ William S. Lind, Keith Nightengale, John Schmitt and Gary I. Wilson, “The Changing Face of War: Into the Fourth Generation,” *Marine Corps Gazette*, November, 2001.

²⁰ Kenneth F. McKenzie, “Elegant Irrelevance: Fourth Generation Warfare,” *Parameters*, Autumn, 1993, pp. 51-60.

²¹ On this point see William S. Lind, “The will doesn’t triumph,” in Terriff, Karp, and Karp, eds., *Global Insurgency and the Future of Armed Conflict*, New York, NY: Routledge Press, 2007, pp. 101-104.

Some prominent historians have been critical of this approach. Dr. Antulio Echevarria finds that “the model of 4GW ... is based on poor history and only obscures what other historians, theorists, and analysts already have worked long and hard to clarify.” His assessment is echoed by others. Professor Lawrence Freedman of King’s College London has noted “...the theory of 4GW suffers from poor use of history and lack of intellectual rigor.”²² We noted the objections to the historical framework, and concluded that the fourth generation framework hides more than it reveals. The Roman era, the Crusades, Europe’s reformation period, or Britain’s imperial history, which is the basis for Callwell’s treatise *Small Wars*, all contain elements of what is now considered fourth generation warfare. Likewise, historians have to wonder why the Philippine insurgency at the turn of the century, the Marine’s Small Wars era, and T.E. Lawrence’s campaign that occurred during the purported Second Generation, are not relevant as case studies.²³

Thus, not surprisingly, British, Israeli and Australian analysts take issue with the concept. Professor Ron Thornton believes the concept’s formulation risks excluding a huge corpus of valuable knowledge. Avi Kober finds it “incoherent and eclectic,” and Mike Evans contends that instead of distinct historical generations and the emergence of something entirely new, we need a rigorous evaluation of the apparent merging of existing forms of conflict.²⁴ That said, the debate stirred up by the theory has been useful as it forced the profession to examine itself, today’s prevailing conventional orthodoxy, and the tendency in the United States to ignore irregular warfare.

Whether this really is something entirely new, “visible and distinctly different from the forms of war that preceded it,” has emerged as challengeable. What has occurred is simply part of war’s evolution, a shift in degree rather than kind, and a return to older and horrific cases. 4GW advocates do not deny the existence of irregular warfare

²² Antulio J. Echevarria II, “Deconstructing the Theory of Fourth-Generation War,” *Contemporary Security Policy*, August 2005, pp. 11-20; Lawrence Freedman, “War Evolves into the Fourth Generation,” *Contemporary Security Policy*, August 2005, pp. 1-10.

²³ For further assessment, see the concluding chapter in Terry Terriff, Aaron Karp and Regina Karp, eds., *Global Insurgency and the Future of Armed Conflict: Debating fourth-generation warfare*, *Op. Cit.*

²⁴ See Rod Thornton, “Fourth Generation: A ‘new form of warfare?’” p. 87; Avi Kober, “The end of Israeli omnipotence?” pp. 147-159; and Mike Evans “Elegant irrelevance revisited: A critique of Fourth Generation Warfare,” p. 72 in Terriff, Karp and Karp, *Global Insurgency*, *Op Cit.*

techniques and the return to medieval warfare. But they do tend to overlook Clausewitz, who noted that war is “more than a chameleon,” with continuous adaptation in character in every age. Very little in what is described as fundamentally different in the 4GW literature is all that inconsistent with a Clausewitzian understanding of war as a contest of human wills.²⁵ The emphasis on impacting one’s political cohesion or will was a fundamental aspect of Clausewitz’s canon, but the idea of achieving this indirectly rather than via the fielded military forces of the opponent has merit, as does the increasingly blurring character of conflict. While it lacks prescriptions, the 4GW school is certainly relevant.

