Journal of Theoretical Politics 16(2): 107–141
DOI: 10.1177/0951629804041117

RECONCILING RATIONALITY WITH DETERRENCE

A RE-EXAMINATION OF THE LOGICAL FOUNDATIONS OF DETERRENCE THEORY

Frank C. Zagare

ABSTRACT

This article argues that classical (or rational) deterrence theory is logically inconsistent, empirically inaccurate and prescriptively deficient. In its stead it offers an alternative theoretical framework – perfect deterrence theory – that makes consistent use of the rationality postulate and is in accord with the empirical literature of deterrence. Perfect deterrence theory's axiomatic base, its empirical expectations and its most significant policy prescriptions are highlighted and contrasted with those of classical deterrence theory. The theory's implications for current policy debates about a national missile defense system, arms control, US policy toward China and Russia, and inter-state negotiations in general, are discussed.

KEY WORDS • arms control • deterrence and rationality • US foreign policy

Classical deterrence theory, or what Glaser (1989) calls the 'punitive retaliation school', constitutes the conventional wisdom in international relations scholarship. An intellectual descendant of balance of power theory, classical (or rational) deterrence theory purports to explain 'the remarkable stability' of the post-war era. More specifically, classical deterrence theory holds that the absence of a superpower war in the aftermath of the Second World War can be traced directly to the existence of a bipolar system *and* the high costs of nuclear conflict (Waltz, 1993: 44).

As numerous analysts (e.g. Trachtenberg, 1991) have noted, though, classical deterrence theory is riddled with logical inconsistencies.[1] The inconsistencies clearly circumscribe the theory's explanatory power and call into question the wisdom of its policy prescriptions.[2] Their existence also helps

1. Or, as some have put it, by puzzles, dilemmas and paradoxes.

2. Walt (1999a) downplays the significance of logical consistency as a criterion for evaluating theories, claiming that both empirical validity and originality are more important. A number of (formal) theorists have taken issue with Walt's reasoning (Bueno de Mesquita and Morrow, 1999; Martin, 1999; Niou and Ordeshook, 1999; Powell, 1999; Zagare, 1999). See Walt (1999b) for the counter-response.

to explain why contradictory claims about the relationship between deterrent threats and the likelihood of inter-state conflict are frequently made by theorists and policy-makers working within the same paradigm and ostensibly sharing a common set of assumptions. As Gaddis (1997: 101) has tactfully observed, 'logic, in this field, [is] not what it [is] elsewhere'.

For example, in an article in *The Washington Post*, Charles Krauthammer (2001) denigrates the claim, made by opponents of the space-based missile defense system proposed by the Bush administration, that an anti-ballistic missile (ABM) system might prompt a Russian first strike. 'It would be', he writes, 'the most massive, genocidal and unprovoked act of war in the history of the human race'. Even at the height of the Cold War, he goes on to argue, the possibility of a pre-emptive Soviet attack was only 'minimally plausible'. On this count, Krauthammer may well be correct. But if an out-of-the-blue attack by a sworn enemy of the United States was never a real possibility, why then is a costly and unproven missile defense system needed now? If the Soviet Union then (and Russia today) was almost certainly deterred by the threat of a massive retaliatory strike, as Krauthammer implies, why would a much smaller and more vulnerable state like North Korea even contemplate a nuclear attack on the United States?

Similarly, Secretary of Defense Donald Rumsfeld has claimed that the high probability of a retaliatory strike is unlikely to deter crazy 'rogue' states like Libya or Iraq (under Saddam Hussein). At the same time, Rumsfeld also asserted that a space-based defense need not be completely effective for it to dissuade a nuclear attack. In summarizing Rumsfeld's argument, Thomas L. Friedman (2001) highlights the inconsistency: 'In short, our perfect missiles that will destroy any rogue state with 100 percent accuracy won't deter them, but our imperfect missile shield, which may have as many holes as a Swiss cheese, will deter them.'

One purpose of this article is to situate the source of contradictions like these in classical deterrence theory's axiomatic base.[3] Its primary purpose, however, is to sketch an alternative theoretical framework, called *perfect deterrence theory*, which is, in fact, grounded in strict logic, and to map out its most important policy implications.

All of which is not to suggest that the problems with classical deterrence theory are restricted to its logical structure. The contention here is that the

3. Powell (1985: 75) correctly points out that contradictory conclusions drawn from a theory suggests *either* 'a fundamental weakness in the theory or that those using the theory do not fully appreciate it'. The argument here is not that Krauthammer's and Rumsfeld's contradictory policy statements are necessarily traceable to classical deterrence theory's inconsistent axiomatic base. Rather the point is that the shaky theoretical foundations of classical deterrence theory make contradictory policy pronouncements such as Krauthammer's or Rumsfeld's almost inevitable.

standard approach to deterrence is also empirically inaccurate (to the extent that its major propositions can be clearly identified).[4] In other words, even if one agrees with Walt (1999a) that empirical accuracy takes precedence over logical consistency, classical deterrence theory falls short of the mark. The empirical deficiencies of classical deterrence theory are also discussed here.

1. Classical Deterrence Theory

Classical deterrence theory can conveniently be divided into two distinct, yet compatible, formulations: *structural* and *decision-theoretic* deterrence theory. Even though these strands in the literature focus on different units of analysis, the assumptions they make, the conclusions they reach and the policy prescriptions they draw are essentially the same.[5]

For structural deterrence theorists, the international system constitutes the principal unit of analysis. The system itself is anarchic: there is no overarching authority to enforce agreements. This 'self-help' system is composed of undifferentiated units (i.e. states)[6] that are rational and egotistical. The units are driven either by their nature to maximize power (Morgenthau, 1973) or by their environment to maximize security (Waltz, 1979).

Structural deterrence theorists hold that the key to international stability lies in the distribution of power in the international system and the absolute cost of war. Although there is controversy among these theorists about

4. One problem with demonstrating this contention conclusively is the existence of numerous contradictory propositions, and policy prescriptions based on them, in the literature of classical deterrence theory. Perhaps the clearest example concerns proliferation. Some of the most prominent classical deterrence theorists favor the selective proliferation of nuclear weapons. Many others, however, argue against this policy.

5. For a more complete description of the axioms, tenets and deficiencies of classical deterrence theory than can be provided here, see Zagare (1996a).

6. Legro and Moravcik, (1999: 13) argue that the assumption of undifferentiated actors with 'fixed and uniformly conflictual' preferences distinguishes realism and, by extension, classical deterrence theory, from other paradigms. Waltz (1979) clearly assumes that states are undifferentiated (or like) units. But as Wohlforth (2000: 183) points out, Waltz also asserts that 'the aims of states may be endlessly varied'. Surely there is some ambiguity (if not a contradiction) here. Waltz (1979: 105) readily admits that states 'differ vastly in their capabilities'. In what real sense, then, can the units be undifferentiated if their critical preferences are also assumed to vary? How can *all* other states be potential threats, as Mearsheimer (1990: 12) asserts, if only some states 'think and act in terms of interests defined as power'(Morgenthau, 1973: 5)? And what are we to make of Waltz's (1993: 47) comment that 'our conviction that the United States was the status quo and the Soviet Union the interventionist power distorted our view of reality' if states are not uniformly motivated? Walt (1999a: 17) matter of factly observes that Waltz's theory contains contradictions.

precisely why,[7] structural deterrence theorists, by definition, contend that balanced bipolar systems are inherently more stable than multipolar systems. Nuclear weapons, which dramatically increase the cost of war, only reinforce the stability of parity relationships. Thus, the 'long peace' of the post-war period (Gaddis, 1986) is easy for structuralists to explain: war becomes unthinkable (i.e. is irrational) once power is balanced and the cost of war is exorbitant.

