
Rationalist Explanations for War

Author(s): James D. Fearon

Source: *International Organization*, Vol. 49, No. 3 (Summer, 1995), pp. 379-414

Published by: The MIT Press

Stable URL: <http://www.jstor.org/stable/2706903>

Accessed: 23-06-2017 03:18 UTC

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.

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



The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *International Organization*

Rationalist explanations for war

James D. Fearon

The central puzzle about war, and also the main reason we study it, is that wars are costly but nonetheless wars recur. Scholars have attempted to resolve the puzzle with three types of argument. First, one can argue that people (and state leaders in particular) are sometimes or always irrational. They are subject to biases and pathologies that lead them to neglect the costs of war or to misunderstand how their actions will produce it. Second, one can argue that the leaders who order war enjoy its benefits but do not pay the costs, which are suffered by soldiers and citizens. Third, one can argue that even rational leaders who consider the risks and costs of war may end up fighting nonetheless.

This article focuses on arguments of the third sort, which I will call rationalist explanations.¹ Rationalist explanations abound in the literature on international conflict, assuming a great variety of specific forms. Moreover, for at least two reasons many scholars have given rationalist explanations a certain pride of place. First, historians and political scientists who have studied the origins of particular wars often have concluded that war can be a rational alternative for leaders who are acting in their states' interest—they find that the expected benefits of war sometimes outweigh the expected costs, however unfortunate

An earlier version of this article was presented at the annual meetings of the American Political Science Association, Washington, D.C., 2–5 September 1993. The article draws in part on chapter 1 of James D. Fearon, "Threats to Use Force: Costly Signals and Bargaining in International Crises," Ph.D. diss., University of California, Berkeley, 1992. Financial support of the Institute on Global Conflict and Cooperation of the University of California is gratefully acknowledged. For valuable comments I thank Eddie Dekel, Eric Gartzke, Atsushi Ishida, Andrew Kydd, David Laitin, Andrew Moravcsik, James Morrow, Randolph Siverson, Daniel Verdier, Stephen Walt and especially Charles Glaser and Jack Levy.

1. Of course, arguments of the second sort may and often do presume rational behavior by individual leaders; that is, war may be rational for civilian or military leaders if they will enjoy various benefits of war without suffering costs imposed on the population. While I believe that "second-image" mechanisms of this sort are very important empirically, I do not explore them here. A more accurate label for the subject of the article might be "rational unitary-actor explanations," but this is cumbersome.

International Organization 49, 3, Summer 1995, pp. 379–414

© 1995 by The IO Foundation and the Massachusetts Institute of Technology

this may be. Second, the dominant paradigm in international relations theory, neorealism, is thought to advance or even to depend on rationalist arguments about the causes of war. Indeed, if no rationalist explanation for war is theoretically or empirically tenable, then neither is neorealism. The causes of war would then lie in the defects of human nature or particular states rather than in the international system, as argued by neorealists. What I refer to here as “rationalist explanations for war” could just as well be called “neorealist explanations.”²

This article attempts to provide a clear statement of what a rationalist explanation for war is and to characterize the full set of rationalist explanations that are both theoretically coherent and empirically plausible. It should be obvious that this theoretical exercise must take place prior to testing rationalist explanations against alternatives—we cannot perform such tests unless we know what a rationalist explanation really is. Arguably, the exercise is also foundational for neorealism. Despite its prominence, neorealist theory lacks a clearly stated and fully conceived explanation for war. As I will argue below, it is not enough to say that under anarchy nothing stops states from using force, or that anarchy forces states to rely on self-help, which engenders mutual suspicion and (through spirals or the security dilemma) armed conflict. Neither do diverse references to miscalculation, deterrence failure because of inadequate forces or incredible threats, preventive and preemptive considerations, or free-riding in alliances amount to theoretically coherent rationalist explanations for war.

My main argument is that on close inspection none of the principal rationalist arguments advanced in the literature holds up as an explanation because none addresses or adequately resolves the central puzzle, namely, that war is costly and risky, so rational states should have incentives to locate negotiated settlements that all would prefer to the gamble of war. The common flaw of the standard rationalist arguments is that they fail either to address or to explain adequately what prevents leaders from reaching *ex ante* (prewar) bargains that would avoid the costs and risks of fighting. A coherent rationalist explanation for war must do more than give reasons why armed conflict might appear an attractive option to a rational leader under some circumstances—it must show why states are unable to locate an alternative outcome that both would prefer to a fight.

To summarize what follows, the article will consider five rationalist arguments accepted as tenable in the literature on the causes of war. Discussed at

2. For the founding work of neorealism, see Kenneth Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979). For examples of theorizing along these lines, see Robert Jervis, “Cooperation Under the Security Dilemma,” *World Politics* 30 (January 1978), pp. 167–214; Stephen Walt, *The Origins of Alliances* (Ithaca, N.Y.: Cornell University Press, 1987); John J. Mearsheimer, “Back to the Future: Instability in Europe After the Cold War,” *International Security* 15 (Summer 1990), pp. 5–56; and Charles Glaser, “Realists as Optimists: Cooperation as Self-Help,” *International Security* 19 (Winter 1994/95), pp. 50–90.

length below, these arguments are given the following labels: (1) anarchy; (2) expected benefits greater than expected costs; (3) rational preventive war; (4) rational miscalculation due to lack of information; and (5) rational miscalculation or disagreement about relative power. I argue that the first three arguments simply do not address the question of what prevents state leaders from bargaining to a settlement that would avoid the costs of fighting. The fourth and fifth arguments do address the question, holding that rational leaders may miss a superior negotiated settlement when lack of information leads them to miscalculate relative power or resolve. However, as typically stated, neither argument explains what prevents rational leaders from using diplomacy or other forms of communication to avoid such costly miscalculations.

If these standard arguments do not resolve the puzzle on rationalist terms, what does? I propose that there are three defensible answers, which take the form of general mechanisms, or causal logics, that operate in a variety of more specific international contexts.³ In the first mechanism, rational leaders may be unable to locate a mutually preferable negotiated settlement due to *private information* about relative capabilities or resolve and *incentives to misrepresent* such information. Leaders know things about their military capabilities and willingness to fight that other states do not know, and in bargaining situations they can have incentives to misrepresent such private information in order to gain a better deal. I show that given these incentives, communication may not allow rational leaders to clarify relative power or resolve without generating a real risk of war. This is not simply a matter of miscalculation due to poor information but rather of specific strategic dynamics that result from the combination of asymmetric information and incentives to dissemble.

Second, rationally led states may be unable to arrange a settlement that both would prefer to war due to *commitment problems*, situations in which mutually preferable bargains are unattainable because one or more states would have an incentive to renege on the terms. While anarchy (understood as the absence of an authority capable of policing agreements) is routinely cited as a cause of war in the literature, it is difficult to find explanations for exactly why the inability to make commitments should imply that war will sometimes occur. That is, what are the specific, empirically identifiable mechanisms by which the inability to commit makes it impossible for states to strike deals that would avoid the costs of war? I identify three such specific mechanisms, arguing in particular that preventive war between rational states stems from a commitment problem rather than from differential power growth per se.

The third sort of rationalist explanation I find less compelling than the first two, although it is logically tenable. States might be unable to locate a peaceful

3. The sense of “mechanism” is similar to that proposed by Elster, although somewhat broader. See Jon Elster, *Political Psychology* (Cambridge: Cambridge University Press, 1993), pp. 1–7; and Jon Elster, *Nuts and Bolts for the Social Sciences* (Cambridge: Cambridge University Press, 1989), chap. 1.

settlement both prefer due to *issue indivisibilities*. Perhaps some issues, by their very natures, simply will not admit compromise. Though neither example is wholly convincing, issues that might exhibit indivisibility include abortion in domestic politics and the problem of which prince sits on the throne of, say, Spain, in eighteenth- or nineteenth-century international politics. Issue indivisibility could in principle make war rational for the obvious reason that if the issue allows only a finite number of resolutions, it might be that none falls within the range that both prefer to fighting. However, the issues over which states bargain typically are complex and multidimensional; side-payments or linkages with other issues typically are possible; and in principle states could alternate or randomize among a fixed number of possible solutions to a dispute. War-prone international issues may often be *effectively* indivisible, but the cause of this indivisibility lies in domestic political and other mechanisms rather than in the nature of the issues themselves.

In the first section of the article I discuss the puzzle posed by the fact that war is costly. Using a simple formalization of the bargaining problem faced by states in conflict, I show that under very broad conditions bargains will exist that genuinely rational states would prefer to a risky and costly fight. The second section argues that rational miscalculations of relative power and resolve must be due to private information and then considers how war may result from the combination of private information and incentives to misrepresent that information in bargaining. In the third section, I discuss commitment problems as the second class of defensible rationalist explanations for war. Throughout, I specify theoretical arguments with simple game-theoretic representations and assess plausibility with historical examples.

Before beginning, I should make it clear that I am not presenting either commitment problems or private information and incentives to misrepresent as wholly novel explanations for war that are proposed here for the first time. The literature on the causes of war is massive, and these ideas, mixed with myriad others, can be found in it in various guises. The main theoretical task facing students of war is not to add to the already long list of arguments and conjectures but instead to take apart and reassemble these diverse arguments into a coherent theory fit for guiding empirical research. Toward this end, I am arguing that when one looks carefully at the problem of explaining how war could occur between genuinely rational, unitary states, one finds that there are really only two ways to do it. The diverse rationalist or neorealist explanations commonly found in the literature fail for two reasons. First, many do not even address the relevant question—what prevents states from locating a bargain both sides would prefer to a fight? They do not address the question because it is widely but incorrectly assumed that rational states can face a situation of deadlock, wherein no agreements exist that both sides would prefer to a war.⁴

4. For an influential example of this common assumption see Glenn Snyder and Paul Diesing, *Conflict Among Nations* (Princeton, N.J.: Princeton University Press, 1977).

Second, the rationalist arguments that do address the question—such as (4) and (5) above—do not go far enough in answering it. When fully developed, they prove to be one of the two major mechanisms developed here, namely, either a commitment problem or a problem arising from private information and incentives to misrepresent. These two mechanisms, I will argue, provide the foundations for a rationalist or neorealist theory of war.

The puzzle

Most historians and political scientists who study war dismiss as naive the view that all wars must be unwanted because they entail destruction and suffering. Instead, most agree that while a few wars may have been unwanted by the leaders who brought them about—World War I is sometimes given as an example—many or perhaps most wars were simply wanted. The leaders involved viewed war as a costly but worthwhile gamble.⁵

Moreover, many scholars believe that wanted wars are easily explained from a rationalist perspective. Wanted wars are thought to be Pareto-efficient—they occur when no negotiated settlements exist that both sides would prefer to the gamble of military conflict. Conventional wisdom holds that while this situation may be tragic, it is entirely possible between states led by rational leaders who consider the costs and risks of fighting. Unwanted wars, which take place despite the existence of settlements both sides preferred to conflict, are thought to pose more of a puzzle, but one that is resolvable and also fairly rare.

The conventional distinction between wanted and unwanted wars misunderstands the puzzle posed by war. The reason is that the standard conception does not distinguish between two types of efficiency—*ex ante* and *ex post*. As long as both sides suffer some costs for fighting, then war is always inefficient *ex post*—both sides would have been better off if they could have achieved the same final resolution without suffering the costs (or by paying lower costs). This is true even if the costs of fighting are small, or if one or both sides viewed the potential benefits as greater than the costs, since there are still costs. Unless states enjoy the activity of fighting for its own sake, as a consumption good, then war is inefficient *ex post*.