Compound Wars

Historians have noted that many if not most wars are characterized by both regular and irregular operations. When a significant degree of strategic coordination between separate regular and irregular forces in conflicts occurs they can be considered “compound wars.” Compound wars are those major wars that had significant regular and irregular components fighting simultaneously under unified direction.²⁶ The complementary effects of compound warfare are generated by its ability to exploit the advantages of each kind of force, and by its ability to increase the nature of the threat posed by each kind of force. The irregular force attacks weak areas, and forces a conventional opponent to disperse his security forces. The conventional force generally induces the adversary to concentrate for defense or to achieve critical mass for decisive offensive operations.

One can see this in the American Revolution where Washington’s more conventional force stood as a “force in being” for much of the war, while the South Carolina campaign was characterized by militia and some irregular combat.²⁷ The Napoleonic era is frequently viewed in terms of its massive armies marching back and forth across Europe. But the French invasion of Spain turned into a quagmire with British

²⁵ Carl von Clausewitz, *On War*, Michael Howard and Peter Paret, eds. and trans., Princeton, NJ: Princeton University Press, 1976, p. 77. See also Christopher Daase, “Clausewitz and Small Wars,” in Hew Strachan and Andreas Herberg-Rothe, eds., *Clausewitz in the Twenty-First Century*, New York, NY: Oxford University Press, 2007.

²⁶ Thomas Huber, *Compound Wars: The Fatal Knot*, Fort Leavenworth, KS: Command and General Staff College, 1996.

²⁷ John Grenier, *First Way of War: American War Making on the Frontier*, Cambridge University Press, 2005; Terry Golway, *Washington’s General: Nathanael Green and the Triumph of the American Revolution*, New York, NY: Henry Holt, 2005.

regulars contesting Napoleon's control of the major cities, while the Spanish guerrillas successfully harassed his lines of communication. Here again, strategic coordination was achieved, but overall in different battlespaces.²⁸ Likewise, the American Civil War is framed by famous battles at Chancellorsville, Gettysburg, Vicksburg, and Antietam. Yet, partisan warfare and famous units like Mosby's 43rd Virginia cavalry provided less conventional capabilities as an economy of force operation.²⁹ T. E. Lawrence's role as an advisor to the Arab Revolt against the Ottomans is another classic case of Compound War that materially assisted General Allenby's thrusts with the British Expeditionary Force against Jerusalem and Damascus. But here again, Lawrence's raiders did not fight alongside the British, they were strategically directed by the British and supplied with advisors, arms and gold only.³⁰

Vietnam is another classic case of the strategic synergy created by compound wars, juxtaposing the irregular tactics of the Viet Cong with the more conventional capabilities of the North Vietnamese Army.³¹ The ambiguity between conventional and unconventional approaches vexed military planners for several years. Even years afterwards, Americans debated what kind of war they actually fought and lost.³²

Upon detailed examination of the case studies presented, this theory did not hold up to its own definition in that we could identify only cases of strategic coordination rather than Huber's claim that forces fought alongside each other. When militia and irregular forces were ever employed with regular forces, as at Cowpens under Morgan's direction, the irregular forces were not employed as such but merely as second-rate conventional forces. This theory offered synergy and combinations at the strategic level, but not the complexity, fusion and simultaneity we foresaw at the operational and even tactical level. Irregular forces in these cases operated largely as a distraction or economy of force measure, in a separate theater or adjacent operating area, including the rear echelon. Because it is based on operationally separate forces, the

²⁸ Charles J. Esdaile, *Fighting Napoleon, Guerrillas, Bandits and Adventurers in Spain 1808-1814*, New Haven, CT: Yale University Press, 2004, pp. 154-155.

²⁹ Jeffrey D. Wert, *Mosby's Rangers*, New York, NY: Simon & Schuster, 1991.

³⁰ B. H. Liddell Hart, *Lawrence of Arabia*, New York, NY: Da Capo Press, 1989.

³¹ Harold G. Moore and Joseph L. Galloway, *We Were Soldiers Once...and Young: Ia Drang—The Battle That Changed the War in Vietnam*, New York, NY: Random House, 1992.