Structural deterrence theorists locate the cause of inter-state conflict in asymmetric power relationships. This is especially so when the cost of conflict is low (Waltz, 1993: 77). In general, structural deterrence theorists see a monotonic relationship between the cost and the probability of war. As Mearsheimer (1990: 19) puts it, 'the more horrible the prospect of war, the less likely it is to occur'.

Given all this, it is not difficult to understand why many structural deterrence theorists argue that quantitative arms races help prevent war (additional weapons, they hold, increase the cost of war),[8] why some contend that qualitative arms races and defensive weapons are destabilizing (because they believe that certain types of weapons will reduce war costs for one or both sides)[9] and why others are in favor of managed nuclear proliferation (again, because nuclear weapons make war more costly).[10] Given the exceedingly low probability of war between nuclear equals, structural deterrence theorists conclude that the gravest threat to peace is an accident or a mishap. In other words, for structural deterrence theorists, the probability of a premeditated (or rational) nuclear war is virtually zero (e.g. Intriligator and Brito, 1981: 256; Waltz, 1990: 740).

Although structural deterrence theory is consistent with the absence of a superpower conflict during certain periods of the Cold War, it is inconsistent with other pertinent empirical realities. As Jervis (1985: 6) notes, 'many events present unexpected puzzles for standard deterrence theory'. To wit, structural deterrence theory is inconsistent with the fact that most major power wars have been waged under parity conditions or with the observation that power imbalances are poor predictors of inter-state conflict.[11] More specifically, unless they make ad hoc arguments that contravene the theory's axiomatic base, structural deterrence theorists are hard put to explain the absence of war before the Soviet Union achieved 'essential equivalence'

7. Compare, for example, Waltz (1964: 882–6); Gaddis (1986: 105–10); and Mearsheimer (1990: 14).

8. See, for example, Gray (1974).

9. See, *inter alia*, Jervis (1978), Scoville (1981) and Van Evera (1984).

10. Among classical deterrence theorists who argue for the selective proliferation of nuclear weapons are Mearsheimer (1990), Waltz (1981) and Van Evera (1990/91).

11. For the relevant citations, see Zagare and Kilgour (2000: 24–6)

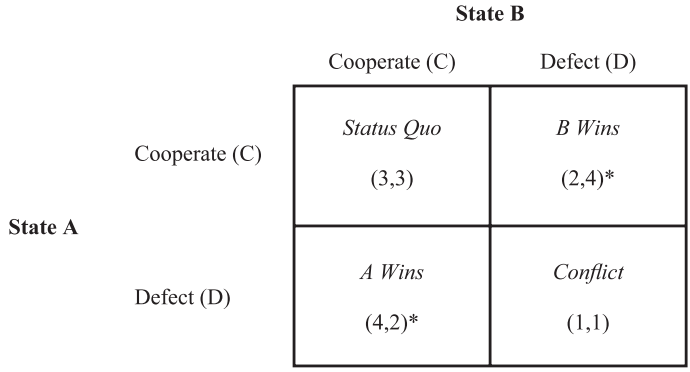

Figure 1. Chicken. Key: (x, y) = payoff to state A, payoff to state B; 4 = best; 3 = next-best; 2 = next-worst; 1 = worst and * = Nash equilibrium.

with the United States during the 1970s. As Waltz (1993: 47; 2000: 13) himself suggests, to explain the absence of a US–Soviet war up to the advent of nuclear parity by claiming that the United States was either a status quo power or a self-deterred democracy unwilling to violate moral precepts contradicts the assumption that egotistical, rational and undifferentiated units populate the inter-state system. Structural deterrence theorists also have difficulty explaining why the contentious relationship of the former Soviet Union and China did not erupt into an all-out conflict, why the United States has not attempted to invade Canada since 1812 or why states in general fail to jump through 'windows of opportunity' (Lebow, 1984).

Decision-theoretic deterrence theorists – who focus on the interplay of outcomes, preferences and rational choices – begin where structural deterrence theorists leave off.[12] In developing either formal or informal rational choice models based on the payoff structure of the game of 'Chicken' (see Figure 1), early decision-theoretic deterrence theorists like Schelling (1960, 1966), Ellsberg (1959) and Kahn (1962) or later theorists like Powell (1987) and Nalebuff (1986) fully embrace *the* central conclusion of structural deterrence theory: that war in the nuclear age is 'irrational'. In Chicken, war (or conflict) is the worst possible outcome for both players. In consequence,

12. Young (1975) calls this most influential approach to deterrence 'manipulative bargaining theory'. Rapoport (1968) pejoratively refers to decision-theoretic deterrence theorists as 'neo-Clauswitzians'. Danilovic (2002) labels the genre 'commitment theory'. The term 'decision-theoretic deterrence theory' is used here in order to include both the seminal first wave of expected utility models of deterrence and those subsequent game-theoretic refinements that share the modal assumptions outlined later.

conflict can never be consistent with rational contingent decision-making. Thus, Jervis (1985: 19) is correct: within the axiomatic confines of classical deterrence theory, 'a rational strategy for the employment of nuclear weapons is a contradiction in terms'.

The reason that a mutually worst outcome is unquestionably irrational is that it can never be part of a (pure strategy) equilibrium outcome in any game with strict preference rankings over outcomes.[13] And only equilibrium outcomes are consistent with rational choices by all the players in a game.

It is clear that a mutually worst outcome can never be part of a (pure strategy) equilibrium since either player can always achieve a more preferred outcome simply by changing its strategy choice.[14] Thus, by assuming that conflict is the worst outcome for both players, decision-theoretic deterrence theorists, *perforce*, presume war to be irrational. In so doing, they take as axiomatic a critical deduction of structural deterrence theory.

By *assuming* that war is irrational, decision-theoretic deterrence theorists *presuppose* the world envisioned by structural deterrence theorists. Hence, decision-theoretic deterrence theory can be interpreted as a micro- (or unit-)level extension of structural deterrence theory, in effect mapping out what optimal strategic behavior would be in the world envisioned by structural deterrence theorists. This means that the conclusions of these decision theorists also have important implications for the empirical accuracy and the logical consistency of structural deterrence theory.

The descriptions and prescriptions of decision-theoretic deterrence theory are well known and will not be rehearsed in detail here.[15] Suffice it to say that, *inter alia*, statesmen have been counseled to seize the initiative by making an 'irrevocable commitment' to a hard line strategy, to avoid defeat by 'linking' one issue to another, to make an opponent's concession more likely by

13. *Conflict* is part of a mixed strategy equilibrium in Chicken. O'Neill (1992: 471–2) argues persuasively that this equilibrium fails as a normative device.

14. In a static (strategic-form) two-person game like Chicken, the standard equilibrium concept is due to Nash (1951). A strategy pair is a *Nash equilibrium* if no player could achieve a better outcome by switching, unilaterally, to another strategy. In a dynamic (extensive-form) game, where the players' choices are sometimes contingent, the central equilibrium concept is *subgame-perfect* (Selten, 1975). Nash equilibria exist in the dynamic context, but they may be based on incredible threats (i.e. on threats of irrational choice), whereas subgame-perfect equilibria require the players to plan to choose rationally at every node of the game tree (Morrow, 1994: 127–8). Nash and subgame-perfect equilibria are the accepted measures of rational behavior in games of *complete* information, in which each player is fully informed about the preferences of its opponent. In games of *incomplete* information in which, for instance, at least one player is uncertain about the other's preferences, rational choices are associated with *Bayesian Nash equilibria* (in static games) and with *perfect Bayesian equilibria* (in dynamic games). See Gibbons (1992) for a discussion.

15. Snyder (1972: 229–31) provides a comprehensive listing.

making conflict more costly[16] or to even feign 'irrationality' in order to force an opponent to concede during a crisis.