From a rationalist perspective, the central puzzle about war is precisely this *ex post* inefficiency. Before fighting, both sides know that war will entail some costs, and even if they expect offsetting benefits they still have an incentive to avoid the costs. The central question, then, is what prevents states in a dispute

5. See, for examples, Geoffrey Blainey, *The Causes of War* (New York: Free Press, 1973); Michael Howard, *The Causes of Wars* (Cambridge, Mass.: Harvard University Press, 1983), especially chap. 1; and Arthur Stein, *Why Nations Cooperate: Circumstance and Choice in International Relations* (Ithaca, N.Y.: Cornell University Press, 1990), pp. 60–64. Even the case of World War I is contested; an important historical school argues that this was a wanted war. See Fritz Fisher, *Germany's Aims in the First World War* (New York: Norton, 1967).

from reaching an *ex ante* agreement that avoids the costs they know will be paid *ex post* if they go to war? Giving a rationalist explanation for war amounts to answering this question.

Three of the most common and widely employed rationalist arguments in the literature do not directly address or answer the question. These are arguments from anarchy, preventive war, and positive expected utility.

Anarchy

Since Kenneth Waltz's influential *Man, the State, and War*, the anarchical nature of the international realm is routinely cited as a root cause of or explanation for the recurrence of war. Waltz argued that under anarchy, without a supranational authority to make and enforce law, "war occurs because there is nothing to prevent it. . . . Among states as among men there is no automatic adjustment of interests. In the absence of a supreme authority there is then the constant possibility that conflicts will be settled by force."⁶

The argument focuses our attention on a fundamental difference between domestic and international politics. Within a well-ordered state, organized violence as a strategy is ruled out—or at least made very dangerous—by the potential reprisals of a central government. In international relations, by contrast, no agency exists that can credibly threaten reprisal for the use of force to settle disputes.⁷ The claim is that without such a credible threat, war will sometimes appear the best option for states that have conflicting interests.

While I do not doubt that the condition of anarchy accounts for major differences between domestic and international politics, and that anarchy encourages both fear of and opportunities for military conflict, the standard framing of the argument is not enough to explain why wars occur and recur. Under anarchy, nothing stops states from using force if they wish. But if using force is a costly option regardless of the outcome, then why is it ever employed? How exactly does the lack of a central authority prevent states from negotiating agreements both sides would prefer to fighting? As it is typically stated, the argument that anarchy provides a rationalist explanation for war does not address this question and so does not solve the problem posed by war's *ex post* inefficiency.

Neither, it should be added, do related arguments invoking the security dilemma, the fact that under anarchy one state's efforts to make itself more secure can have the undesired but unavoidable effect of making another state

6. The quotation is drawn from Kenneth Waltz, *Man, the State, and War: A Theoretical Analysis* (New York: Columbia University Press, 1959), p. 188.

7. For a careful analysis and critique of this standard argument on the difference between the international and domestic arenas, see R. Harrison Wagner, "The Causes of Peace," in Roy A. Licklider, ed., *Stopping the Killing: How Civil Wars End* (New York: New York University Press, 1993), pp. 235–68 and especially pp. 251–57.

less secure.⁸ By itself this fact says nothing about the availability or feasibility of peaceful bargains that would avoid the costs of war. More elaborate arguments are required, and those that are typically given do not envision bargaining and do not address the puzzle of costs. Consider, for instance, a spiral scenario in which an insecure state increases its arms, rendering another so insecure that it decides to attack. If the first state anticipated the reaction producing war, then by itself this is a deadlock argument; I argue against these below. If the first state did not anticipate war and did not want it, then the problem would seem to be miscalculation rather than anarchy, and we need to know why signaling and bargaining could not have solved it. As Robert Jervis has argued, anarchy and the security dilemma may well foster arms races and territorial competition.⁹ But with the exception of occasional references to the preemptive war problem, the standard security dilemma arguments do not explicitly address the question of why the inability to make commitments should necessarily make for war between rational states.¹⁰

Below I will argue that anarchy is indeed implicated as a cause of specific sorts of military conflict (e.g., preventive and preemptive war and in some cases war over strategic territory). In contrast to the standard arguments, however, showing how anarchy figures in a coherent rationalist explanation entails describing the specific mechanism by which states' inability to write enforceable contracts makes peaceful bargains both sides would prefer unattainable.

Preventive war

It frequently is argued that if a declining power expects it might be attacked by a rising power in the future, then a preventive war in the present may be rational. Typically, however, preventive war arguments do not consider whether the rising and declining powers could construct a bargain, perhaps across time, that would leave both sides better off than a costly and risky preventive war would.¹¹ The incentives for such a deal surely exist. The rising state should not want to be attacked while it is relatively weak, so what stops it from offering concessions in the present and the future that would make the declining state prefer not to attack? Also, if war is inefficient and bargains both sides prefer to a fight will exist, why should the declining power rationally fear being attacked in the future? The standard argument supposes that an

8. See John H. Herz, "Idealist Internationalism and the Security Dilemma," *World Politics* 2 (January 1950), pp. 157–80; and Jervis, "Cooperation Under the Security Dilemma." Anarchy is implicated in the security dilemma externality by the following logic: but for anarchy, states could commit to use weapons only for nonthreatening, defensive purposes.

9. Jervis, "Cooperation Under the Security Dilemma."

10. For an analysis of the security dilemma that takes into account signaling, see Andrew Kydd, "The Security Dilemma, Game Theory, and World War I," paper presented at the annual meeting of the American Political Science Association, Washington, D.C., 2–5 September 1993.

11. The most developed exception I know of is found in Stephen Van Evera, "Causes of War," Ph.D. diss., University of California, Berkeley, 1984, pp. 61–64.

anticipated shift in the balance of power can by itself be enough to make war rational, but this is not so.

Positive expected utility

Perhaps the most common informal rationalist explanation found in the literature is that war may occur when two states each estimate that the expected benefits of fighting outweigh the expected costs. As Bruce Bueno de Mesquita argued in an influential formalization of this claim, war can be rational if both sides have positive expected utility for fighting; that is, if the expected utility of war (expected benefits less costs) is greater than the expected utility of remaining at peace.¹²

Informal versions of the expected utility argument typically fail to address the question of how or under what conditions it can be possible for two states both to prefer the costly gamble of war to any negotiated settlement. Formal versions have tended to avoid the question by making various restrictive and sometimes nonrationalist assumptions. To support these claims, I need to be more precise about the expected utility argument.

When will there exist bargains both sides prefer to war?

This section considers the question of whether and when two rationally led states could both prefer war to any negotiated settlement.

Consider two states, A and B, who have preferences over a set of issues represented by the interval $X = [0, 1]$. State A prefers issue resolutions closer to 1, while B prefers outcomes closer to 0. Let the states' utilities for the outcome $x \in X$ be $u_A(x)$ and $u_B(1 - x)$, and assume for now that $u_A(\cdot)$ and $u_B(\cdot)$ are continuous, increasing, and weakly concave (that is, risk-neutral or risk-averse). Without losing any generality, we can set $u_i(1) = 1$ and $u_i(0) = 0$ for both states ($i = A, B$). For concreteness we might think of x as representing the proportion of all territory between A and B that is controlled by A.

In order to say whether the set X contains negotiated settlements that both sides would prefer to conflict, it must be possible to say how the states evaluate the military option versus those outcomes. Almost all analysts of war have

12. See Bruce Bueno de Mesquita, *The War Trap* (New Haven, Conn.: Yale University Press, 1981), and "The War Trap Revisited: A Revised Expected Utility Model," *American Political Science Review* 79 (March 1985), pp. 157–76. For a generalization that introduces the idea of a bargaining range, see James D. Morrow, "A Continuous-Outcome Expected Utility Theory of War," *Journal of Conflict Resolution* 29 (September 1985), pp. 473–502. Informal versions of the expected utility argument are everywhere. For example, Waltz's statement that "A state will use force to attain its goals if, after assessing the prospects for success, it values those goals more than it values the pleasures of peace" appears in different ways in a great many works on war. See Waltz, *Man, the State, and War*, p. 60.

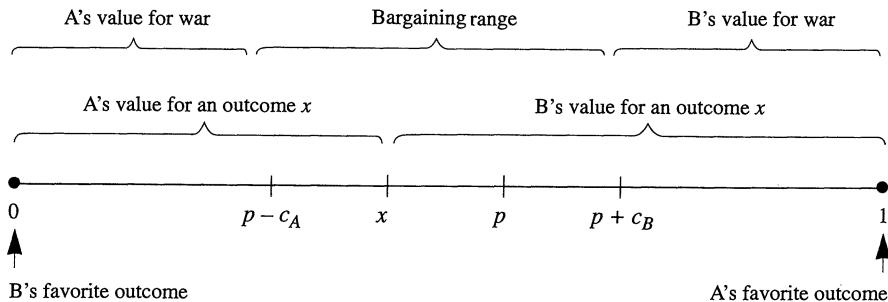


FIGURE 1. *The bargaining range*

stressed that war is a gamble whose outcome may be determined by random or otherwise unforeseeable events.¹³ As Bueno de Mesquita argued, this makes expected utility a natural candidate.¹⁴ Suppose that if the states fight a war, state A prevails with probability $p \in [0, 1]$, and that the winner gets to choose its favorite outcome in the issue space. It follows that A's expected utility for war is $pu_A(1) + (1 - p)u_A(0) - c_A$, or $p - c_A$, where c_A is state A's utility for the costs of a war. Similarly, state B's expected utility for war will be $1 - p - c_B$. Since we are considering rationalist theories for war, we assume that c_A and c_B are both positive. War is thus represented as a costly lottery. (Note that in this formulation the terms c_A and c_B capture not only the states' values for the costs of war but also the value they place on winning or losing on the issues at stake. That is, c_A reflects state A's costs for war relative to any possible benefits. For example, if the two states see little to gain from winning a war against each other, then c_A and c_B would be large even if neither side expected to suffer much damage in a war.)

We can now answer the question posed above. The following result is easily demonstrated: given the assumptions stated in the last two paragraphs, there always exists a set of negotiated settlements that both sides prefer to fighting.¹⁵ Formally, there exists a subset of X such that for each outcome x in this set, $u_A(x) > p - c_A$ and $u_B(1 - x) > 1 - p - c_B$. For example, in the risk-neutral case where $u_A(x) = x$ and $u_B(1 - x) = 1 - x$, both states will strictly prefer any peaceful agreement in the interval $(p - c_A, p + c_B)$ to fighting. This interval represents the bargaining range, with $p - c_A$ and $p + c_B$ as the reservation levels that delimit it. A risk-neutral case is depicted in Figure 1.

This simple but important result is worth belaboring with some intuition. Suppose that two people (or states) are bargaining over the division of \$100—if they can agree on a split they can keep what they agree to. However, in contrast

13. See, for classic examples, Thucydides, *The Peloponnesian War* (New York: Modern Library, 1951), pp. 45 and 48; and Carl von Clausewitz, *On War* (Princeton, N.J.: Princeton University Press, 1984), p. 85.

14. Bueno de Mesquita, *The War Trap*.

15. A proof is given in the Appendix.

to the usual economic scenarios, in this international relations example the players also have an outside option.¹⁶ For a price of \$20, they can go to war, in which case each player has a 50-percent chance of winning the whole \$100. This implies that the expected value of the war option is \$30 ($0.5 \cdot 100 + 0.5 \cdot 0 - 20$) for each side, so that if the players are risk-neutral, then neither should be willing to accept less than \$30 in the bargaining. But notice that there is still a range of peaceful, bargained outcomes from (\$31, \$69) to (\$69, \$31) that make both sides strictly better off than the war option. Risk aversion will tend to increase the range yet further; indeed, even if the leaders pay no costs for war, a set of agreements both sides prefer to a fight will still exist provided both are risk-averse over the issues. In effect, the costs and risks of fighting open up a “wedge” of bargained solutions that risk-neutral or risk-averse states will prefer to the gamble of conflict. The existence of this *ex ante* bargaining range derives from the fact that war is inefficient *ex post*.