³² The best source on the war is Andrew F. Krepinevich, *The Army in Vietnam*, Baltimore, MD: Johns Hopkins University Press, 1986.

compound concept did not capture the merger or blurring modes of war we had identified in recent case studies or our projections.

Unrestricted Warfare

A pair of Chinese Colonels are notorious for their conception of Unrestricted Warfare—or “war beyond limits.”³³ This pair of Chinese political officers caused quite a stir by suggesting an immoral and potentially violent mutation in human conflict, one that was beyond the pale of most Western military scholars or practitioners. But a closer reading of their text reveals a lot of useful and even obvious conclusions. Well ahead of their time, the authors recognized the potential implications of globalization. Their conception of unrestricted warfare is really best translated as war “beyond limits,” and this translation serves to expand not just the forms that warfare takes, but the boundaries of the domains or dimensions of warfare that most Western military officers might hold.

The two Colonels did not suggest that war was without moral restraints or beyond any limits at all. They sought to expand the definition and understanding of *war beyond just its traditional military domain*. Like many insightful security analysts in Europe and the United States, Colonels Qiao and Wang also understood the strains that the conventional nation-state was under due to globalization. In their words:

The great fusion of technologies is impelling the domains of politics, economics, the military, culture, diplomacy, and religion to overlap each other. The connection points are ready, and the trend towards the merging of the various domains is very clear. All of these things are rendering more and more obsolete the idea of confining warfare to the military domain and of using the number of casualties as a means of the intensity of a war.³⁴

Their concept, which they overstated as “a completely new method of warfare” was titled “modified combined war that goes beyond limits” [“*pian zhen shi chao xian zube zhan*”]. This concept exploits the benefits of “combinations” in types of organizations and among the various domains of national power. While in the past, the Great Captains were masters of combinations, these were all achieved within the military

³³ Liang, *Unrestricted Warfare*.

³⁴ *Ibid.*, p. 162.

domain. In *Unrestricted Warfare*, future Great Captains must master the ability to “combine” all of the resources of war at their disposal and use them as means to prosecute the war. These resources must include information warfare, financial warfare, trade warfare, and other entirely new forms of war. These distinctions do not appear startling today, given our current efforts to harness all instruments of national power in Iraq. But the idea that “warfare is no longer an activity confined only to the military sphere,” remains outside the orthodoxy of Western military thinking even now.³⁵

In terms of the “beyond-limits” thinking, the authors noted that the United States is already effective at “supra-domain combinations,” or the combining of battlefields, or what American analysts call instruments of power. Their text highlights the U.S. mastery of supra-domain combinations against Iraq during Desert Storm by establishing a large coalition, by political actions at the United Nations (UN), by our sweeping military campaign, and the subsequent decade of pressure by the U.S. military coupled with economic sanctions.

Essential Principles

The authors generated a list of new principles appropriate to “beyond-limits combined war.” These include Omni-directionality, Synchrony, and Asymmetry.³⁶ These are defined below:

Omni-directionality – requires that commanders observe a potential battlefield without mental preconditions or blind spots. The designing of plans, employment measures, and combinations must make use of all war resources which can be mobilized. The commander is enjoined to make no distinction between what is or is not the battlefield. All the traditional domains, (ground, seas, air, and outer space) as well as politics, economics, culture, and moral factors are to be considered battlefields.

Synchrony – enjoins on commanders to link the disaggregated nature of multiple battlefields in different domains with consideration of the temporal dimension. In other words, “conducting actions in different spaces in the same period of time” to achieve desired effects. Instead of phases, with the accumulated results of multiple battles,

³⁵ *Ibid.*, p. 190.

³⁶ *Ibid.*, pp. 204-216.

strategic results can now be attained rapidly by simultaneous action or at designated times.

Asymmetry – here the authors recognized that asymmetry manifests itself to some extent in every aspect of warfare. However, asymmetry has been sought in operational terms within traditional military dimensions. In war beyond limits, the spectrum for overlooking the normal rules is much wider.