While provocative, these stratagems are of dubious empirical validity. In a recent review of the deterrence literature, Huth (1999: 74) finds that 'early arguments about the strategic advantages of the manipulation of risk and commitment strategies have not been fully supported by empirical research' (see also Danilovic, 2001, 2002). Making the same point, Betts (1987: 30) observes that 'the view that apparent recklessness and irrevocable commitment are more effective is usually more comfortable to pure strategists than to presidents'. Perhaps Jervis (1988: 80) put it best: 'although we often model superpower relations as a game of Chicken, in fact the United States and USSR have not behaved like reckless teenagers'.

Like structural deterrence theorists, decision-theoretic deterrence theorists are hard put to explain, in a logically consistent way, the long peace that followed the Second World War. To see why, consider now the Rudimentary Asymmetric Deterrence Game depicted in Figure 2. In this, perhaps the simplest deterrence situation that one can imagine, there are two players – state A and state B – and only three outcomes: *Status Quo*, *A Wins* and *Conflict*.

The Rudimentary Asymmetric Deterrence Game is a model of an asymmetric or one-sided deterrence situation: state B wishes to deter state A but not the other way around. Thus, in the extensive-form game depicted in Figure 2, state A begins play at decision node 1 by deciding whether to *cooperate* (C) and accept the status quo or to *defect* (D) and demand its alteration. If state A chooses C, the game ends and the outcome is the *status quo*. But if state A defects, state B must decide at decision node 2 whether to concede (C) the issue – in which case the outcome is *A Wins* – or deny (D) the demand and precipitate *Conflict*.

For the sake of argument, let us accept, for the moment, two core assumptions of classical deterrence theory: (1) that *Conflict* is the worst possible outcome; and (2) that the players are instrumentally rational. Next, we ask what instrumentally rational players would do when presented with the choices in the Rudimentary Asymmetric Deterrence Game.

To answer this question, the game tree of Figure 2 is examined using *backwards induction*. To apply this procedure, one works backwards up the game tree and determines, first, what an instrumentally rational state B would do at decision node 2; then, using this information, the rational choice of state A at node 1 is specified.

16. Both structural and decision-theoretic deterrence theorists, therefore, recommend policies that raise the cost of conflict.

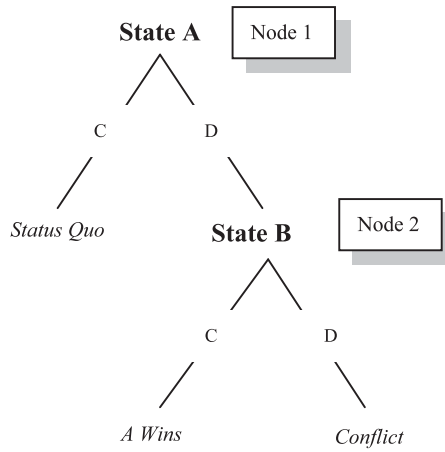

Figure 2. Rudimentary Asymmetric Deterrence Game

At node 2, state B is faced with a choice between conceding (i.e. choosing C), which brings about outcome *A Wins*, and denying state A's demand (i.e. choosing D), which brings about *Conflict*. But if *Conflict* is assumed to be the *worst* possible outcome, an instrumentally rational state B can *only* choose to concede since, by assumption, *A Wins* is the more preferred outcome.

Given that an instrumentally rational state B will choose to concede at node 2, what should state A do at node 1? State A's choice is either to cooperate, in which case the outcome will be the *Status Quo*, or to defect, in which case the outcome will be *A Wins* – because an instrumentally rational state B will choose to concede at node 2. If state A prefers *A Wins* to the *Status Quo*, that is if it has an incentive to upset the *Status Quo* and, therefore, needs to be deterred, it will rationally choose D. In other words, given these two core assumptions of classical deterrence theory, the *Status Quo* is unstable and deterrence rationally fails. Or, to put it in a slightly different way, the theory's assumptions are logically incompatible with the possibility of deterrence success. Nonetheless, classical deterrence theorists contend that bipolar nuclear relationships are exceedingly stable.

The same conclusion could be drawn as well from an analysis of Chicken. Since the *Status Quo* is also not part of a (pure strategy Nash) equilibrium in Chicken, policies that unconditionally support the status quo in this game are also incompatible with rational choice. Thus, it remains true that the assumptions that delineate decision-theoretic deterrence theory are inconsistent with the persistence of peace throughout the Cold War. Moreover, since decision-theoretic deterrence theory is axiomatically derivative from structural deterrence theory, this deficiency of decision-theoretic deter-

rence theory casts doubt on the latter approach to deterrence as well. As Van Gelder (1989: 159) observes, the lack of congruence between the assumptions and the conclusions of classical deterrence theory 'threatens the very foundations of nuclear deterrence as a rational strategy'.

For obvious reasons, decision-theoretic deterrence theorists have attempted mightily to reconcile the canons of rational choice with empirical reality (i.e. the absence of a superpower war) within the confines of the paradigm. But they have been unable to square this circle (Zagare, 1990). For example, Gauthier (1984) constructs a rational choice model based on Chicken in which the players maximize their utility by choosing *not* to upset the status quo. The policy, however, is inconsistent with Selten's (1975) perfectness criterion, meaning that it is rooted in irrational choices and incredible threats. Powell's (1987) model overcomes this limitation but at the cost of special assumptions. First, he assumes that all-out attacks are *always* reciprocated.[17] One direct consequence of this assumption is that first strikes can never be rational in the model – because they always lead to a player's worst outcome, conflict. Hence, at the strategic level, Powell's model postulates, rather than derives, stability.[18] Second, he assumes that irrational threats are executed probabilistically by a disinterested third party, Nature. As even Powell (1987: 725) admits, 'relying on Nature to impose the irrational sanctions does not really solve the credibility problem' and calls into question the logical foundations of decision-theoretic deterrence theory. Finally, in Howard's (1971) alternative game-theoretic framework, the status quo emerges as an equilibrium in the 'metagame' of Chicken. However, since this 'metaequilibrium' is based on an incredible threat (Harsanyi, 1974), it too fails to provide a rational basis for explaining the absence of a superpower war.[19]

To summarize briefly: classical (or rational) deterrence theory is riddled by empirical inaccuracies and logical inconsistencies. Even if its 'originality' is stipulated, classical deterrence theory clearly fails to satisfy even the relaxed standards Walt (1999a) offers for evaluating theory. It is no small wonder then that one prominent theorist concludes that many of the theory's policy prescriptions are 'contrary to common sense' (Jervis, 1979: 292) or that another finds them to be just plain 'bizarre' (Rapoport, 1992).

17. This is a common assumption in the formal literature of deterrence. For other examples, see Fearon (1994b: 590); Bueno de Mesquita et al. (1997: 17); or Kydd (1997: 379).

18. The same could be said of the mainstream strategic literature. As Powell (1985: 83) notes elsewhere, 'in the theory [of deterrence that takes invulnerable second-strike forces as a given] the risk of an unrestricted nuclear attack is assumed away'.

19. For a more detailed examination of these and related unsuccessful attempts to reconcile the absence of a superpower conflict with the axioms of decision-theoretic deterrence theory, see Zagare (1990) and Zagare and Kilgour (2000: Ch. 2).