Three substantive assumptions are needed for the result, none of which seems particularly strong. First, the states know that there is some true probability p that one state would win in a military contest. As discussed below, it could be that the states have conflicting estimates of the likelihood of victory, and if both sides are optimistic about their chances this can obscure the bargaining range. But even if the states have private and conflicting estimates of what would happen in a war, if they are rational, they should know that there can be only one true probability that one or the other will prevail (perhaps different from their own estimate). Thus rational states should know that there must in fact exist a set of agreements all prefer to a fight.

Second, it is assumed that the states are risk-averse or risk-neutral over the issues. Because risk attitude is defined relative to an underlying metric (such as money in economics), the substantive meaning of this assumption depends on the bargaining context. Loosely, it says that the states prefer a fifty-fifty split or share of whatever is at issue (in whatever metric it comes, if any) to a fifty-fifty chance at all or nothing, where this refers to the value of winning or losing a war. In effect, the assumption means that leaders do not like gambling when the downside risk is losing at war, which seems plausible given the presumption that state leaders normally wish to retain territory and power. A risk-acceptant leader is analogous to a compulsive gambler—willing to accept a sequence of gambles that has the expected outcome of eliminating the state and regime. Even if we admitted such a leader as rational, it seems doubtful that many have held such preferences (Hitler being a possible exception).

16. On the theory of bargaining with outside options, see Martin J. Osborne and Ariel Rubinstein, *Bargaining and Markets* (New York: Academic Press, 1990), chap. 3; Motty Perry, “An Example of Price Formation in Bilateral Situations,” *Econometrica* 50 (March 1986), pp. 313–21; and Robert Powell, “Bargaining in the Shadow of Power” (University of California, Berkeley, 1993, mimeographed). See also the analyses in R. Harrison Wagner, “Peace, War, and the Balance of Power,” *American Political Science Review* 88 (September 1994), pp. 593–607; and Wagner, “The Causes of Peace.”

Finally, it was assumed that a continuous range of peaceful settlements (from 0 to 1) exists. In other words, the issues in dispute are perfectly divisible, so that there are always feasible bargains between the states' reservation levels $p - c_A$ and $p + c_B$. This third assumption immediately suggests a tenable rationalist explanation for war. Perhaps something about the nature of some international issues, such as which successor will sit on a throne, does not admit finely graded divisions and compromise. If so, then small costs for fighting and bad luck may make for rational war over such issues.

But we would immediately like to know what about the nature of an issue makes it impossible to divide up. On more thought, this seems empirically implausible. In the first place, most issues states negotiate over are quite complex—they have many dimensions of concern and allow many possible settlements. Second, if states can simply pay each other sums of money or goods (which they can, in principle), or make linkages with other issues, then this should have the effect of making any issues in dispute perfectly divisible. Before the age of nationalism, princes often bought, sold, and partitioned land.¹⁷ In the nineteenth century the United States purchased the Louisiana territory from France, and Alaska from Russia, and as late as 1898 President McKinley explored the possibility of buying Cuba from Spain in order to avoid a war over it.¹⁸ Third, if something about the nature of an issue means that it can be settled in only, say, two ways, then some sort of random allocation or alternation between the two resolutions could in principle serve to create intermediate bargains. Mafia dons, for example, apparently have avoided costly internal wars by using lotteries to allocate construction contracts among families.¹⁹

In practice, creating intermediate settlements with cash, with linkages to other issues, or with randomization or alternation often seems difficult or impossible for states engaged in a dispute. For example, the immediate issue that gave rise to the Franco-Prussian war was a dispute over which prince would take the Spanish throne. It doubtless occurred to no one to propose that the two candidates alternate year by year, or three years for the Hapsburg and one for the Hohenzollern, or whatever. In this case as in many others, the issue could in principle have been made more continuous and was not for other reasons—here, alternating kings would have violated so many conventions and norms as to have been domestically unworkable. To give a more realistic

17. See, for example, Evan Luard, *War in International Society* (New Haven, Conn.: Yale University Press, 1992), p. 191. Schroeder notes that "patronage, bribes, and corruption" were "a major element" of eighteenth-century international relations. See Paul Schroeder, *The Transformation of European Politics, 1763-1848* (Oxford: Oxford University Press, 1994), p. 579.

18. On Cuba, see Ernest May, *Imperial Democracy* (New York: Harper and Row, 1961), pp. 149-50. On the Louisiana purchase, military threats raised in the U.S. Senate apparently made Napoleon more eager to negotiate the sale. See E. Wilson Lyon, *Louisiana in French Diplomacy* (Norman: University of Oklahoma Press, 1934), pp. 179 and 214ff.

19. Diego Gambetta, *The Sicilian Mafia: The Business of Private Protection* (Cambridge, Mass.: Harvard University Press, 1993), p. 214.

example, nineteenth- and twentieth-century leaders cannot divide up and trade territory in international negotiations as easily as could rulers in the seventeenth and eighteenth centuries, due in part to domestic political consequences of the rise of nationalism; contrast, for example, the Congress of Vienna with the negotiations following World War I.

So in principle the indivisibility of the issues that are the subject of international bargaining can provide a coherent rationalist explanation for war. However, the real question in such cases is what prevents leaders from creating intermediate settlements, and the answer is likely to be other mechanisms (often domestic political) rather than the nature of the issues themselves.²⁰ Both the intrinsic complexity and richness of most matters over which states negotiate and the availability of linkages and side-payments suggest that intermediate bargains typically will exist.

It is thus not sufficient to say that positive expected utility by itself supplies a coherent or compelling rationalist explanation for war. Provided that the issues in dispute are sufficiently divisible, or that side-payments are possible, there should exist a set of negotiated agreements that have greater utility for both sides than the gamble of war does. The reason is that the *ex post* inefficiency of war opens up an *ex ante* bargaining range.

So, to explain how war could occur between rationally led states, we need to answer the following question. Given the existence of an *ex ante* bargaining range, why might states fail either to locate or to agree on an outcome in this range, so avoiding the costs and risks of war?

War due to private information and incentives to misrepresent

Two commonly employed rationalist explanations in the literature directly address the preceding question. Both turn on the claim that war can be and often is the product of rational miscalculation. One explanation holds that a state's leaders may rationally overestimate their chance of military victory against an adversary, so producing a disagreement about relative power that only war can resolve. The other argues that rationally led states may lack information about an adversary's willingness to fight over some interest and so may challenge in the mistaken belief that war will not follow.

In this section I argue that while these ideas point toward a tenable rationalist explanation for war, neither goes far enough and neither works by itself. Both neglect the fact that states can in principle communicate with each other and so avoid a costly miscalculation of relative power or will. The cause of

20. In one of the only articles on this problem, Morrow proposes a private information explanation for states' failures to link issues in many disputes. See James D. Morrow, "Signaling Difficulties with Linkage in Crisis Bargaining," *International Studies Quarterly* 36 (June 1992), pp. 153–72.

war cannot be simply lack of information, but whatever it is that prevents its disclosure. I argue that the fact that states have incentives to misrepresent their positions is crucial here, explaining on rationalist terms why diplomacy may not allow rational states to clarify disagreements about relative power or to avoid the miscalculation of resolve.

The mainstream international relations literature recognizes the existence of both private information and incentives to misrepresent, but typically views them as background conditions to be taken for granted rather than as key elements of an explanation of how rationally led states might end up at war. For example, Jack Levy's impressive review of the literature on the causes of war contains nothing on the role of incentives to misrepresent and discusses private information largely in the context of misperceptions of other states' intentions (which are linked to psychological biases). This is an accurate reflection of where these factors stand in the mainstream literature.²¹

Disagreements about relative power

Geoffrey Blainey's well-known and often-cited argument is that "wars usually begin when two nations disagree on their relative strength."²² It is easy to see how a disagreement about relative strength—understood as conflicting estimates of the likelihood of military victory—can eliminate any *ex ante* bargaining range. Recall the example given above, where two states bargain over the division of \$100, and each has the outside option of going to war. If each expects that it surely would prevail at war, then each side's expected value for the war option is \$80 ($1 \cdot 100 + 0 \cdot 0 - 20$). So given these expectations, neither side will accept less than \$80 in the bargaining, implying that no negotiated outcome is mutually preferred to war. More generally, suppose that state A expects to win with probability p , state B expects to win with probability r , and p and r sum to greater than one. Such conflicting expectations will certainly shrink and could eliminate any *ex ante* bargaining range.

But how could rationally led states have conflicting expectations about the likely outcome of military conflict? In the extreme case, how could both sides rationally expect to win? The literature barely addresses this question in

21. See Jack Levy, "The Causes of War: A Review of Theories and Evidence," in Philip E. Tetlock et al., eds., *Behavior, Society, and Nuclear War*, vol. 1 (Oxford: Oxford University Press, 1989), pp. 209–333. Recent work using limited-information game theory to analyze crisis bargaining places the strategic consequences of private information at the center of the analysis. See, for examples, Bruce Bueno de Mesquita and David Lalman, *War and Reason* (New Haven, Conn.: Yale University Press, 1992); James D. Fearon, "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review* 88 (September 1994), pp. 577–92; James D. Morrow, "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining," *American Journal of Political Science* 33 (November 1989), pp. 941–72; Barry Nalebuff, "Brinksmanship and Nuclear Deterrence: The Neutrality of Escalation," *Conflict Management and Peace Science* 9 (Spring 1986), pp. 19–30; and Robert Powell, *Nuclear Deterrence Theory: The Problem of Credibility* (Cambridge: Cambridge University Press, 1990).

22. Blainey, *The Causes of War*, p. 246.

explicit terms. Blainey, whom the literature views as advancing a rationalist explanation for war, in fact explains disagreements about relative power as a consequence of human *irrationality*. He says that mutual optimism about victory in war is the product of “moods which cannot be grounded in fact” and which “permeate what appear to be rational assessments of the relative military strength of two contending powers.” Mutual optimism is said to result from a “process by which nations evade reality,” which hardly sounds like a rationalist explanation.²³

Conflicting expectations about the likely outcome of military conflict may be explained in three ways. First, as Blainey suggests, emotional commitments could irrationally bias leaders’ military estimates. They might, for instance, come to believe nationalist rhetoric holding that their soldiers are more courageous and spirited than those of the adversary.²⁴ Second, the world is a very complex place, and for this reason military analysts in different states could reach different conclusions about the likely impact of different technologies, doctrines, and tactics on the expected course of battle. Third, state leaders might have private information about militarily relevant factors—military capabilities, strategy, and tactics; the population’s willingness to prosecute a long war; or third-state intentions. If a state has superior (and so private) information about any such factor, then its estimate of the probable course of battle may differ from that of an adversary.

Under a strict but standard definition of rationality, only the third explanation qualifies as an account of how rationally led states could have conflicting estimates of the probability of winning in war. As argued by John Harsanyi, if two rational agents have the same information about an uncertain event, then they should have the same beliefs about its likely outcome.²⁵ The claim is that given identical information, truly rational agents should reason to the same conclusions about the probability of one uncertain outcome or another. Conflicting estimates should occur only if the agents have different (and so necessarily private) information.²⁶

23. Ibid., p. 54. Blainey also blames patriotic and nationalistic fervor, leaders’ (irrational) tendency to surround themselves with yes-men, and crowd psychology.