This concept poses an expansion of thinking about what constitutes war, one that challenges our conventional thinking. The authors rightly identified a number of implications of this concept, including the fact that Western military officers would have difficulty grasping the scope and nature of such a war. As other American authors have now noted, the U.S. focus on the operational level of war was essentially a preoccupation with battles, not the broader nature of war.³⁷ War “beyond limits,” on the other hand:

...will impose demands which will mean that most of the warriors will be inadequately prepared, or will feel as though they are in the dark: the war will be fought and won in a war beyond the battlefield; the struggle for victory will take place on a battlefield beyond the battlefield.³⁸

A pair of Air Force scholars reinforced our own understanding of this conception of future warfare with their discussion of “combinational warfare” based on their interpretation of Chinese strategic thinking.³⁹ Of course, supra-domain operations are also suggested by American strategists as well, including Yale Professor Paul Bracken, who has written about forms of economic warfare that exploit modern computer networks, banking systems and information operations.⁴⁰

Related to this discussion is the Chinese strategic concept of *shashou jiang* or *shashoujian* most often translated as “Assassin’s Mace.” This

³⁷ Antulio J. Echevarria, II, “Towards an American Way of War,” Carlisle, PA: Army War College, Strategic Studies Institute, March 2004.

³⁸ *Unrestricted Warfare*, p. 153.

³⁹ James Callard and Peter Faber, “An Emerging Synthesis for a New Way of War: Combinational Warfare and Future Innovation,” *Georgetown Journal of International Affairs*, Winter/Spring 2002, pp. 61-68.

⁴⁰ Paul Bracken, “Financial Warfare,” *E-Notes*, Philadelphia, PA: Foreign Policy Research Institute, September 2007.

concept relates to a perceived emphasis on advanced weapons and methods to attack identified vulnerabilities in Western military operations by Chinese military research and developmental efforts. U.S. government reports emphasize the development of these purportedly nefarious weapons and methods.⁴¹ Within the Chinese literature, these systems and modes are usually described as the People's Liberation Army's (PLA) trump card to Western military superiority and a necessary response for an inferior force to defeat a superior military power. A number of information warfare and missile development programs have been described as potential Assassin's Mace projects within the PLA. However, most of the PLA literature on this topic focuses on mimicking or offsetting the so-called American "Revolution in Military Affairs" and the exploitation of information technology.

The 2005 National Defense Strategy

Our development of a new warfighting construct to deal with emerging threats was significantly influenced by the publication of the inaugural National Defense Strategy in March 2005. The strategy broadens the scope of military planners to think past traditional threats that conveniently array themselves in open terrain for attack by America's preferred mode of precise stand-off warfare. The NDS acknowledged that America's military predominance influences the behavior of its enemies, and that our preeminence forces adversaries away from *traditional* forms of warfare. Instead these potential adversaries shift from opposing us conventionally and prepare more nontraditional or asymmetric capabilities and methods.

The NDS goes on to frame and define an array of *traditional*, *irregular*, *catastrophic*, and *disruptive* capabilities and methods that could threaten U.S. interests:

- **Traditional** challenges are posed by states employing recognized military capabilities and forces in well-understood forms of military competition and conflict.

⁴¹ See the Department of Defense Annual Report to Congress, Military Power of the People's Republic of China, 2006. Additionally, for a deep cultural exploration of this strategic concept see Jason E. Bruzdinski, "Demystifying *Shashoujian*: China's "Assassin's Mace" Concept," in Larry Wortzel and Andrew Scobell, eds., *Civil-Military Change in China*, Carlisle, PA: Army War College, 2004, pp. 309-364.

- **Irregular** challenges come from those employing “unconventional” methods to counter the *traditional* advantages of stronger opponents.
- **Catastrophic** challenges involve the acquisition, possession, and use of weapons of mass destruction (WMD) or methods producing WMD-like effects.
- **Disruptive** challenges may come from adversaries who develop and use breakthrough technologies to negate current U.S. advantages in key operational domains.⁴²

We found enormous value in the NDS and its commendable approach. The clearest benefits of this framework were highest at the strategic level. The wider set of challenges allowed OSD to examine its investment portfolio to preserve America’s current competitive advantage in conventional operations while also enhancing capabilities in Irregular war, deflecting acts against catastrophic terrorism, and avoiding strategic surprise in the Science and Technology arena.