Within the confines of the theory's axiomatic base, the only way to explain the 'remarkable stability' of the Cold War period is to assume, simultaneously, that the players are at once rational and irrational. The players are rational when they are being deterred – presumably because they fear the costs of initiating a conflict – but they must also be presumed to be irrational when they are deterring an opponent and threatening to retaliate – presumably because they do not fear the cost of conflict.[20] As Brodie (1959: 293) has observed, 'for the sake of deterrence before hostilities, the enemy must expect us to be vindictive and irrational if he attacks us'. Or as Achen (1987: 92) puts it, 'far from leaning too heavily on rational choice postulates, ''rational deterrence theory'' necessarily assumes that nations are not always self-interestedly rational'.[21]

Clearly, players who can be both rational and irrational can also fail to be deterred by a perfectly accurate retaliatory threat and, at the same time, can be deterred by an imperfect missile shield. As well, a first strike against the United States by a potent hostile state like the Soviet Union can be judged to be only 'minimally plausible', while a similar attack by a much smaller and much more vulnerable rogue state is justification enough to develop a costly, provocative and unproven ABM system. For classical deterrence theorists, all things are possible once logical consistency is abandoned.[22]

2. Perfect Deterrence Theory

In this section, an alternative approach to deterrence – perfect deterrence theory – is outlined and its major policy implications highlighted (Zagare

20. Powell's (1987) model is an exception to this statement. The states in Powell's model always make rational choices. To maintain logical consistency, however, Powell must also assume that irrational threats are carried out probabilistically by Nature or some other autonomous force.

21. See also Powell (1985: 80).

22. Trachtenberg (1991: 32) traces the logical problems of classical deterrence theory to two 'fundamentally inconsistent' ideas: (1) that nuclear war is absurd and (2) that the threat of nuclear war can be used for political advantage. By uncritically accepting the Chicken analogy, decision-theoretic deterrence theorists implicitly accept the absurdity of nuclear war. Yet the coercive bargaining techniques they championed rest on the supposition that nuclear threats could serve a political purpose. Trachtenberg (1991: 4) finds that classical deterrence theorists like Brodie and Schelling 'were attracted to both approaches, often at the same time'. Thus it is not at all surprising that he also finds a 'pervasive' and 'fundamental' tension between these conflicting notions (1991: 32) in the strategic discourse of the 1950s and early 1960s. Trachtenberg's penetrating observation still holds. Witness Krauthammer's (2002) pronouncement that 'the iron law of the nuclear age is this: nuclear weapons are instruments of madness; their actual use would be a descent into madness, but the threat to use them is not madness. On the contrary, it is exceedingly logical'.

Table 1. Classical Deterrence Theory and Perfect Deterrence Theory Compared

	Classical deterrence theory	Perfect deterrence theory
Assumptions		
States	Undifferentiated	May be differentiated
Actors	Egotistical	Egotistical
Credibility	Constant	Not fixed
Rationality	Sometimes	Always
Irrational threats	May be executed	Never executed
Theoretical characteristics		
Logically consistent	No	Yes
Empirical validity	Uncertain	Consistent with extant empirical literature
Propositions		
Status quo	Unimportant/ignored	Significant
Strategic deterrence	Robust/all but certain	Fragile/contingent
Relationship between conflict costs and deterrence success	Strictly positive and monotonic	Non-monotonic
Asymmetric power relationships	Unstable	Potentially very stable
Parity relationships	Very stable	Potentially unstable
Capability	Sufficient for deterrence success	Necessary, but not sufficient, for deterrence success
Limited conflicts and escalation spirals	Unexplained	Placed in theoretical context
Policies		
Overkill capability	Supports	Opposes
Minimum deterrence	Opposes	Supports
'Significant' arms reductions	Opposes	Supports
Proliferation	Supports	Opposes
Negotiating stances	Coercive, based on increasing war costs and inflexible bargaining tactics	Conditionally cooperative, based on reciprocity

and Kilgour, 2000). Unlike classical deterrence theory, this 'common sense' approach to deterrence makes consistent use of the rationality postulate. A number of 'plausibility probes' suggests that it is also consistent with the empirical record (Quackenbush and Zagare, 2001; Quackenbush, 2003; Senese and Quackenbush, 2003). Table 1 summarizes the most important differences between perfect deterrence theory and classical deterrence theory.

Some of the axioms of perfect deterrence theory and classical deterrence theory are the same. For example, states are assumed to be rational and

egotistical. States, however, are not *necessarily* assumed to be undifferentiated (see later). There are, as well, a number of other critical differences between the two theories. First, in perfect deterrence theory, an outright attack is not assumed to culminate automatically in war.[23] The players (i.e. states) always have an opportunity *not* to retaliate. Since response options are not necessarily executed, the possibility of an unrestricted first strike is not assumed away. Second, in perfect deterrence theory, only players can execute deterrent threats. Thus, in perfect deterrence theory, an opponent's threat, and not some impersonal force, is the principal source of the risks run by the players. Third, only rational (i.e. credible) threats can be carried out. This stricture ensures that the deductions of perfect deterrence theory remain consistent with the rationality postulate, in general, and with Selten's perfectness criterion, in particular[24] – hence, the theory's name. Finally, while credibility is fixed and constant in the formal renditions of classical deterrence theory, it is measured on a continuum in perfect deterrence theory.

The last feature of perfect deterrence theory requires special comment. Note that in classical deterrence theory, deterrent threats are, perforce, *always* presumed to be incredible. This reflects the fact that decision-theoretic deterrence theorists, by definition, take conflict to be the mutually worst outcome.[25] If credibility is equated with instrumental rationality, as it is in both the formal and the wider strategic literature (see Section 3.2), a threat that leads to a threatener's worst outcome can *never* be credible – because it can never be rational to carry out such a threat (Zagare, 1990; Morrow, 2000). And threats that are *always* incredible can never vary.[26]

It is important to point out that as long as credibility is considered a constant, no logical relationship between it and stable deterrence can ever be established. But even if credibility were allowed to vary in classical deterrence theory, a fundamental problem would still exist: as long as incredible threats can be executed, the very possibility of exploring the theoretical

23. See footnotes 17 and 18.

24. See footnote 14.

25. Structural deterrence theorists are quite comfortable with this assumption. See, for example, Waltz (1993: 53–54).

26. As Walt (1999b: 123) correctly points out, classical deterrence theorists (e.g. Schelling, 1966) have, in fact, speculated about mechanisms that enhance credibility or circumstances that make threats more or less credible. But these discussions should be kept separate from the theory that gives rise to them. Many of the prescriptions developed by decision-theoretic deterrence theorists require states to credibly (i.e. rationally) threaten war (or conflict) in order to deter aggression. But in the brinkmanship models that underpin these prescriptions, it is always irrational to execute a deterrent threat. In other words, it is logically inconsistent to treat patently self-abnegating (i.e. absurd) threats as rational, as credible or as variable (Trachtenberg, 1991: 32).

relationship between the credibility of threats and the operation of deterrence is precluded.

By contrast, in perfect deterrence theory, credibility can indeed vary. As well, irrational threats cannot be carried out. Thus, unlike classical deterrence theory, perfect deterrence theory is well situated to explore the logical connection between threat credibility and the dynamics of dyadic inter-state relationships. It is, therefore, a more general theory.

At first blush, the axiomatic differences between classical deterrence theory and perfect deterrence theory might appear to be minor, perhaps even insignificant. But, as illustrated in the films *Sliding Doors* and *Run Lola Run* and as will be shown later, small differences in initial assumptions can have important theoretical consequences and significant policy implications.[27]

The major conclusions of perfect deterrence theory are drawn from an examination of three simple (incomplete information) deterrence models. Two of the models are of a direct deterrence relationship in which at least one state is attempting to deter the other. The Generalized Mutual Deterrence Game is a model of those direct deterrence situations in which each of two states threatens the other. By contrast, in the Unilateral Deterrence Game, one player (Defender) prefers to preserve the status quo while the other (Challenger) prefers to upset it. The third model, called the Asymmetric Escalation Game, explores extended deterrence relationships in which one state seeks to deter an attack against a third party. The fundamental axioms of perfect deterrence theory (see earlier) are brought to bear in the analysis of each model.

Space and other considerations clearly preclude a formal analysis of the models here.[28] But formal demonstration is not the purpose of this essay. Rather, the aim is to show that important theoretical differences, with significant policy implications, flow from an ostensibly minor alteration of classical deterrence theory's axiomatic base.