24. See Ralph K. White, *Nobody Wanted War: Misperception in Vietnam and Other Wars* (New York: Doubleday/Anchor), chap. 7; Blainey, *The Causes of War*, p. 54; and Richard Ned Lebow, *Between Peace and War: The Nature of International Crises* (Baltimore, Md.: Johns Hopkins University Press, 1981), p. 247.

25. John C. Harsanyi, “Games with Incomplete Information Played By ‘Bayesian’ Players, Part III,” *Management Science* 14 (March 1968), pp. 486–502.

26. Aumann observed an interesting implication of this doctrine: genuinely rational agents cannot “agree to disagree,” in the sense that it cannot be commonly known that they are rational and that they hold different estimates of the likelihood of some uncertain event. See Robert Aumann, “Agreeing to Disagree,” *The Annals of Statistics* 4 (November 1976), pp. 1236–39. Emerson Niou, Peter Ordeshook, and Gregory Rose note that this implies that rational states cannot agree to disagree about the probability that one or the other would win in a war in *The Balance of Power: Stability in the International System* (Cambridge: Cambridge University Press, 1989), p. 59.

It follows that the second explanation for disagreements about relative power listed above—the complexity of the world—is not a rationalist account. Instead, it is an account that explains conflicting military estimates as a consequence of bounded rationality. In this view, leaders or military analysts with the same information about military technology, strategy, political will, etc. might reason to different conclusions about the likely course of a war because of differential ability to cope with complexity of the problem. This is entirely plausible, but it is a bounded rationality explanation rather than a fully rationalist one.²⁷

The rationalist account of how disagreements about the probability of winning might arise also seems empirically plausible. States certainly have private information about factors affecting the likely course of battle—for example, they jealously guard military secrets and often have superior information about what an ally will or will not fight for. Nonetheless, while private information about militarily relevant capabilities provides a first step, it does not provide a coherent rationalist explanation for war. The problem is that even if leaders have such private information, they should understand that their own estimates based on this information are suspect because they do not know the other side's private information. In principle, both sides could gain by sharing information, which would yield a consensus military estimate (absent bounded rationality). And, as shown above, doing so could not help but reveal bargains that both would prefer to a fight.²⁸

So the question of how rationally led states can disagree about relative power devolves to the question of what prevents states from sharing private information about factors that might affect the course of battle. Before turning to this question, I will consider the second common explanation for how a rational miscalculation may produce war.

War due to the miscalculation of an opponent's willingness to fight

Many wars have been given the following so-called rationalist explanation: state A transgressed some interest of state B in the erroneous belief that B would not fight a war over the matter. Though rationally led, state A lacked information about B's willingness to fight and simply happened to guess wrong, causing a war. Thus, some say that Germany miscalculated Russian and/or British willingness to fight in 1914; Hitler miscalculated Britain and France's

27. On bounded rationality, see Herbert A. Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69 (February 1955), pp. 99–118.

28. This analysis runs exactly parallel to work in law and economics on pretrial bargaining in legal disputes. Early studies explained costly litigation as resulting from divergent expectations about the likely trial outcome, while in more recent work such expectations derive from private information about the strength of one's case. For a review and references, see Robert D. Cooter and Daniel L. Rubinfeld, "Economic Analysis of Legal Disputes and Their Resolution," *Journal of Economic Literature* 27 (September 1989), pp. 1067–97.

willingness to resist his drive to the east; Japanese leaders in 1941 miscalculated U.S. willingness to fight a long war over control in the South Pacific; North Korea miscalculated U.S. willingness to defend South Korea; the United States miscalculated China's willingness to defend North Korea; and so on. In each case, the argument would hold that lack of information led a more-or-less rational actor to guess wrong about the extent of the bargaining range.

Blainey has argued that if states agree on relative power they are very unlikely to go to war against each other.²⁹ It is worth pointing out that in the preceding argument, war can occur despite complete agreement on relative power across states. To show how and for later use, I will introduce a simple model of international bargaining. As in the empirical examples just mentioned, in the model one state unilaterally chooses some revision of the status quo. The second state can then either acquiesce to the revision or can go to war to reverse it.

Formally, suppose there is a status quo resolution of the issues $q \in X$ and that state A has the opportunity to choose any outcome $x \in X$, presenting state B with a fait accompli. On observing what state A did (which might be nothing, i.e., $x = q$), state B can choose whether to go to war or to acquiesce to A's revision of the status quo.

If neither state has any private information, so that all payoffs are common knowledge, state A does best to push the outcome just up to B's reservation level $p + c_B$, which makes B just willing to acquiesce rather than go to war. With complete information, then, the states avoid the inefficient outcome of war.³⁰ On the other hand, if state B has private information about either its capabilities (which affect p) or its value for the issues at stake relative to the costs of conflict (c_B), then state A may not know whether a particular "demand" x will yield war or peace. Lacking this information, state A faces a trade-off in deciding whether and how much territory to "grab": The larger the grab, the greater the risk of war, but the better off A will be if state B acquiesces.

Suppose, for example, that A and B share a common estimate of p —they agree about relative power—but that A is unsure about B's costs for fighting. Under very broad conditions, if A cannot learn B's private information and if A's own costs are not too large, then state A's optimal grab produces a positive chance of war.³¹ Intuitively, if A is not too fearful of the costs of war relative to

29. Blainey, *The Causes of War*.

30. This take-it-or-leave-it model of international bargaining is proposed and analyzed under conditions of both complete and incomplete information in James D. Fearon, "Threats to Use Force: The Role of Costly Signals in International Crises," Ph.D. diss., University of California, Berkeley, 1992, chap. 1. Similar results for more elaborate bargaining structures are given in my own work in progress. See James D. Fearon, "Game-Theoretic Models of International Bargaining: An Overview," University of Chicago, 1995. Powell has analyzed an alternative model in which both sides must agree if the status quo is to be revised. See Powell, "Bargaining in the Shadow of Power."

31. See Claim 2 in the Appendix.

what might be gained in bargaining, it will run some risk of war in hopes of gaining on the ground. So Blainey's suggestion that a disagreement about relative power is necessary for war is incorrect—all that is necessary is that the states in dispute be unable to locate or agree on some outcome in the bargaining range. Since the bargaining range is determined not just by relative power but also by states' values for the issues at stake relative to the costs of fighting, uncertainty about the latter can (and apparently does) produce war.

Once again, it is entirely plausible that state leaders have private information about their value for various international interests relative to their costs of fighting over them.³² Thus it seems we have a second tenable rationalist explanation for war, again based on the concept of private information. But as in the case of disagreements about relative power, the explanation fails as given because it does not explain why states cannot avoid miscalculating a potential opponent's willingness to fight. In the model, why cannot state A simply ask state B whether it would fight rather than acquiesce to a particular demand? To give a concrete example, why did German leaders in 1914 not simply ask their British and Russian counterparts what they would do if Austria were to attack Serbia? If they could have done so and if the answers could have been believed, the Germans might not have miscalculated concerning Russian and, more importantly, British willingness to fight. In consequence they might have avoided the horrendous costs of World War I.

To recap, I have argued that in a rationalist framework, disagreements about relative power and uncertainty about a potential opponent's willingness to fight must have the same source: leaders' private information about factors affecting the likely course of a war or their resolve to fight over specific interests. In order to avoid war's *ex post* inefficiency, leaders have incentives to share any such private information, which would have the effect of revealing peaceful settlements that lie within the bargaining range. So, to explain how war could occur between states led by rational leaders who consider the costs of fighting, we need to explain what would prevent them from sharing such private information.

Incentives to misrepresent in bargaining

Prewar bargaining may fail to locate an outcome in the bargaining range because of strategic incentives to withhold or misrepresent private information. While states have an incentive to avoid the costs of war, they also wish to obtain a favorable resolution of the issues. This latter desire can give them an incentive to exaggerate their true willingness or capability to fight, if by doing so they might deter future challenges or persuade the other side to make concessions. States can also have an incentive to conceal their capabilities or resolve, if they are concerned that revelation would make them militarily (and

32. For examples and discussion on this point, see Fearon, "Threats to Use Force," chap. 3.

hence politically) vulnerable or would reduce the chances for a successful first strike. Similarly, states may conceal their true willingness to fight in order to avoid appearing as the aggressor.

Combined with the fact of private information, these various incentives to misrepresent can explain why even rational leaders may be unable to avoid the miscalculations of relative will and power that can cause war. This section first considers why this is so theoretically and then discusses two empirical examples.

A drawback of the simple bargaining model given above was that state B had no opportunity to try to communicate its willingness to fight to state A. It is easy to imagine that if communication were possible—say, if B could announce what interests in X it considered vital enough to fight over—this might at least lower the chance of war by miscalculation. To check this, we give state B an initial opportunity to make a foreign policy announcement f , which can be any statement about its foreign policy or what it considers to be vital or peripheral interests. (Assume as before that A is uncertain about B's capabilities or costs for fighting.)

If the announcement itself has no effect on either side's payoffs, then it can be shown that in any equilibrium in which state A does not choose randomly among demands, A will make the same demand regardless of what state B says, and the *ex ante* risk of war will remain the same as in the game without communication by state B. To gain an intuition for these results, suppose that A conditioned its behavior on f , grabbing more or less depending on what B announced. Then regardless of B's true willingness to fight, B does best to make the announcement that leads to the smallest grab by A—that is, B has an incentive to misrepresent its actual willingness to resist. But then A learns nothing from the announcement.³³

This conclusion is slightly altered if the leaders of B can render the announcement f costly to make.³⁴ In practice, five common methods include building weapons, mobilizing troops, signing alliance treaties, supporting troops in a foreign land, and creating domestic political costs that would be paid if the announcement proves false. Of course, signaling by means of domestic political audience costs lies outside a purely unitary rational-actor framework, since this presumes a state run by an agent on behalf of a principal (the “audience”) rather than a unitary state with a perfectly secure leadership. In the latter case, leaders may be able to make foreign policy announcements

33. See the Appendix for proofs of these claims. Cheap talk announcements can affect outcomes in some bargaining contexts. For an example from economics, see Joseph Farrell and Robert Gibbons, “Cheap Talk Can Matter in Bargaining,” *Journal of Economic Theory* 48 (June 1989), pp. 221–37. These authors show how cheap talk might credibly signal a willingness to negotiate seriously that then affects subsequent terms of trade. For an example from international relations, see James D. Morrow, “Modeling the Forms of International Cooperation: Distribution Versus Information,” *International Organization* 48 (Summer 1994), pp. 387–423.

34. The conclusion is likewise altered if the possibility of repeated interactions in sufficiently similar contexts is great enough that reputation building can be supported.

credible only by engaging an international reputation, taking financially costly mobilization measures, or bearing the costs and risks of limited military engagements.³⁵

Even when the signal is costly, however, this will not in general completely eliminate all risk of war by miscalculation—indeed, it may even increase it. The reason concerns the nature of the signals that states have incentives to send. To be genuinely informative about a state's actual willingness or ability to fight, a signal must be costly in such a way that a state with lesser resolve or capability might not wish to send it. Actions that generate a real risk of war—for example, troop mobilizations that engage a leadership's reputation before international or domestic audiences—can easily satisfy this constraint, since states with high resolve are less fearful of taking them. In other words, a rational state may choose to run a real risk of (inefficient) war in order to signal that it will fight if not given a good deal in bargaining.³⁶

The July crisis of World War I provides several examples of how incentives to misrepresent can make miscalculations of resolve hard to dispel. Soon after German leaders secretly endorsed Austrian plans to crush Serbia, they received both direct and indirect verbal indications from St. Petersburg that Russia would fight rather than acquiesce.³⁷ For example, on 21 July, the Russian Foreign Minister told the German ambassador that “Russia would not be able to tolerate Austria-Hungary’s using threatening language to Serbia or taking military measures.”³⁸ Such verbal statements had little effect on German leaders’ beliefs, however, since they knew Russian leaders had a strategic incentive to misrepresent. On 18 July in a cable explaining Berlin’s policy to Ambassador Lichnowsky in London, Secretary of State Jagow wrote that “there is certain to be some blustering in St. Petersburg.”³⁹ Similarly, when on 26 July Lichnowsky began to report that Britain might join with France and Russia in the event of war, German Chancellor Bethmann Hollweg told his personal assistant of the “danger that France and England will commit their support to Russia in order not to alienate it, perhaps without really believing

35. On signaling costs in crises and audience costs in particular, see Fearon, “Threats to Use Force,” and “Domestic Political Audiences and the Escalation of International Disputes.” For an excellent analysis of international signaling in general, see Robert Jervis, *The Logic of Images in International Relations* (Princeton, N.J.: Princeton University Press, 1970).