But operationally, the NDS and the four challenges prospect did not satisfy our understanding of the kinds of threats the Marine Corps would face in the future, and it did not match our assessment of the ongoing merger or blurring of modes of war. The interpretation given to the NDS as the QDR was developed reinforced this assessment, as QDR implementation efforts continued to create very distinct and separate threats. This occurred despite the notion that the challenger categories could and would overlap, and the explicit statement in the NDS that:

[R]ecent experience indicates that the most dangerous circumstances arise when we face a complex of challenges. Finally, in the future, the most capable opponents may seek to combine truly disruptive capacity with *traditional, irregular, or catastrophic* forms of warfare.⁴³

Despite the useful framework set forth in the QDR, tomorrow’s conflicts will not be easily categorized into simple classifications of conventional or irregular wars. In fact, some of today’s best thinking acknowledges the blurring of lines between modes of war. Our greatest challenge will not come from a state that selects one approach, but from

⁴² Donald Rumsfeld, *The National Defense Strategy of the United States of America*, Washington, D.C., March 2005, pp. 2-4.

⁴³ *Ibid.*, p. 4.

states or groups that select from the whole menu of tactics and technologies to meet its own strategic culture and geography. As Dr. Mike Evans, now of the Australian Defence Academy, wrote well before the QDR:

The possibility of continuous sporadic armed conflict, its engagements blurred together in time and space, waged on several levels by a large array of national and sub-national forces means that war is likely to transcend neat divisions into distinct categories.⁴⁴

Many other analysts have captured these trends, with Russian, Australian, and American authors talking about “multi-modal” and “multi-variants” forms of war. But Dr. Evans was exceptionally insightful, noting:

...that British, French and Russian defense experts now speak of the rise of multi-variant warfare. They speak of a spectrum of conflict marked by unrestrained Mad Max wars in which symmetric and asymmetric wars merge and in which Microsoft coexists with machetes and stealth technology is met by suicide bombers.⁴⁵

Other Australian scholars pointed out the increasingly complex nature of the operating environment, particularly the presence of large numbers of civilians, dense urban environments and complex information activities. The Australian warfighting concept paper *Complex Warfighting* was material to our thinking at the time.⁴⁶ The Australian Army’s capstone concept was very forward looking, and captured the complexity of the terrain in future conflicts in terms of physical terrain, human terrain and informational terrain. This concept also captured the diffusion or blurring of conflict types, combatants/noncombatants, and war/peacetime. It also highlighted the

⁴⁴ Michael Evans, “From Kadesh to Kandahar: Military Theory and the Future of War,” *Naval War College Review*, Summer 2003, p. 136. In addition to Dr. Evans, Dr. Steve Blank, a research fellow at the Army’s Strategic Studies Institute, has followed these same trends. See Stephen Blank, “The War That Dare Not Speak its Name,” *Journal of International Security Affairs*, Spring 2005.

⁴⁵ Evans, “From Kadesh to Kandahar,” p. 140.

⁴⁶ Australian Army, *Complex Warfighting*, Future Land Warfare Branch, 2004. This paper was approved in 2005 as the Australian Army’s official Future Land Operational Concept. The principal author of this paper was LtCol David Kilcullen, Australian Army (now retired) who gave invaluable assistance.

implications of the “virtual theatres” of conflict that have developed from global communications technology.