Some of the specific differences between classical deterrence theory and perfect deterrence theory are apparent from an analysis of the Generalized Mutual Deterrence Game under incomplete information.[29] Since this model presumes undifferentiated actors (i.e. each player is dissatisfied with the

27. For a formal demonstration, see Bueno de Mesquita (1985).

28. For the formal analysis, see Zagare and Kilgour (2000) and the citations therein.

29. At the start of this game, both players simultaneously choose to cooperate (C) or to defect (D). If both choose either C or D, the game ends. But if one chooses C and the other chooses D, the player choosing *C* is provided with another opportunity to defect, i.e. to retaliate. If, at the end of the game, both players have chosen C, the status quo prevails; if both have chosen D, conflict (or war) results. If one player chooses C and the other D, the defecting player gains an advantage. For the interested reader, the extensive form of the Generalized Mutual Deterrence Game is given in the Appendix.

status quo), it provides a particularly apt context in which to compare the deductions of classical deterrence theory with those of perfect deterrence theory.[30]

In the Generalized Mutual Deterrence Game, there are a number of conditions under which the survival of the status quo is consistent with rational choice. Not all of these conditions, however, are equally probable. The status quo is most likely to endure when a Sure-Thing Deterrence Equilibrium exists.[31] Under this equilibrium, neither player has an incentive to challenge the other. Hence, peace is at hand.

Since many of perfect deterrence theory's policy prescriptions flow from the strategic properties of the Sure-Thing Deterrence Equilibrium (and analogous equilibria in related models),[32] it will be instructive to highlight briefly some of its salient strategic characteristics:

- For a Sure-Thing Deterrence Equilibrium to exist, both players *must* project highly credible threats. In the absence of this condition, the status quo is unlikely to survive rational play. Recall that decision-theoretic deterrence theorists implicitly presume that all deterrent threats lack credibility. As a result, they are, *logically*, unable to make this, perhaps obvious but nonetheless incontrovertible, connection between threat credibility and deterrence stability.
- The likelihood that a Sure-Thing Deterrence Equilibrium will exist is increased, ceteris paribus, as the cost of conflict is increased. Significantly, however, an increase in costs does not *always* increase the likelihood that a Sure-Thing Deterrence Equilibrium will exist, suggesting that there are distinct limits to the stabilizing impact of weapons of mass destruction and that an overkill capability, recommended by many classical deterrence theorists, is just that, overkill.
- A Sure-Thing Deterrence Equilibrium is more likely to exist and, consequently, the status quo is more likely to survive when the status quo is highly valued by the players. While this observation might appear to be self-evident, it is noteworthy that classical deterrence theorists, who tend to focus on threat capability, all but ignore the impact of satisfaction

30. The Generalized Mutual Deterrence Game is analyzed in detail in Kilgour and Zagare (1991) and Zagare and Kilgour (2000).

31. The Sure-Thing Deterrence Equilibrium is a perfect Bayesian equilibrium. This means that it arises in a game of incomplete information (see footnote 14). Henceforth, unless qualified, the term 'equilibrium' should be taken to imply a perfect Bayesian equilibrium.

32. For example, the strategic properties of Certain Deterrence Equilibrium in the Unilateral Deterrence Game are quite similar to those of the Sure-Thing Deterrence Equilibrium in the Generalized Mutual Deterrence Game (Zagare and Kilgour, 2000: 149–50).

(or dissatisfaction) with the status quo on deterrence stability.[33] In consequence, their policy prescriptions tend to slight the importance of diplomatic initiatives in preserving peace (e.g. Kagan, 1995).

To be complete and accurate, it is important to point out that, in the Generalized Mutual Deterrence Game, there is a theoretical possibility, albeit remote, of the status quo persisting when the conditions taken as axiomatic by decision-theoretic deterrence theorists are approached. In the Generalized Mutual Deterrence Game, a Bluff Equilibrium exists whenever both players have threats that are all but incredible. Under a Bluff Equilibrium, both players generally, but not always, initiate conflict. In consequence, there may be times in which no challenge to the status quo is made. It is unlikely in the extreme, however, for the status quo to survive rational play under a Bluff Equilibrium over the long haul.

Within the confines of the Generalized Mutual Deterrence Game, then, the *only* explanation of the 'long peace' that is consistent with the canons of rationality and with the axioms of classical deterrence theory is that luck prevailed.[34] Of course, since this dubious argument is inconsistent with the core conclusion that, under parity, the stability of nuclear relationships is extremely robust, few classical deterrence theorists make it. In consequence, classical deterrence theorists have opted to explain the stability of the post-war period by either sacrificing logical consistency or by making special assumptions that require a disinterested actor to carry out all deterrent threats (see earlier). The recommendations that flow from the classical view of deterrence include policies that favor an overkill capability and promote proliferation and that guard against arms reductions that are carried 'too far' (Intriligator and Brito, 1984). As well, during a crisis, statesmen and women are encouraged to seek an advantage by reducing flexibility, by exercising implacability or by behaving recklessly.

In this context it should also be noted that it is possible for the status quo to survive rational play in the Unilateral Deterrence Game *even when all deterrent threats are minimally credible*.[35] One might argue, therefore, that this result squares classical deterrence theory with the restrictions of rationality. It would – except for the fact that it requires yet another assumption

33. Classical deterrence theory is frequently characterized as apolitical. One reason is that classical deterrence theory holds 'the fundamental conflict of interest underlying a crisis as fixed' (Powell, 1985: 96).

34. The same is true of Chicken. Under the mixed strategy equilibrium, the *Status Quo* occurs sometimes but not necessarily often.

35. For a detailed analysis of the Unilateral Deterrence Game, see Zagare and Kilgour (1993a, 2000).

that lies outside the theory: differentiated actors, which explains why some classical deterrence theorists selectively adhere to the assumption that all states are similarly motivated.

To be somewhat more specific, in the Unilateral Deterrence Game, the players have distinct roles and distinct motivations: one player, the 'Defender', hopes to preserve the status quo while the other, the 'Challenger', would prefer to overturn it.[36] In classical deterrence theory, however, all states are considered alike.[37] Thus, in order to use this simple game model to construct an explanation for the absence of a superpower conflict, one must necessarily cast off yet another axiom of classical deterrence theory. All of which is simply another way of saying that any theory that posits egotistical, rational and undifferentiated actors, all of whom lack a credible retaliatory threat, is inconsistent with the actual stability of the post-war international system.

By contrast, an explanation of stability of the post-war system arises quite naturally in perfect deterrence theory: all-out conflict was avoided simply because each side's retaliatory threat was sufficiently capable and credible to deter either superpower from attacking the other. While this explanation might appear unexceptional, it runs counter to the conventional wisdom. As will be seen later, perfect deterrence theory's policy implications also stand in stark contrast to classical deterrence theory's: states should, inter alia, develop a minimum deterrent capability, pursue arms control agreements, cap military spending, avoid proliferation policies and, in crisis, seek compromise by adopting firm-but-flexible negotiating stances and tit-for-tat military deployments. Walt (1999a: 25), then, is factually incorrect in asserting that perfect deterrence theory reinvents 'the central elements of deterrence theory without improving on it'.

This is not to say that perfect deterrence theory's policy recommendations are necessarily new or unique. Indeed, it is quite difficult to offer completely novel policy prescriptions on most national security issues. Clearly, proponents and opponents of almost every proposed weapon system or deployment stance have staked out just about every conceivable policy position, pro

36. In this game, Challenger begins play by either cooperating or defecting. If Challenger cooperates, the game ends and the status quo prevails. But if Challenger defects, Defender must decide whether to concede or to resist. When Defender concedes, Challenger gains an advantage. When Defender resists, Challenger either gives in or holds firm. In the former case, Defender gains an advantage; in the latter, conflict occurs.