36. For developed models that make this point, see James Fearon, “Deterrence and the Spiral Model: The Role of Costly Signals in Crisis Bargaining,” paper presented at the annual meeting of the American Political Science Association, 30 August–2 September 1990, San Francisco, Calif.; Fearon, “Domestic Political Audiences and the Escalation of International Disputes”; Morrow, “Capabilities, Uncertainty, and Resolve”; Nalebuff, “Brinkmanship and Nuclear Deterrence”; and Powell, *Nuclear Deterrence Theory*.

37. Luigi Albertini, *The Origins of the War of 1914*, vol. 2 (London: Oxford University Press, 1953), pp. 183–87.

38. Ibid., p. 187.

39. Ibid., p. 158. For the full text of the cable, see Karl Kautsky, comp., *German Documents Relating to the Outbreak of the World War* (New York: Oxford University Press, 1924), doc. no. 71, p. 130.

that for us mobilization means war, thinking of it as a bluff which they answer with a counterbluff.”⁴⁰

At the same time, the Chancellor had an incentive to misrepresent the strength and nature of German support for Austria’s plans. Bethmann correctly anticipated that revealing this information would make Germany appear the aggressor, which might undermine Social Democratic support for his policies in Germany as well as turn British public opinion more solidly against his state.⁴¹ This incentive led the Chancellor to avoid making direct or pointed inquiries about England’s attitude in case of war. The incentive also led him to pretend to go along with the British Foreign Secretary’s proposals for a conference to mediate the dispute.⁴² In consequence, Lord Grey may not have grasped the need for a stronger warning to Germany until fairly late in the crisis (on 29 July), by which time diplomatic and military actions had made backing off more difficult for both Austria and Germany.

In July 1914, incentives to misrepresent private information fostered and supported miscalculations of willingness to fight. Miscalculations of relative power can arise from this same source. On the one hand, states at times have an incentive to exaggerate their capabilities in an attempt to do better in bargaining. On the other hand, they can also have the well-known incentive to withhold information about capabilities and strategy. Presumably because of the strongly zero-sum aspect of military engagements, a state that has superior knowledge of an adversary’s war plans may do better in war and thus in prewar bargaining—hence, states rarely publicize war plans. While the theoretical logic has not been worked out, it seems plausible that states’ incentives to conceal information about capabilities and strategy could help explain some disagreements about relative power.

The 1904 war between Japan and Russia serves to illustrate this scenario. On the eve of the war, Russian leaders believed that their military could almost certainly defeat Japan.⁴³ In this conviction they differed little from the view of most European observers. By contrast, at the imperial council of 4 February that decided for war, the Japanese chief of staff estimated a fifty-fifty chance of

40. Konrad Jarausch, “The Illusion of Limited War: Chancellor Bethmann Hollweg’s Calculated Risk,” *Central European History* 2 (March 1969), pp. 48–76. The quotation is drawn from p. 65.

41. See L. C. F. Turner, *Origins of the First World War* (New York: Norton, 1970), p. 101; and Jarausch, “The Illusion of Limited War,” p. 63. Trachtenberg writes that “one of Bethmann’s basic goals was for Germany to avoid coming across as the aggressor.” See Marc Trachtenberg, *History and Strategy* (Princeton, N.J.: Princeton University Press, 1991), p. 90.

42. Albertini concludes that “on the evening of the 27th all the Chancellor sought to do was to throw dust in the eyes of Grey and lead him to believe that Berlin was seriously trying to avert a conflict, that if war broke out it would be Russia’s fault and that England could therefore remain neutral.” See Albertini, *The Origins of the War of 1914*, vol. 1, pp. 444–45. See also Turner, *Origins of the First World War*, p. 99.

43. See J. A. White, *The Diplomacy of the Russo-Japanese War* (Princeton, N.J.: Princeton University Press, 1964), pp. 142–43; and Ian Nish, *The Origins of the Russo-Japanese War* (London: Longman, 1985), pp. 241–42.

prevailing, if their attack began immediately.⁴⁴ Thus Japanese and Russian leaders disagreed about relative power—their estimates of the likelihood of victory summed to greater than 1.

Moreover, historical accounts implicate this disagreement as a major cause of the war: Russia's refusal to compromise despite repeated offers by the Japanese was motivated in large measure by their belief that Japan would not dare attack them. The Japanese Cabinet finally decided for war after the Tsar and his advisers failed to make any real compromises over Korea or Manchuria in a series of proposals exchanged in 1903. The Tsar and his top advisers were hardly eager to fight, not because they expected to lose but because they saw an Asian war as a costly diversion of resources to the wrong theater.⁴⁵ Nonetheless, they refused to make concessions from what they viewed as a position of great military strength. They believed that Japan would have to settle for less, given its relative military weakness.⁴⁶

The disagreement arose in substantial part from Japanese private information about their military capabilities and how they compared with Russia's. A far superior intelligence service had provided the Japanese military with a clear picture of Russian strengths and weaknesses in Northeast Asia and enabled them to develop an effective offensive strategy. According to John Albert White, due to this intelligence "the Japanese government apparently faced the war with a far more accurate conception of their task than their enemy had."⁴⁷ In addition, compared with the Russians or indeed with any European power, Japanese leaders had much better knowledge of the fighting ability of the relatively untested Japanese army and of the effect of the reforms, training, and capital development of the previous decade.⁴⁸

If by communicating this private information the Japanese could have led the Russians to see that their chances of victory were smaller than expected, they might have done so. Almost all historians who have carefully examined the case agree that the Japanese government was not bent on war for its own sake—they

44. J. N. Westwood, *Russia Against Japan, 1904–5: A New Look at the Russo–Japanese War* (Albany: State University of New York Press, 1986), p. 22. Estimates varied within the Japanese leadership, but with the exception of junior-level officers, few seem to have been highly confident of victory. For example, as the decision for war was taken the Japanese navy requested a two-week delay to allow it to even the odds at sea. See Nish, *The Origins of the Russo–Japanese War*, pp. 197–200 and 206–7.

45. See, for example, David Walder, *The Short Victorious War: The Russo–Japanese Conflict, 1904–5* (London: Hutchinson, 1973), pp. 53–56; and Nish, *The Origins of the Russo–Japanese War*, p. 253.

46. See White, *The Diplomacy of the Russo–Japanese War*, chaps. 6–8; Nish, *The Origins of the Russo–Japanese War*, p. 241; and Lebow, *Between Peace and War*, pp. 244–46.

47. White, *The Diplomacy of the Russo–Japanese War*, p. 139. Nish writes that "many Russians certainly took a view of [the Japanese military] which was derisory in comparison with themselves. It may be that this derived from a deliberate policy of secrecy and concealment which the Japanese army applied because of the historic coolness between the two countries." See Nish, *The Origins of the Russo–Japanese War*, p. 241.

48. The British were the major exception, who as recent allies of Japan had better knowledge of its capabilities and level of organization. See Nish, *The Origins of the Russo–Japanese War*, p. 241.

were willing to compromise if the Russians would as well.⁴⁹ However, it was unthinkable for the Japanese to reveal such information or convince the Russians even if they did. In the first place, the Japanese could not simply make announcements about the quality of their forces, since the Russians would have had no reason to believe them. Second, explaining how they planned to win a war might seriously compromise any such attempt by changing the likelihood that they would win; there is a trade-off between revealing information about resolve or capabilities to influence bargaining and reducing the advantages of a first strike.

In sum, the combination of private information about relative power or will and the strategic incentive to misrepresent these afford a tenable rationalist explanation for war. While states always have incentives to locate a peaceful bargain cheaper than war, they also always have incentives to do well in the bargaining. Given the fact of private information about capabilities or resolve, these incentives mean that states cannot always use quiet diplomatic conversations to discover mutually preferable settlements. It may be that the only way to surmount this barrier to communication is to take actions that produce a real risk of inefficient war.

This general mechanism operates in at least two other empirically important ways to produce conflict in specific circumstances. First, private information about the costs of fighting or the value leaders place on international interests can give them an incentive to cultivate a reputation for having lower costs or more far-flung vital interests than they actually do. If cutting a deal in one dispute would lead other states to conclude the leader's costs for using force are high, then the leader might choose a costly war rather than suffer the depredations that might follow from making concessions. The U.S. interventions in Korea and Vietnam are sometimes explained in these terms, and states surely have worried about such inferences drawn by other states for a long time.⁵⁰ The same logic operates when a small state or group (for example, Finland or the Chechens) chooses to fight a losing war against a larger one (for example, the Soviet Union or Russia) in order to develop a reputation for being hard to subjugate. In both cases, states employ war itself as a costly signal of privately known and otherwise unverifiable information about willingness to fight.

Second, since incentives to misrepresent military strength can undermine diplomatic signaling, states may be forced to use war as a credible means to reveal private information about their military capabilities. Thus, a rising state may seek out armed conflict in order to demonstrate that it is more powerful

49. See, for example, William Langer, "The Origins of the Russo-Japanese War," in Carl Schorske and Elizabeth Schorske, eds., *Explorations in Crisis* (Cambridge: Harvard University Press, 1969), p. 44.

50. For some examples, see Fearon, "Threats to Use Force," chap. 3. For a formal version of reputational dynamics due to private information, see Barry Nalebuff, "Rational Deterrence in an Imperfect World," *World Politics* 43 (April 1991), pp. 313–35.

than others realize, while a state in apparent decline may fight in hope of revealing that its capabilities remain better than most believe. In both instances, the inefficient outcome of war derives from the fact that states have private information about their capabilities and a strategic incentive to misrepresent it to other states.

War as a consequence of commitment problems

This section considers a second and quite different rationalist mechanism by which war may occur even though the states in dispute share the same assessment of the bargaining range. Even if private information and incentives to misrepresent it do not tempt states into a risky process of discovery or foster costly investments in reputation, states may be unable to settle on an efficient bargained outcome when for structural reasons they cannot trust each other to uphold the deal.

In this class of explanations, the structural condition of anarchy reemerges as a major factor, although for nonstandard reasons. In the conventional argument, anarchy matters because no hegemonic power exists to threaten states with “jail” if they use force. Without this threat, states become suspicious and worried about other states’ intentions; they engage in self-help by building weapons; and somehow uncertainty-plus-weapons leads them ultimately to attack each other (the security dilemma or spiral model). Below, I show that anarchy does indeed matter but for more specific reasons and in more specific contexts. Anarchy matters when an unfortunate combination of state preferences and opportunities for action imply that one or both sides in a dispute have incentives to renege on peaceful bargains which, if they were enforceable, would be mutually preferred to war. I will consider three such unfortunate situations that can claim some empirical plausibility.