American insights were also solicited, and the netwar concepts regarding new adversary organizational models were thoroughly reviewed and absorbed into our thinking.⁴⁷ Dr. Richard Harknett argued for an increasingly multidimensional character of war, but also greater lethality, based on “the combination of existing and new forms of organization with existing and new forms of destructive capability.”⁴⁸ Harknett shares Evans’ concerns for unique combinations, or as the latter put it, “a world of asymmetric and ethnopolitical warfare—in which machetes and Microsoft merge, and apocalyptic millenarians wearing Reeboks and Raybans dream of acquiring weapons of mass destruction.”⁴⁹

Hybrid Threats and Challengers

Thus, we have to conclude that the future does not portend a suite of distinct challengers with alternative or different methods but their convergence into multi-modal or Hybrid Wars. “Hybrid Wars” blend the lethality of state conflict with the fanatical and protracted fervor of irregular warfare. The term “Hybrid” captures both their organization and their means. Organizationally, they may have a hierarchical political structure, coupled with decentralized cells or networked tactical units. Their means will also be hybrid in form and application. In such conflicts, future adversaries (states, state-sponsored groups, or self-funded actors) will exploit access to modern military capabilities including encrypted command systems, man-portable air to surface missiles, and other modern lethal systems, as well as promote protracted insurgencies that employ ambushes, improvised explosive devices (IEDs), and coercive assassinations. This could include states blending high-tech capabilities, like anti-satellite weapons, with terrorism and cyber-warfare directed against financial targets. Conflicts will include hybrid organizations like Hezbollah and Hamas, employing a diverse set of capabilities. Additionally, states can shift their conventional units to irregular formations and adopt new tactics, as Iraq’s *Fedayeen* did in 2003.

In such conflicts we will face major states capable of supporting covert and indirect means of attack, as well as Thomas Friedman’s

⁴⁷ John Arquilla and David Ronfeldt, eds., *In Athena’s Camp: Preparing for Conflict in the Information Age*, Santa Monica, CA: RAND, 1997.

⁴⁸ Quoted in Blank, p. 32.

⁴⁹ Michael Evans, “From Kadesh to Kandahar,” p. 136.

“super-empowered” fanatics capable of highly lethal attacks undercutting the sinews of global order.⁵⁰ Cunning savagery, continuous improvisation and rampant organizational adaptation will mark this form of warfare. Such wars will not be conventional, low in intensity or short—and as General Rupert Smith notes in *The Utility of Force*, these conflicts can be timeless.⁵¹

These Hybrid Wars are polymorphous by their nature as are their antagonists.⁵² Hybrid Wars can be conducted by both states and a variety of non-state actors. *Hybrid Wars incorporate a range of different modes of warfare including conventional capabilities, irregular tactics and formations, terrorist acts including indiscriminate violence and coercion, and criminal disorder.* These multi-modal activities can be conducted by separate units, or even by the same unit but are generally operationally and tactically directed and coordinated within the main battlespace to achieve synergistic effects. The effects can be gained at all levels of war.

At the strategic level, many wars have had regular and irregular components. However, in most conflicts, these components occurred in different theaters or in distinctly different formations. In Hybrid Wars, these forces become blurred into the same force in the same battlespace. While they are operationally integrated and tactically fused, the irregular component of the force attempts to become operationally decisive rather than just protract the conflict, provoke overreactions or extend the costs of security for the defender.

Unlike in Maoist or compound wars, the purpose of the multi-modal approach is not to facilitate the progression of the opposition force through phases nor is it to help set up a conventional force for decisive battle. Hybrid opponents, in contrast, seek victory by the fusion of irregular tactics and the most lethal means available in order to attack and attain their political objectives. The disruptive component of Hybrid Wars does not come from high-end or revolutionary technology but from criminality. Criminal activity is used to sustain the hybrid force or to facilitate the disorder and disruption of the target nation. The goal may include protracted conflicts with a greatly diffused set of force

⁵⁰ “Super-empowered individuals” is from Thomas L. Friedman, *The Lexus and the Olive Tree: Understanding Globalization*, New York, NY: Anchor Books, 1999, pp. 14-15.

⁵¹ Rupert Smith, *The Utility of Force*, p. 8. General Smith noted “our conflicts tend to be timeless, even unending.”