37. The assumption of undifferentiated actors is explicit in structural deterrence theory. It is clearly implicit in the Chicken analogy and most models developed by decision-theoretic deterrence theorists. (Ellsberg's 1959 model is an important exception.)

or con. Thus, to expect any theory of bilateral conflict to prescribe policies that have not as yet been imagined elsewhere is unrealistic.[38]

What one should expect, though, is that any theory of inter-state conflict initiation maintains logical consistency. Indeed, it is the absence of this attribute in classical deterrence theory that helps to explain the existence of many conflicting prescriptions in the policy literature. Policy recommendations that flow from political imperatives rather than from strategic principles and a consistent axiomatic base are bound to contradict one another. Perfect deterrence theory, by contrast, offers a consistent perspective in which to view the dynamics of deterrence and a clear logic supporting its policy prescriptions. This is, perhaps, its most important feature.

3. Theoretical Propositions, Empirical Expectations and Policy Implications

At this point, one might well ask what are perfect deterrence theory's most important theoretical propositions; what empirical expectations arise from its axiomatic base; and what policy prescriptions follow from a strategic theory that respects logical consistency? In this section, each of these questions is answered in the context of the theory's principal variables, which include threat capability and credibility, the cost of conflict and satisfaction (or dissatisfaction) with the status quo. Among the more specific queries raised are: What conditions are most conducive to deterrence success? What is the precise connection between threat credibility and the stability of the status quo? How do the dynamics of direct deterrence relationships differ from those of extended deterrence relationships? Which extended deterrence deployment stances are most efficacious? What negotiating style and crisis management technique is most conducive to peace? Are limited conflicts possible and, if so, when? Why do some conflicts escalate while others do not?

3.1. Capability

In perfect deterrence theory, capable threats are threats that would hurt.[39] Actions that hurt are those that leave a player worse off than if the action were not executed. Operationally, this means that one player's threat is

38. This is one reason why Martin (1999: 83) criticizes Walt (1999a) for 'extracting a few isolated propositions from models and deriding them for being insufficiently original'. As Martin points out, one of the frequently unrecognized benefits of logically consistent theories is their ability 'to generate integrated, coherent complexes of assumptions and propositions'.

39. This definition is, in fact, Schelling's (1966: 7).

capable only if the other, the *threatened* player, prefers the status quo to the outcome that results when and if the threat is carried out. In other words, a threat will lack capability whenever the threatened player prefers to act even when a deterrent threat is acted upon (Zagare, 1987).

When defined in this way, a threat may lack capability for one of two reasons. First, the threatening player may not have the physical ability to carry the threat out. For example, capable nuclear threats require both nuclear weapons and the means to deliver them. In addition, threats may lack capability if the threatened state calculates that the cost of conflict is less than the cost of doing nothing. The US threat against Japan in 1941 may have been incapable for precisely this reason (Snyder and Diesing, 1977), as may have been Poland's threat against Germany in 1939 or Hungary's against the Soviet Union in 1956.

There is considerable opinion in the theoretical literature of international relations that threat capability constitutes a *sufficient* condition for deterrence success.[40] Quinlan (2000/2001: 142), for example, all but accepts the sufficiency of capability for stabilizing hostile bilateral relationships. Speaking of the strategic relationship of the United States and the Soviet Union during the Cold War, he writes: 'the prodigious size to which the two nuclear armouries grew imposed a massive caution almost irrespective of the precise credibility of doctrine for use'.[41] Existential deterrence theorists like Bundy (1983), who hold that the mere existence of nuclear weapons virtually assures strategic stability, also see a highly capable retaliatory threat as sufficient for avoiding crises and war.

Not so in perfect deterrence theory, where deterrence may fail even when threats are capable all around. However, in perfect deterrence theory, capability emerges as the only condition absolutely *necessary* for deterrence success; when one or both states in a mutual deterrence relationship lack capability, deterrence is bound to fail. Since weak states, almost by definition, usually lack the ability to hurt a larger, stronger opponent, it should come as no surprise that there is strong evidence for the proposition that interstate conflict initiators are generally stronger than their opponents (Bueno de Mesquita, 1981: 155–6). These data are consistent with – indeed, they provide compelling systematic empirical support for – an important conclusion of perfect deterrence theory. As well, Harvey's (1998: 691) more recent empirical study 'indirectly supports' perfect deterrence theory's conclusions about the crucial role of capability in deterrence relationships.

40. See Levy (1988: 489–90) for a detailed discussion.

41. Quinlan (2000/2001: 151) fails to apply the same logic to other nuclear states. To promote stability he recommends that both India and Pakistan stop short of a full-blown deployment of their nuclear weapons. Quinlan's prescription is clearly a contradiction: some (e.g. Singh, 1998) would probably label it a double standard.

3.2. Credibility

In the strategic literature, credible threats are frequently equated with threats that ought to be believed (e.g. Smoke, 1987: 93); threats can be believed only when they are rational to carry out (Betts, 1987: 12): thus only rational threats can be credible (Lebow, 1981: 15). In perfect deterrence theory, the formal definition of credibility is consistent with the theoretical linkage between threats that are credible and threats that are both believable and rational: credible threats are precisely those that are consistent with Selten's (1975) perfectness criterion, i.e. with threats that the threatener prefers to execute.

While it is perhaps not surprising to learn that perfect deterrence theory holds that a capable retaliatory threat is a necessary (but not a sufficient) condition for deterrence success, it may, in fact, be surprising to learn that credible threats are neither necessary nor sufficient for deterrence to succeed. This means, inter alia, that deterrence may fail even when all retaliatory threats are capable; and deterrence may succeed even when all retaliatory threats are incredible.

To demonstrate that a credible threat is not sufficient to ensure successful deterrence, consider once more the Rudimentary Asymmetric Deterrence Game of Figure 2 (p. 114). But assume now that state B's threat lacks capability – because the threat is insufficiently hurtful, state A actually prefers *Conflict* to the *Status Quo*. Also assume that state B's threat is credible, that state B prefers *Conflict* to *A Wins*. Given the latter assumption, state B will rationally execute its threat if and when it faces a choice at decision node 2. Nonetheless, deterrence will fail because the rational choice of state A at decision node 1 is to contest the *Status Quo*. In other words, in the absence of a necessary condition (i.e. a capable threat), a credible threat is insufficient for ensuring deterrence success.

Nonetheless, given capable threats, deterrence is most likely to prevail, ceteris paribus, when all threats are highly credible, a straightforward and seemingly unexceptional result that simply cannot be derived from models that presume that all retaliatory threats are forever incredible. The inability of classical deterrence theory to come to such an obvious conclusion without violating the canons of logic speaks to the inadequacies of its theoretical underpinnings.

Although credible threats are not quite the 'magic ingredient' of deterrence, as Freedman (1989: 96) asserts, they come close. Still, it is possible for deterrence to succeed even when a defender's threat is incredible. The key to this possibility, however, is not the characteristics of the defender's threat but of the *challenger's*. In a number of deterrence situations, a challenger whose retaliatory threat is itself not credible is unable to deter a defender from retaliating. In consequence, the challenger is deterred and

the status quo survives rational play. It is not inconceivable, then, for deterrence to succeed even when a defender prefers not to execute its end-game threat. Thus, in perfect deterrence theory, a credible threat is also not a necessary condition for successful deterrence.

That deterrence might rationally work even when a defender's threat is incredible is an important insight into the interactive nature of deterrence relationships. It is also an insight that is clearly missed by theorists who focus exclusively on the characteristics of a defender's retaliatory threat.[42] In consequence, they produce a misleading, perhaps even a distorted, understanding of the dynamics of deterrence.