It should be stressed that in standard security dilemma and spiral model arguments the suspicions and lack of trust engendered by anarchy are understood to originate either from states’ inability to observe each other’s motivations (that is, from private information about greed or desire for conquest) or from the knowledge that motivations can change.⁵¹ By contrast, in the arguments given below, states have no private information and motivations never change; thus states understand each other’s motivations perfectly. This is not to argue that private information about the value a leadership places on expansion is unimportant in international politics—it surely is. Indeed, private information about motivation and various incentives to misrepresent it might exacerbate any of the three specific commitment problems discussed below.

51. See, for examples, Robert Jervis, *Perception and Misperception in International Politics* (Princeton, N.J.: Princeton University Press, 1976), pp. 62–67; Barry Posen, *The Sources of Military Doctrine* (Ithaca, N.Y.: Cornell University Press, 1984), pp. 16–17; and Charles Glaser, “The Political Consequences of Military Strategy,” *World Politics* 44 (July 1992), p. 506.

However, when they do so this is a matter of an interaction between informational and commitment problems rather than of anarchy per se. Our first task should be to isolate and specify the mechanisms by which anarchy itself might cause war.

Preemptive war and offensive advantages

Consider the problem faced by two gunslingers with the following preferences. Each would most prefer to kill the other by stealth, facing no risk of retaliation, but each prefers that both live in peace to a gunfight in which each risks death. There is a bargain here that both sides prefer to “war”—namely, that each leaves the other alone—but without the enforcement capabilities of a third party, such as an effective sheriff, they may not be able to attain it. Given their preferences, neither person can credibly commit not to defect from the bargain by trying to shoot the other in the back. Note that no matter how far the shadow of the future extends, iteration (or repeat play) will not make cooperation possible in strategic situations of this sort. Because being the “sucker” here may mean being permanently eliminated, strategies of conditional cooperation such as tit-for-tat are infeasible.⁵² Thus, if we can find a plausible analogy in international relations, this example might afford a coherent rationalist explanation for war.

Preemptive war scenarios provide the analogy. If geography or military technology happened to create large first-strike or offensive advantages, then states might face the same problem as the gunslingers. To demonstrate this theoretically, I consider how offensive advantages affect the bargaining range between two states engaged in a dispute.

There are at least three ways of interpreting offensive advantages in a formal context. First, an offensive advantage might mean that a state’s odds of winning are better if it attacks rather than defends. Second, an offensive advantage might mean that the costs of fighting are lower for an attacking state than for a defending state. It can be shown that no commitment problem operates in this second case, although lowering the costs of war for attackers does narrow the de facto bargaining range. Third, offensive advantages might mean that military technology and doctrine increase the variance of battlefield outcomes. That is, technology and doctrine might make total victory or total defeat more likely, while rendering stalemate and small territorial changes less likely. In this case, offensive advantages can actually reduce the expected utility of war for both sides, thus increasing the bargaining range and perhaps making war less rather

52. For dynamic game models that demonstrate this, see Robert Powell, “Absolute and Relative Gains in International Relations Theory,” *American Political Science Review* 85 (December 1991), pp. 1303–20; and James D. Fearon, “Cooperation and Bargaining Under Anarchy,” (University of Chicago, 1994, mimeographed). On tit-for-tat and the impact of the shadow of the future, see Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); and Kenneth Oye, ed., *Cooperation Under Anarchy* (Princeton, N.J.: Princeton University Press, 1986).

than more likely. Intuitively, if states care most of all about security (understood as survival), then offensive advantages make war less safe by increasing the risk of total defeat.⁵³

A commitment problem of the sort faced by the gunslingers arises only under the first interpretation, in which “offensive advantage” refers to an increase in a state’s military prospects if it attacks rather than defends. To demonstrate this, let p_f be the probability that state A wins a war if A attacks; p_s the probability that A wins if A strikes second or defends; and p the chance that A wins if both states mobilize and attack at the same time. Thus, an offensive advantage exists when $p_f > p > p_s$.

Since states can always choose to attack if they wish, a peaceful resolution of the issues is feasible only if neither side has an incentive to defect unilaterally by attacking. In the risk-neutral case, we must have $x > p_f - c_A$ for state A to prefer not to attack and $1 - x > 1 - p_s - c_B$ for state B to prefer the bargained outcome x . Thus stable outcomes that both sides prefer to conflict exist provided that $p_f - c_A < p_s + c_B$, implying a de facto bargaining range represented by the interval $(p_f - c_A, p_s + c_B)$.

Notice that as p_f increases above p , and p_s decreases below it, this interval shrinks and may even disappear. Thus, first-strike advantages narrow the de facto bargaining range, while second-strike (or defensive) advantages increase it. The reason is that when first-strike advantages are large, both states must be given more from the peacetime bargain in order to allay the greater temptation of unilateral attack.

In the extreme case, if $p_f - c_A > p_s + c_B$ or $p_f - p_s > c_A + c_B$, no self-enforcing peaceful outcomes exist. This does not mean that no bargains exist that both sides would prefer to war. Since by definition both states cannot enjoy the advantage of going first, agreements that both sides prefer to fighting are always available in principle. The problem is that under anarchy, large enough first-strike incentives (relative to cost–benefit ratios) can make all of these agreements unenforceable and incredible as bargains.

Does this prisoners’ dilemma logic provide an empirically plausible explanation for war? Though I lack the space to develop the point, I would argue that first-strike and offensive advantages probably are an important factor making war more likely in a few cases, but not because they make mobilization and attack a dominant strategy, as in the extreme case above. In the pure preemptive war scenario leaders reason as follows: “The first-strike advantage is so great that regardless of how we resolve any diplomatic issues between us, one side will always want to attack the other in an effort to gain the (huge) advantage of going first.” But even in July 1914, a case in which European

53. This argument about military variance runs counter to the usual hypothesis that offensive advantages foster war. For a discussion and an empirical assessment, see James D. Fearon, “Offensive Advantages and War since 1648,” paper presented at the annual meeting of the International Studies Association, 21–25 February 1995. On the offense–defense balance and war, see Jervis, “Cooperation Under the Security Dilemma”; and Van Evera, “Causes of War,” chap. 3.

leaders apparently held extreme views about the advantage of striking first, we do not find leaders thinking in these terms.⁵⁴ It would be rather surprising if they did, since they had all lived at peace but with the same military technology prior to July 1914. Moreover, in the crisis itself military first-strike advantages did not become a concern until quite late, and right to the end competed with significant political (and so strategic) disadvantages to striking first.⁵⁵

Rather than completely eliminating enforceable bargains and so causing war, it seems more plausible that first-strike and offensive advantages exacerbate other causes of war by narrowing the bargaining range. If for whatever reason the issues in dispute are hard to divide up, then war will be more likely the smaller the set of enforceable agreements both sides prefer to a fight. Alternatively, the problems posed by private information and incentives to misrepresent may be more intractable when the de facto bargaining range is small.⁵⁶ For example, in 1914 large perceived first-strike advantages meant that relatively few costly signals of intent were sufficient to commit both sides to war (chiefly, for Germany/Austria and Russia). Had leaders thought defense had the advantage, the set of enforceable agreements both would have preferred would have been larger, and this may have made costly signaling less likely to have destroyed the bargaining range.

I should note that scholars have sometimes portrayed the preemptive war problem differently, assuming that neither state would want to attack unilaterally but that each would want to attack if the other was expected to also. This is a coordination problem known as “stag hunt” that would seem easily resolved by communication. At any rate, it seems farfetched to think that small numbers of states (typically dyads) would have trouble reaching the efficient solution here, if coordination were really the only problem.⁵⁷

Preventive war as a commitment problem

Empirically, preventive motivations seem more prevalent and important than preemptive concerns. In his diplomatic history of Europe from 1848 to 1918, A.J.P. Taylor argued that “every war between the Great Powers [in this period] started as a preventive war, not a war of conquest.”⁵⁸ In this subsection

54. For the argument about leaders’ views on first-strike advantages in 1914, see Stephen Van Evera, “The Cult of the Offensive and the Origins of the First World War,” *International Security* 9 (Summer 1984), pp. 58–107.

55. See, for example, Trachtenberg, *History and Strategy*, p. 90.

56. This is suggested by results in Roger Myerson and Mark Satterthwaite, “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory* 29 (April 1983), pp. 265–81.

57. Schelling suggested that efficient coordination in stag hunt-like preemption problems might be prevented by a rational dynamic of “reciprocal fear of surprise attack.” See Thomas Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1960), chap. 9. Powell has argued that no such dynamic exists between rational adversaries. See Robert Powell, “Crisis Stability in the Nuclear Age,” *American Political Science Review* 83 (March 1989), pp. 61–76.

58. Taylor, *The Struggle for Mastery in Europe, 1848–1918* (London: Oxford University Press, 1954), p. 166. Carr held a similar view: “The most serious wars are fought in order to make one’s

I argue that within a rationalist framework, preventive war is properly understood as arising from a commitment problem occasioned by anarchy and briefly discuss some empirical implications of this view.⁵⁹

The theoretical framework used above is readily adapted for an analysis of the preventive war problem. Whatever their details, preventive war arguments are necessarily dynamic—they picture state leaders who think about what may happen in the future. So, we must modify the bargaining model to make it dynamic as well. Suppose state A will have the opportunity to choose the resolution of the issues in each of an infinite number of successive periods. For periods $t = 1, 2, \dots$, state A can attempt a fait accompli to revise the status quo, choosing a demand x_t . On seeing the demand x_t , state B can either acquiesce or go to war, which state A is assumed to win with probability p_t . For simplicity, assume the states are risk-neutral; that the winner of a war gets to implement its favorite issue resolution for all subsequent periods; and that the states discount future payoffs by a per-period factor $\delta \in (0, 1)$.

This model extends the one-period bargaining game considered above to an infinite-horizon case in which military power can vary over time. An important observation about the multiperiod model is that war remains a strictly inefficient outcome. If the states go to war in period t , expected payoffs from period t on are $(p_t/(1 - \delta)) - c_A$ for state A and $((1 - p_t)/(1 - \delta)) - c_B$ for state B. It is straightforward to show that there will always exist peaceful settlements in X such that both states would prefer to see one of these settlements implemented in every period from t forward rather than go to war.

The strategic dilemma is that without some third party capable of guaranteeing agreements, state A may not be able to commit itself to future foreign policy behavior that makes B prefer not to attack at some point. Consider the simple case in which A's chance of winning a war begins at p_1 and then will increase to $p_2 > p_1$ in the next period, where it will remain for all subsequent periods. Under anarchy, state A cannot commit itself not to exploit the greater bargaining leverage it will have starting in the second period. In the unique subgame perfect equilibrium, A will demand $x_t = p_2 + c_B(1 - \delta)$ in the second period and in all subsequent periods t . That is, it will choose a resolution of the issues that makes state B just willing to acquiesce, given the new distribution of military power. This means that in the first period, when state B is still relatively strong, B is choosing between going to war and acquiescing to the demand x_1 , which would yield it a total payoff of $1 - x_1 + \delta(1 - x_2)/(1 - \delta)$. The most state

own country militarily stronger or, more often, to prevent another country from becoming militarily stronger.” See E. H. Carr, *The Twenty Years’ Crisis, 1919–1939* (New York: Harper and Row, 1964), pp. 111–12.

59. To my knowledge, Van Evera is the only scholar whose treatment of preventive war analyzes at some length how issues of credible commitment intervene. The issue is raised by both Snyder and Levy. See Van Evera, “Causes of War,” pp. 62–64; Jack Snyder, “Perceptions of the Security Dilemma in 1914,” in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, eds., *Psychology and Deterrence* (Baltimore, Md.: Johns Hopkins University Press, 1985), p. 160; and Jack Levy, “Declining Power and the Preventive Motivation for War,” *World Politics* 40 (October 1987), p. 96.