⁵² Christopher Coker, “Cultural Ruthlessness and the War Against Terror,” *Australian Army Journal*, Vol. III, No. 1, 2006 p. 148.

capabilities to wear down resistance, or the actual defeat of a conventionally-oriented government.

This concept draws upon many schools of thought. From the 4GW school, it uses the concept of the blurring nature of conflict and the loss of the State's monopoly of violence. The concepts of omnidimensionality and combinations were crucial ideas adopted from Chinese analysts. From John Arquilla and T.X. Hammes we took in the power of networks. From the proponents of Compound Wars, the concept absorbs the synergistic benefit of mixing conventional and unconventional capabilities, but at lower and more integrated levels. From the Australian experts, we have accepted the growing complexity and disaggregated nature of the operational environment, as well as the opportunistic nature of future adversaries.

Since our initial research, we have gained new insights from a variety of sources, but none more critical than terrorism expert John Robb, whose Open Source Warfare concept points out the increasingly vulnerable nature of modern urban complexes.⁵³ Equally critical have been insights from Georgetown University professor Bruce Hoffman whose work on terrorism and insurgency has been highly influential over the past several decades, especially on the rising importance of urban insurgency. He has identified the growing tendency of insurgents today to adopt conventional means of greater and greater lethality as part of something he called "stand off insurgency." In such insurgencies, the insurgent exploits modern means including the IED, advanced munitions, unmanned aerial vehicles (UAVs) and even precision guided missiles to gain greater stand off against conventional counterinsurgent forces.⁵⁴ From both British and American Army theorists we have incorporated the role that criminal behavior, smuggling and narcotics play within today's conflict.⁵⁵

⁵³ John Robb espouses Open Source Warfare which aptly captures the entrepreneurial element of today's enemies. John Robb, *Brave New War: The Next Stage of Terrorism and the End of Globalization*, Hoboken, NJ: Wiley, 2007.

⁵⁴ Bruce Hoffman, "The 'Cult of the Insurgent': its tactical and strategic implications," *Australian Journal of International Affairs*, Vol. 61, Issue 3, September 2007, pp. 312-329.

⁵⁵ Lieutenant Colonel Andrew Ross, Royal Marines, "Joint Concept for Disorder, Insurgency, Criminality and Terrorism, J-DICT," Presentation to the Marine Irregular Warfare II Conference, Quantico, VA, July 6, 2005.

Subsequent Supporting Research

Subsequent to the development of this concept and its publication in 2005, a number of authors and analysts have come to the same conclusions. OSD strategists who worked on the QDR regret not having fully documented and explored the Hybrid Warfare phenomena that they believed to pose the most significant threat to U.S. interests in the future.⁵⁶ OSD policy makers are in the process of extrapolating from the last major defense review, and are now gauging the importance of future conflicts in terms of complex irregular wars or hybrid conflicts.⁵⁷

Key scholars are also acknowledging the blending or blurring character of future conflicts. The most historically astute of strategic observers have admitted that while the future is hard to predict, there are clear tendencies and trends that mark tomorrow's path. Professor Colin Gray, of the University of Bristol, has grudgingly admitted that with regard to future conflict in this century, that the one feature "we can predict with confidence is that there is going to be a blurring, a further blurring, of warfare categories."⁵⁸

The American intelligence community has devoted some effort into studying the concept. Numerous studies have been initiated to study the nature of "disruptive challenges" in particular, with CETO asked to provide briefings and insights into these projects. The Director of National Intelligence's long-range conflict team assessed the potential complexity and synergy of hybrid approaches, and published a well received paper about disruptive approaches. Other analysts continue to study the Chinese concept of "unrestricted warfare."⁵⁹ Chinese strategists continue to evolve in their study of future conflict, and adopted their People's War thinking to incorporate both low and high

⁵⁶ Nathan Freier, "Strategic Competition & Resistance in the 21st Century: Irregular, Catastrophic, Traditional, & Hybrid Challenges in Context," Carlisle, PA: Army War College, 2007, p. 96.

⁵⁷ Dr. Thomas Mahnken, paper delivered to the International Institute for Strategic Studies, Global Strategy Review, Geneva Switzerland, September 6, 2007.