It is equally significant that perfect deterrence theory finds that mutual deterrence can (but need not) fail, even when both players have capable and credible retaliatory threats. The reason is that even when deterrence is consistent with the strictures of rationality, there are frequently other rational possibilities, some of which are associated with an all-out conflict.[43] This conclusion contrasts sharply with classical deterrence theory's supposition that parity and high war costs virtually eliminate the possibility of a (rational) deterrence breakdown. In other words, in contrast to classical deterrence theory, perfect deterrence theory finds that bilateral deterrence relationships are fragile and fraught with peril. That mutual deterrence is not necessarily robust has important implications for the wisdom of even 'selective' proliferation policies (see later).

To put this in a slightly different way, in the set of inter-related models that forms the basis of perfect deterrence theory, it is almost always the case that the conditions that make peace a real possibility are exactly the same as those that are associated with all-out conflict. From the vantage point of perfect deterrence theory, then, wars do not arise as the inevitable consequence of impersonal forces that lie beyond human intervention or control.[44] Rather,

42. For example, Lebow (1981: 85) writes that 'four conditions emerge as crucial to successful deterrence. Nations must (1) define their commitment clearly, (2) communicate its existence to possible adversaries, (3) develop the means to defend it, or to punish adversaries who challenge it, and (4) demonstrate their resolve to carry out the actions this entails'. Of the four conditions that Lebow argues are necessary for deterrence success, only the third, which can be interpreted as threat capability, emerges as a necessary condition in perfect deterrence theory.

43. More technically, multiple equilibria almost always exist.

44. According to Powell (1985: 84), 'the fact that ''the participants are not fully in control of events'' is fundamental to much of strategic nuclear deterrence theory'. And Trachtenberg (1990/1991: 120), who concurs, comments that the supposition that a major war could occur when statesmen lose control of events 'is one of the most basic and most common notions in contemporary American strategic thought'. Trachtenberg explicitly associates the theory of inadvertent war with classical deterrence theorists like Schelling (1966) and Quester (1966), observing that 'many important conclusions about the risk of nuclear war, and thus about the political meaning of nuclear forces, rest on this fundamental idea'.

they result from choices made by fallible human beings acting rationally, though not necessarily wisely. The good news is that this means that skillful diplomacy and adroit statesmanship may sometimes save the day.[45] The bad news, of course, is that peace can never be all-but-guaranteed, as some classical deterrence theorists suggest.

3.3. Status Quo Evaluations

In classical deterrence theory, states are generally thought of as undifferentiated actors. As such, they have identical interests and aspirations, and an equal motivation to overturn the status quo (Legro and Moravsik, 1999: 13).[46] As Mearsheimer (1990: 12) writes, 'all other states are potential threats'. There are no exceptions to this dictum. This means that there can be no variation in the utility (or disutility) states derive from the existing order. *All* states are presumed to be perpetually dissatisfied (see also Mearsheimer [2001: 2]).

By contrast, in perfect deterrence theory, the players are not necessarily assumed to be undifferentiated. Some, in theory, may be content with the prevailing status quo and, consequently, may lack the motivation to upset it. But even when both players are dissatisfied, the extent of their dissatisfaction may be different. In other words, in perfect deterrence theory, the value of the status quo is an important strategic variable: as satisfaction with the status quo increases, ceteris paribus, so does the likelihood of deterrence success.[47]

Again, because classical deterrence theorists tend to treat the value of the status quo as a constant, they are unable to derive logically this obvious conclusion. This explains why most classical deterrence theorists favor coercive policies that increase the cost of conflict, overlooking, in the process, initiatives that may enhance the prospects for peace by eliminating a

45. Perfect deterrence theory, therefore, is consistent with the argument (e.g. Trachtenberg, 1990/1991: 143) that with different leaders, or with different policies, wars like the First World War can be avoided.

46. Realism, whether classical or neo-, loses much of its explanatory power if only some states are taken to be power maximizers, or if only some states are motivated by structural insecurity. Nonetheless, some decision-theoretic deterrence theorists (e.g. Ellsberg, 1959) do differentiate actors.

47. Perfect deterrence theory is connected, theoretically, with power transition theory (Organski and Kugler, 1980), which sees the international system as hierarchical rather than anarchistic. In a hierarchical system, the dominant state and its allies are generally content with the status quo. Thus the assumption of differentiated actors is not ad hoc in perfect deterrence theory, as it is in most manifestations of classical deterrence theory. For a discussion of the linkage between power transition theory and perfect deterrence theory, see Zagare (1996b).

common (and empirically recognized)[48] root cause of war: dissatisfaction with the status quo. As Van Gelder (1989: 163) observes, 'it is too often forgotten that [successful deterrence] requires not only that the expected utility of acting be relatively low, but that the expected utility of refraining be acceptably high'.

3.4. The Cost of Conflict

In both classical deterrence theory and perfect deterrence theory, the costs associated with conflict have significant strategic implications. But there are important differences in the conclusions the two theoretical frameworks reach about the impact of increased costs on the likelihood of deterrence success. In classical deterrence theory, deterrence success becomes more and more likely as these costs increase. As already mentioned, the monotonic relationship between the costs of conflict and the probability of deterrence success leads many classical deterrence theorists to recommend an overkill capability. One reason for this straightforward connection between cost and stability is that most classical deterrence theorists assume fixed preferences: players *always* prefer an advantage to the status quo; and they *always* prefer not to execute their deterrent threat (i.e. threats are always incredible). Thus, any increase in the cost of conflict always has the same impact relative to other outcomes.

In perfect deterrence theory, by contrast, the cost of conflict is gauged against two other important strategic variables. The first is the value of the status quo (see earlier). One consequence of this variable relationship is that, in perfect deterrence theory, there is a *minimum* cost threshold below which deterrence cannot succeed. This is the point separating threats that are capable from those that are not.

The second reference point is the value of concession. In perfect deterrence theory, the players may, or may not, prefer to concede rather than execute a deterrent threat. This is the reason why there is also a *maximum* threshold beyond which further increases in the cost of conflict do not contribute to the probability of direct deterrence success. Rather than an overkill capability, then, the logic of perfect deterrence theory is consistent with a policy of *minimum deterrence*, which rests on a threat that is costly enough to deter an opponent but that is not so costly that the threat itself is rendered incredible.

An equally important difference is that there is no simple monotonic relationship in perfect deterrence theory between the cost of conflict and the stability of the status quo, as there is in classical deterrence theory. In perfect deterrence theory, there are circumstances under which an increase

48. For chapter and verse, see Geller and Singer (1998: 64–5; 89–92).

in conflict costs will actually undermine a deterrence relationship. More specifically, extended deterrence becomes more and more difficult to maintain as conflict costs rise, simply because defenders become more and more reluctant to respond to an indirect challenge. In perfect deterrence theory, therefore, increased conflict costs can, under some circumstances, be stabilizing but, under others, may have the opposite consequence.

3.5. Negotiating Styles and Crisis Management Techniques

Classical deterrence theory and perfect deterrence theory also differ about the best way to conduct diplomacy. Recall that decision-theoretic deterrence theorists proffer coercive bargaining tactics that either reduce flexibility or that increase an opponent's conflict costs in order to increase the probability of an opponent's concession. By contrast, perfect deterrence theory recommends an approach rooted in reciprocity. Conditionally cooperative strategies like 'tit-for-tat' that reciprocate both cooperation and noncooperation are associated, both theoretically and empirically, with successful compromise.

In perfect deterrence theory, establishing reciprocity is tantamount to establishing credibility which, in turn, makes deterrence more likely. Thus, it is reassuring that in Huth's (1988) statistical analysis of extended deterrence relationships, firm-but-flexible negotiating styles and tit-for-tat deployments are highly correlated with extended deterrence success. Huth defines a firm-but-flexible diplomatic stance as a signal that the defender is willing to compromise, but not capitulate.[49] And a tit-for-tat policy involves an actual response-in-kind during a crisis or mobilization. Thus, the essence of both a firm-but-flexible bargaining approach and a tit-for-tat response to an actual provocation is reciprocity, the norm that signals credibility when promised or threatened, and demonstrates it when practiced.