A can do for B is to set x_1 to zero, so the largest possible payoff to state B for acquiescing in the first period is $1 + \delta(1 - x_2)/(1 - \delta)$. But this can be less than B's payoff for attacking in the first period, and indeed it will be whenever the following condition holds: $\delta p_2 - p_1 > c_B(1 - \delta)^2$. In words (roughly), if B's expected decline in military power is too large relative to B's costs for war, then state A's inability to commit to restrain its foreign policy demands after it gains power makes preventive attack rational for state B. Note also that A's commitment problem meshes with a parallel problem facing B. If B could commit to fight in the second period rather than accept the rising state's increased demands, then B's bargaining power would not fall in the second period, so that preventive war would be unnecessary in the first.

Several points about this rationalist analysis of preventive war are worth stressing. First, preventive war occurs here despite (and in fact partially because of) the states' agreement about relative power. Preventive war is thus another area where Blainey's argument misleads. Second, contrary to the standard formulation, the declining state attacks not because it fears being attacked in the future but because it fears the peace it will have to accept after the rival has grown stronger. To illustrate, even if Iraq had moved from Kuwait to the conquest of Saudi Arabia, invasion of the United States would not have followed. Instead, the war for Kuwait aimed to prevent the development of an oil hegemon that would have had considerable bargaining leverage due to U.S. reliance on oil.⁶⁰

Third, while preventive war arises here from states' inability to trust each other to keep to a bargain, the lack of trust is not due to states' uncertainty about present or future motivations, as in typical security-dilemma and spiral-model accounts. In my argument, states understand each other's motivations perfectly well—there is no private information—and they further understand that each would like to avoid the costs of war—they are not ineluctably greedy. Lack of trust arises here from the situation, a structure of preferences and opportunities, that gives one party an incentive to renege. For example, regardless of expectations about Saddam Hussein's future motivation or intentions, one could predict with some confidence that decreased competition among sellers of oil would have led to higher prices. My claim is not that uncertainty about intentions is unimportant in such situations—it surely is—but that commitment and informational problems are distinct mechanisms and that a rationalist preventive war argument turns crucially on a commitment problem.

Finally, the commitment problem behind preventive war may be undermined if the determinants of military power can reliably be transferred between states. In the model, the rising state can actually have an incentive to transfer away or

60. According to Hiro, President Bush's main concern at the first National Security Council meeting following the invasion of Kuwait was the potential increase in Iraq's economic leverage and its likely influence on an "already gloomy" U.S. economy. See Dilip Hiro, *Desert Shield to Desert Storm: The Second Gulf War* (London: Harper-Collins, 1992), p. 108.

otherwise limit the sources of its new strength, since by doing so it may avoid being attacked. While such transfers might seem implausible from a realist perspective, the practice of “compensation” in classical balance-of-power politics may be understood in exactly these terms: states that gained territory by war or other means were expected to (and sometimes did) allow compensating gains in order to reduce the incentive for preventive war against them.⁶¹

Preventive motivations figured in the origins of World War I and are useful to illustrate these points. One of the reasons that German leaders were willing to run serious risks of global conflict in 1914 was that they feared the consequences of further growth of Russian military power, which appeared to them to be on a dangerous upward trajectory.⁶² Even if the increase in Russian power had not led Russia to attack Austria and Germany at some point in the future—war still being a costly option—greater Russian power would have allowed St. Petersburg to pursue a more aggressive foreign policy in the Balkans and the Near East, where Austria and Russia had conflicting interests. Austrian and German leaders greatly feared the consequences of such a (pro-Slav) Russian foreign policy for the domestic stability of the Austro-Hungarian Empire, thus giving them incentives for a preventive attack on Russia.⁶³

By the argument made above, the states should in principle have had incentives to cut a multiperiod deal both sides would have preferred to preventive war. For example, fearing preventive attack by Austria and Germany, Russian leaders might have wished to have committed themselves not to push so hard in the Balkans as to endanger the Dual Monarchy. But such a deal would be so obviously unenforceable as to not be worth proposing. Leaving aside the serious monitoring difficulties, once Russia had become stronger militarily, Austria would have no choice but to acquiesce to a somewhat more aggressive Russian policy in the Balkans. And so Russia would be drawn to pursue it, regardless of its overall motivation or desire for conquest of Austria-Hungary.

While German leaders in July 1914 were willing to accept a very serious risk that Russia might go to war in support of Serbia, they seem to have hoped at the start of the crisis that Russia would accept the Austrian demarche.⁶⁴ Thus, it is hard to argue that the preventive logic itself produced the war. Rather, as is

61. On compensation, see Edward V. Gulick, *Europe's Classical Balance of Power* (New York: Norton, 1955), pp. 70–72; and Paul W. Schroeder, *The Transformation of European Politics, 1763–1848*, pp. 6–7.

62. See Trachtenberg, *History and Strategy*, pp. 56–59; Albertini, *The Origins of the War of 1914*, vol. 2, pp. 129–30; Turner, *Origins of the First World War*, chap. 4; James Joll, *The Origins of the First World War* (London: Longman, 1984), p. 87; and Van Evera, “The Cult of the Offensive and the Origins of the First World War,” pp. 79–85.

63. Samuel Williamson, “The Origins of World War I,” *Journal of Interdisciplinary History* (Spring 1988), pp. 795–818 and pp. 797–805 in particular; and D. C. B. Lieven, *Russia and the Origins of the First World War* (New York: St. Martins, 1983), pp. 38–49.

64. Jack S. Levy, “Preferences, Constraints, and Choices in July 1914,” *International Security* 15 (Winter 1990/91), pp. 234–36.

probably true for other cases in which these concerns appear, the preventive logic may have made war more likely in combination with other causes, such as private information, by making Berlin much more willing to risk war.⁶⁵ How preventive concerns impinge on international bargaining with private information is an important topic for future research.

Commitment, strategic territory, and the problem of appeasement

The objects over which states bargain frequently are themselves sources of military power. Territory is the most important example, since it may provide economic resources that can be used for the military or be strategically located, meaning that its control greatly increases a state's chances for successful attack or defense. Territory is probably also the main issue over which states fight wars.⁶⁶

In international bargaining on issues with this property, a commitment problem can operate that makes mutually preferable negotiated solutions unattainable. The problem is similar to that underlying preventive war. Here, both sides might prefer some package of territorial concessions to a fight, but if the territory in question is strategically vital or economically important, its transfer could radically increase one side's future bargaining leverage (think of the Golan Heights). In principle, one state might prefer war to the status quo but be unable to commit not to exploit the large increase in bargaining leverage it would gain from limited territorial concessions. Thus the other state might prefer war to limited concessions (appeasement), so it might appear that the issues in dispute were indivisible. But the underlying cause of war in this instance is not indivisibility per se but rather the inability of states to make credible commitments under anarchy.⁶⁷

As an example, the 1939 Winter War between Finland and the Soviet Union followed on the refusal of the Finnish government to cede some tiny islands in the Gulf of Finland that Stalin seems to have viewed as necessary for the defense of Leningrad in the event of a European war. One of the main reasons the Finns were so reluctant to grant these concessions was that they believed they could not trust Stalin not to use these advantages to pressure Finland for

65. Levy argues that preventive considerations are rarely themselves sufficient to cause war. See Levy, "Declining Power and the Preventive Motivation for War."

66. See for example Kalevi J. Holsti, *Peace and War: Armed Conflicts and International Order 1648–1989* (Cambridge: Cambridge University Press, 1991); and John Vasquez, *The War Puzzle* (Cambridge: Cambridge University Press, 1993).

67. The argument is formalized in work in progress by the author, where it is shown that the conditions under which war will occur are restrictive: the states must be unable to continuously adjust the odds of victory by dividing up and trading the land. In other words, the smallest feasible territorial transfer must produce a discontinuously large change in a state's military chances for war to be possible. See also Wagner, "Peace, War, and the Balance of Power," p. 598, on this commitment problem.

more in the future. So it is possible that Stalin's inability to commit himself not to attempt to carry out in Finland the program he had just applied in the Baltic states may have led or contributed to a costly war both sides clearly wished to avoid.⁶⁸

Conclusion

The article has developed two major claims. First, under broad conditions the fact that fighting is costly and risky implies that there should exist negotiated agreements that rationally led states in dispute would prefer to war. This claim runs directly counter to the conventional view that rational states can and often do face a situation of deadlock, in which war occurs because no mutually preferable bargain exists.

Second, essentially two mechanisms, or causal logics, explain why rationally led states are sometimes unable to locate or agree on such a bargain: (1) the combination of private information about resolve or capability and incentives to misrepresent these, and (2) states' inability, in specific circumstances, to commit to uphold a deal. Historical examples were intended to suggest that both mechanisms can claim empirical relevance.

I conclude by anticipating two criticisms. First, I am not saying that explanations for war based on irrationality or "pathological" domestic politics are less empirically relevant. Doubtless they are important, but we cannot say how so or in what measure if we have not clearly specified the causal mechanisms making for war in the "ideal" case of rational unitary states. In fact, a better understanding of what the assumption of rationality really implies for explaining war may actually raise our estimate of the importance of particular irrational and second-image factors.

For example, once the distinction is made clear, bounded rationality may appear a more important cause of disagreements about relative power than private information about military capabilities. If private information about capabilities was often a major factor influencing the odds of victory, then we would expect rational leaders to update their war estimates during international crises; a tough bargaining stand by an adversary would signal that the adversary was militarily stronger than expected. Diplomatic records should then contain evidence of leaders reasoning as follows: "The fact that the other side is not backing down means that we are probably less likely to win at war than we initially thought." I do not know of a single clear instance of this sort of updating in any international crisis, even though updating about an opponent's resolve, or willingness to fight, is very common.

68. See Max Jakobson, *The Diplomacy of the Winter War: An Account of the Russo-Finnish Conflict, 1939–1940* (Cambridge, Mass.: Harvard University Press, 1961), pp. 135–39; and Van Evera, "Causes of War," p. 63. Private information and incentives to misrepresent also caused problems in the bargaining here. See Fearon, "Threats to Use Force," chap. 3.

Second, one might argue that since both anarchy and private information plus incentives to misrepresent are constant features of international politics, neither can explain why states fail to strike a bargain preferable to war in one instance but not another. This argument is correct. But the task of specifying the causal mechanisms that explain the occurrence of war must precede the identification of factors that lead the mechanisms to produce one outcome rather than another in particular settings. That is, specific models in which commitment or information problems operate allow one to analyze how different variables (such as power shifts and cost-benefit ratios in the preventive war model) make for war in some cases rather than others.

This is the sense in which these two general mechanisms provide the foundations for a coherent rationalist or neorealist theory of war. A neorealist explanation for war shows how war could occur given the assumption of rational and unitary ("billiard ball") states, the assumption made throughout this article. Consider any particular factor argued in the literature to be a cause of war under this assumption—for example, a failure to balance power, offensive advantages, multipolarity, or shifts in relative power. My claim is that showing how any such factor could cause war between rational states requires showing how the factor can occasion an unresolvable commitment or information problem in specific empirical circumstances. Short of this, the central puzzle posed by war, its costs, has not been addressed.