⁵⁸ Colin S. Gray, *Another Bloody Century: Future Warfare*, London, UK: Weidenfeld & Nicolson, 2006.

⁵⁹ Daniel Flynn, *The Changing Character of Conflict—Disrupting the American Way of War (U)*, Langley, VA: NIC 2007; Mathew J. Burrows, "Intelligence Community Perspective On the Maturing URW Threat," in Ronald Luman, ed., *Proceedings of the 2007 Unrestricted Warfare Symposium*, Johns Hopkins University/Applied Physics Lab, 2007.

tech, and have extended the concept and the battlespace into the civilian and non-military realm.⁶⁰ These developments bear watching.

British and Australian officers have moved ahead and begun the hard work of drawing out implications and the desired counter-capabilities required to effectively operate against hybrid threats. The British have gone past American and Marine concept writers and already incorporated hybrid threats within their construct for irregular war.⁶¹ Australian security analysts continue to be on the front lines of inquiry in this area.⁶²

Theorists responsible for some of the most cutting edge thinking in alternative modes of war and associated organizational implications continue to explore the blurring of conflict types. Dr. John Arquilla, an expert in irregular warfare tactics and networked forms of organization, has concluded that:

Networks have even shown a capacity to wage war toe-to-toe against nation-states—with some success ... The range of choices available to networks thus covers an entire spectrum of conflict, posing the prospect of a significant blurring of the lines between insurgency, terror, and war.⁶³

Other American and international scholars at the Naval War College in Newport, RI and at King's College London have endorsed and extended the concept.⁶⁴ Max Boot, an Olin Fellow at the Council of Foreign Relations, recently concluded his lengthy study of war and technology with the observation that:

The boundaries between 'regular' and 'irregular' warfare are *blurring*. Even non-state groups are increasingly gaining access to the kinds of weapons that were once the exclusive preserve of states. And even states will increasingly turn to

⁶⁰ See Pen Guangqian and Yao Youzhi, eds., *The Science of Military Strategy* (English version), Beijing: Military Science Publishing House, 2005, p. 117.

⁶¹ Rear Admiral Chris Parry, *Countering Irregular Activity Within A Comprehensive Approach*, Joint Doctrine Note 2/07, United Kingdom, March 2007, p. 1-15.

⁶² Brigadier Michael G. Krause, "Square Pegs for Round Holes?," Australian Army, Australian Land Warfare Studies Centre, *Working Paper No. 132*, June 2007.

⁶³ Arquilla, "The end of war as we knew it," p. 369.

⁶⁴ Mackubin Thomas Owens, "Reflections on Future War," *Naval War College Review*, Forthcoming; David Betz, "A Real Revolution in Military Affairs: Online Continuous Learning for the Operational Warfighter," paper delivered at the Marine Training and Education Command Conference, Pedagogy for the Long War, October 29-31, 2007.

unconventional strategies to blunt the impact of American power.⁶⁵

To conclude this chapter, tomorrow's conflicts will not be easily categorized into simple classifications of conventional or irregular. Numerous security analysts have acknowledged the blurring of lines between modes of war. Conventional and irregular forces, combatants and noncombatants, and even the physical/kinetic and virtual dimensions of conflict are blurring. As Dr. Mike Evans has recently noted in his overview of future conflict:

Armed conflict also began to reflect a bewildering mixture of modes—conventional and unconventional activity merged—while many combatants simultaneously employed modern Kalashnikov assault rifles, pre-modern machetes and post-modern cellular phones in their operations.⁶⁶

⁶⁵ Max Boot, *War Made New: Technology, Warfare, and the Course of History, 1500 to Today*, New York, NY: Random House, 2006, p. 472.

⁶⁶ Michael Evans, "From the Long Peace to the Long War: Armed Conflict and Military Education and Training in the 21st Century," Australian Defence College, Occasional Paper No. 1, 2007, p. 6.

[INTENTIONALLY LEFT BLANK]