Appendix

Proofs of several claims made in the text are provided here. Throughout, "equilibrium" refers to the solution concept "perfect Bayesian equilibrium."⁶⁹

CLAIM 1. *Under the conditions on $u_A(\cdot)$ and $u_B(\cdot)$ given in the text, there will exist a set of issue resolutions such that both states prefer any one of these resolutions to a war.*

Proof. It is sufficient to show that there exists an interval $[a, b] \subseteq [0, 1]$ such that for all $x \in [a, b]$, $u_A(x) \geq p - c_A$ and $u_B(1 - x) \geq 1 - p - c_B$. Choose an ϵ such that $0 < \epsilon < \min\{c_A, c_B\}$ and define $a = \max[0, p - \epsilon]$ and $b = \min[p + \epsilon, 1]$. Consider any $x' \in [a, b]$. By weak concavity, $u_A(x') \geq x'$ for all $x' \in [0, 1]$. Further, $x' > p - c_A$, since $x' \geq a \geq p - \epsilon > p - c_A$, by definitions. Thus $u_A(x') > p - c_A$. A similar argument shows that $u_B(1 - x') > 1 - p - c_B$ for all $x' \in [a, b]$. \square

Take-it-or-leave-it international bargaining

The take-it-or-leave-it game is structured as follows. Nature draws a cost of war for state B, c_B , from a cumulative distribution $H(z)$ on the nonnegative real numbers with a

69. See Drew Fudenberg and Jean Tirole, *Game Theory* (Cambridge, Mass.: MIT Press, 1991), chap. 8.

strictly positive density function $h(z)$ and a nondecreasing hazard rate $h(z)/(1 - H(z))$.⁷⁰ State B observes c_B but A does not. State A moves first, choosing a demand $x \in [0, 1]$. B observes the demand and chooses whether to fight or not. As discussed in the text, payoffs are $(p - c_A, 1 - p - c_B)$ if B fights and $(x, 1 - x)$ if B does not fight.

CLAIM 2. *The take-it-or-leave-it game has a generically unique perfect Bayesian equilibrium in which A demands x^* and B fights if and only if $c_B < x - p$. The demand x^* is defined as follows: (i) $x^* = p$ if $h(0) > 1/c_A$; (ii) $x^* = 1$ if*

$$\frac{h(1-p)}{(1-H(1-p))} < \frac{1}{(1-p+c_A)};$$

(iii) otherwise, x^* is the unique solution to

$$\frac{h(x-p)}{(1-H(x-p))} = \frac{1}{(x-p+c_A)}.$$

Moreover, the ex ante risk of war is always positive for small enough c_A greater than zero.

Proof. That B fights if and only if $c_B < x - p$ is immediately implied by subgame perfection.⁷¹ Thus in any equilibrium the probability that B fights given demand x is $\Pr(1-x < 1-p-c_B) = \Pr(c_B < x-p) = H(x-p)$. So A's expected utility for demanding x is always $u_A(x) \equiv H(x-p)(p-c_A) + (1-H(x-p))x$. Differentiation shows that $u'_A(x)$ is nonnegative when

$$\frac{h(x-p)}{(1-H(x-p))} \leq \frac{1}{(x-p+c_A)}.$$

Since the left-hand side is nondecreasing in x and the right-hand side is strictly decreasing for $x > p$, we can conclude that (i) if $h(0) > 1/c_A$ then $x^* = p$ maximizes $u_A(x)$; (ii) if

$$\frac{h(1-p)}{(1-H(1-p))} < \frac{1}{(1-p+c_A)}$$

then $x^* = 1$ maximizes $u_A(x)$; (3) any $x^* \in [p, 1]$ that solves

$$\frac{h(x-p)}{(1-H(x-p))} = \frac{1}{(x-p+c_A)}$$

will be a unique maximum of $u_A(x)$. Since the ex ante probability of war is $H(x^* - p)$, it follows that only in case (i) can this equal zero, and for small enough c_A case (i) cannot obtain. \square

70. This condition is satisfied for a broad range of distributions. See Fudenberg and Tirole, *Game Theory*, p. 267.

71. The assumption that type $c_B = x - p$ chooses not to fight is immaterial.

*Take-it-or-leave-it international bargaining
with communication (cheap talk)*

I now modify the take-it-or-leave-it game by allowing state B to choose (after observing its costs for war) an announcement f from a large but finite set of possible speeches, F . After the announcement, the game proceeds exactly as before with identical payoffs. To avoid some measure-theoretic complications, I will consider a finite approximation of the game analyzed above. Nature chooses state B's type from the set $T = \{c_0, c_1, c_2, \dots, c_i, \dots, c_n\}$, $n \gg 0$, according to a discrete prior distribution $h(\cdot)$, where $h(c_i) = \Pr(c_B = c_i)$ and $h(c_i) > 0$ for all $c_i \in T$. The elements of T satisfy $c_0 = 0$ and (by convention) $c_i < c_j$ for all $i < j$, where $i, j \in N \equiv [0, 1, 2, \dots, n]$. Let $H(c_i) \equiv \sum_{j=0}^{i-1} h(c_j)$ denote the prior probability that B's cost is strictly less than c_i , letting $H(c_0) = 0$.

In the game with talk, state B has a *message strategy* that gives a probability distribution on all possible messages in F for each type in the set T . Let $m(f, c_i)$ be the probability that type c_i announces speech f in a given equilibrium. State A's *demand strategy* now associates with each f a probability distribution over $[0, 1]$. Let $x(f)$ be the demand made by A on hearing f whenever A does not mix given the announcement. In a given equilibrium of the game with talk, state A will form posterior beliefs about B's type for each message f . Let these be denoted $h(c_i, f) \equiv \Pr(c_B = c_i | f)$, with $H(c_i, f) \equiv \sum_{j=0}^{i-1} h(c_j, f)$.

For convenience, I will assume that if B is indifferent between rejecting and accepting a demand x , B accepts for sure.

PROPOSITION. *If x^* is the unique equilibrium demand in the game without talk, then in any equilibrium of the game with talk in which state A uses a pure strategy, (1) state A demands x^* regardless of the announcement; and (2) the ex ante risk of war is the same as in the game without talk.*

The proof follows from several lemmas.

LEMMA 1. *In any equilibrium A's payoff is at least p , and in no equilibrium will A respond to any message with a demand that is sure to yield war.*

Proof of Lemma 1. If A sets $x = p$, all types of B accept for sure in any equilibrium (by subgame perfection), so A can assure itself p in this way. If in some equilibrium, A chooses x following a message f such that war certainly follows, then A receives $p - c_A$, but then A could deviate to $x' = p$ and do strictly better. \square

LEMMA 2. *In any equilibrium, the demand x is in the support of A's demand strategy given a message f only if there exists a $c_i \in T$ such that $x = p + c_i$.*

Proof of Lemma 2. If not, then A might on hearing f choose a demand x' such that $p + c_i < x' < p + c_{i+1}$ for some i . But then A could increase its payoff on hearing f by deviating to $x'' = p + c_{i+1}$, since doing so has no effect on the risk of war. \square

LEMMA 3. *In any equilibrium in which A does not mix, $x(f) = k$, a constant, for all messages $f \in F'$, where F' is the set of messages sent with positive probability in the given equilibrium.*

Proof of Lemma 3. Let $T_f \equiv \{c_i : m(f, c_i) > 0\}$. Suppose to the contrary that in some equilibrium there exist two distinct messages f and f' such that $x(f) < x(f')$. By Lemma 1, both demands must be accepted with positive probability, implying that there are types $c_i \in T_f$ such that $1 - x(f) \geq 1 - p - c_i$ and types $c_j \in T_{f'}$ such that $1 - x(f') \geq 1 - p - c_j$. But then any such $c_j \in T_{f'}$ can do strictly better by deviating to f , which gives it $1 - x(f) > 1 - x(f')$. But this implies that $x(f')$ is certainly rejected, contradicting Lemma 1. \square

Proof of the proposition. Suppose $x^* = p + c_k$ is the unique equilibrium demand in the game without talk. Then c_k is the only element of T such that for all $j \in N$,

$$H(c_k)(p - c_A) + (1 - H(c_k))(p + c_k) \geq H(c_j)(p - c_A) + (1 - H(c_j))(p + c_j). \quad (1)$$

Suppose to the contrary of the proposition that in the game with talk there is some other demand $x' = p + c_l$, $l \neq k$, such that A demands x' on hearing any message $f \in F'$ (by Lemma 3, any equilibrium without mixing by A must have this form). Then it must be that for each $f \in F'$ and for all $j \in N$,

$$H(c_l, f)(p - c_A) + (1 - H(c_l, f))(p + c_l) \geq$$

$$H(c_j, f)(p - c_A) + (1 - H(c_j, f))(p + c_j). \quad (2)$$

By Bayes's rule, $h(c_j, f) = h(c_j)m(f, c_j)/\Pr(f)$ and $H(c_j, f) = (1/\Pr(f)) \sum_{i=0}^{l-1} h(c_i, f)$, where $\Pr(f) = \sum_{c \in T} h(c)m(f, c)$ is the probability that B chooses f in the equilibrium. Substitution into equation (2) and multiplication of both sides by $\Pr(f)$ yields, for all $j \in N$,

$$\begin{aligned} & \left[\sum_{i=0}^{l-1} h(c_i)m(f, c_i) \right] (p - c_A) + \left[\Pr(f) - \sum_{i=0}^{l-1} h(c_i)m(f, c_i) \right] (p + c_l) \\ & \geq \left[\sum_{i=0}^{j-1} h(c_i)m(f, c_i) \right] (p - c_A) + \left[\Pr(f) - \sum_{i=0}^{j-1} h(c_i)m(f, c_i) \right] (p + c_j). \end{aligned} \quad (3)$$

Since equation (3) holds for each $f \in F'$, we can sum both sides over all $f \in F'$ and the inequalities still hold. Thus, for all $j \in N$,

$$\begin{aligned} & \left[\sum_{i=0}^{l-1} h(c_i) \sum_{f \in F'} m(f, c_i) \right] (p - c_A) + \left[\sum_{f \in F'} \Pr(f) - \sum_{i=0}^{l-1} h(c_i) \sum_{f \in F'} m(f, c_i) \right] (p + c_l) \\ & \geq \left[\sum_{i=0}^{j-1} h(c_i) \sum_{f \in F'} m(f, c_i) \right] (p - c_A) + \left[\sum_{f \in F'} \Pr(f) - \sum_{i=0}^{j-1} h(c_i) \sum_{f \in F'} m(f, c_i) \right] (p + c_j). \end{aligned} \quad (4)$$

Since $\sum_{f \in F'} \Pr(f) = 1$ and $\sum_{f \in F'} m(f, c_i) = 1$, equation (4) simplifies to yield, for all $j \in N$,

$$H(c_l)(p - c_A) + (1 - H(c_l))(p + c_l) \geq H(c_j)(p - c_A) + (1 - H(c_j))(p + c_j). \quad (5)$$

But this contradicts equation (1), the hypothesis that the game without talk has a unique equilibrium, $x^* = p + c_k$ where $c_k \neq c_l$. This proves the first part of the proposition. To see the second part, notice that if A demands x^* after any message sent with positive probability, then all types $c_j < x^* - p$ will fight and all types $c_j \geq x^* - p$ will accept the demand, so the *ex ante* risk of war is just $H(x^* - p)$, as in the game without talk. \square

Remarks. (1) I conjecture that if there is unique equilibrium demand x^* in the game without talk, then in no equilibrium of the game with talk will A ever mix over demands, so that the proposition should extend to all equilibria of the game with talk. While I have found no counterexamples, I have not been able to demonstrate conclusively that A cannot mix in some equilibrium. (2) Cheap talk can indeed matter in this game in a very limited way—equilibria may exist in which both A's beliefs and the risk of war differ depending on the message sent. However, as shown in the proposition, in any such equilibrium the variation in A's beliefs with different messages affects neither A's behavior nor the *ex ante* (i.e., premessage) risk of war.