There is, as well, a large empirical literature that is consistent not only with Huth's findings but also with the theoretical expectations of perfect deterrence theory about the pervasiveness of reciprocal behavior in inter-state interactions.[50] This evidence attests to perfect deterrence theory's explanatory and predictive power. By contrast, this widely observed norm is difficult, if not impossible, for classical deterrence theorists to explain. In their models,

49. It is telling that Trachtenberg (1990/1991: 143) concludes that had Germany taken a firm-but-flexible approach in 1914, the First World War could have been avoided. Trachtenberg argues that Russia would not have mobilized had German Chancellor Bethmann Holweg made it clear to the Russian leadership 'that war was not inevitable, that a political settlement was within reach, that Austria could be led to moderate her demands on Serbia, but that he needed a little time to bring her around'.

50. For a review, see Cashman (1993: Ch. 6) or Zagare and Kilgour (2000: Ch. 10).

which are based on preference structures derived from Chicken, mutual cooperation and mutual defection can never be part of a (pure strategy) equilibrium. Indeed, in Chicken, each player's optimal strategy is always the *reverse* of the other's (see Figure 1), which is why these models tend to speak to the question of which side can expect to win or lose in a crisis (see, for instance, Powell, 1987). Ties, however, which involve reciprocity, are extremely rare events in brinkmanship models, making *both* war and peace unfathomable. Clearly, the pertinent theoretical puzzle for classical deterrence theory is explaining why peace is so often observed, why crises do not occur all the time.

3.6. Extended Deterrence Deployment Strategies

Using the Asymmetric Escalation Game as a guide, perfect deterrence theory also speaks to the efficacy of a number of competing extended deterrence deployment policies.[51] As in the Unilateral Deterrence game, the players in this game are also clearly differentiated: one is called Challenger and the other is called Defender. Only Challenger can contest the status quo. Should a challenge be issued, Defender can respond in one of three ways: it can concede, it can respond-in-kind or it can escalate. Defender's initial response opportunity constitutes its first- (or tactical-) level threat. Should Defender respond-in-kind and Challenger subsequently escalate, Defender has a second- (or strategic-) level threat: to counter-escalate. In this game, Challenger has but one threat: to counter-escalate should Defender escalate first.

In the Asymmetric Escalation Game there is an easy and natural way to distinguish all-or-nothing deployments (like massive retaliation) from limited-war stances (like flexible response): A limited-war approach requires that Defender possess both a capable and a credible first-level threat to respond-in-kind. By contrast, under an all-or-nothing deployment, Defender's first-level threat to respond-in-kind is non-existent (i.e. devoid of credibility).[52]

As one would expect from a 'common-sense' approach to deterrence, all-or-nothing deployments are not particularly conducive to deterrence success. They are effective in precisely two situations: (1) when a challenger lacks a capable strategic threat, as perhaps the Soviet Union did during the Eisenhower administration; or (2) when a potential challenger's strategic level threat is not very credible. In each case, a challenger will be unable to

51. The Asymmetric Escalation Game is analyzed in Zagare and Kilgour (1993b, 1995, 1998, 2000).

52. See Gacek (1994) for a lucid discussion of the differences between the all-or-nothing and limited-war approaches to deterrence.

deter a defender from a disproportionate escalatory response. Hence, no challenge is made and the status quo survives. However, in the absence of either condition, all-or-nothing deployment policies are unlikely to deter a determined challenger.

In perfect deterrence theory, more flexible, limited-war deployments fare better than all-or-nothing stances. A number of distinct deterrence equilibria exist under the conditions that define this type of deployment stance, thereby increasing the chance that the status quo will survive rational play. Again, one might expect as much since, under a limited-war deployment, Defender's threat to respond-in-kind is not inherently incredible, as it would be under an all-or-nothing deployment – like the one adopted by Great Britain prior to the First World War, by France prior to the Second World War or by the United States throughout the 1950s.

It is interesting to observe that the most plausible (and, hence, the most probable) deterrence equilibria are consistent with a 'no-first-use' limited-war deployment policy.[53] In other words, deterrence is most likely to succeed precisely when a defender never intends to escalate first. This does not mean, however, that a defender's escalatory threat is without strategic significance. For the challenger to be completely deterred under a no-first-use policy, a defender's strategic level threat must be highly credible as well. To put this differently, for deterrence to work, a defender must be able to deter its opponent not only from initiating a low-level conflict but also from escalating if and when the defender chooses to respond-in-kind.

That deterrence success is consistent with rationality under certain limited-war deployments should not be taken to suggest that extended deterrence relationships are robustly stable, as classical deterrence theory suggests. As is the case in direct deterrence situations, other rational possibilities always exist under precisely the conditions that give rise to stable deterrence. Thus, even under ideal conditions, a rupture of both direct and extended deterrence always remains a distinct theoretical possibility. In consequence, non-escalatory response options are necessary, but not sufficient, for stabilizing extended deterrence relationships.

3.7. Limited Conflicts and Escalation Spirals

The Asymmetric Escalation Game model can also be used to locate the conditions under which limited conflicts and escalation spirals are most likely to

53. By contrast, a 'warfighting' deployment in which defender intends to either respond-in-kind or escalate immediately is very unlikely to stabilize the status quo. For the specifics, see Zagare and Kilgour (2000: Ch. 8). For a discussion of the variety of deployment policies consistent with a limited war approach, see Daalder (1991).

occur, thereby placing two important, yet distinct, real world processes into a more general theoretical context than is provided by classical deterrence theory.

In perfect deterrence theory, it is easy to explain those limited conflicts that occur in the context of an unequal power relationship, as would be the case in a dispute between a major and a lesser (though not necessarily minor) power: such conflicts are to be expected if and when a stronger, yet circumspect, challenger confronts a weaker opponent that lacks the wherewithal (i.e. the capability) to ward off a confrontation. Prussia's brief war with Austria in 1866 is a good example. But what about limited conflicts that take place when both players have the capability to hurt one another? In perfect deterrence theory, these less-than-all-out conflicts are generally associated with the existence of a Constrained Limited-Response Equilibrium.

For a Constrained Limited-Response Equilibrium to exist and, hence, for a limited war to occur, a potential challenger must be uncertain not only about a defender's willingness to respond-in-kind when confronted but also about its willingness to endure an all-out conflict. In other words, limited conflicts under parity conditions require considerable uncertainty, especially on a challenger's part.

Clearly, uncertainty abounds in international politics. Thus it will sometimes happen that a challenger, expecting an immediate concession, will contest the status quo, only to encounter unexpected resistance. The key to distinguishing those conflicts that remain limited from those that escalate to a higher level is the inference that a challenger draws from an unanticipated response – such as China's after UN forces crossed the 38th parallel in 1950.

Limited conflicts transpire precisely when a challenger concludes that a defender, who offers unexpected measured resistance, is also determined to oppose any and all additional aggressive acts.[54] In other words, when a challenger infers that a defender is likely to match, tit-for-tat, any other untoward act, it will assiduously avoid further provocations – lest it finds itself in an all-out conflict it would prefer to avoid.[55] Such was clearly the case after China intervened in the Korean conflict. Fearing a wider war not only with China but perhaps with the Soviet Union as well, US Secretary of Defense George Marshall then decided to 'use all available political, economic and psychological action to limit the war' (quoted in Gacek, 1994: 57).

54. This result is fully consistent with Fearon's (1994a) finding that many of the breakdowns of general deterrence that do not escalate to a higher level involve a challenger who is highly uncertain, ex ante, about a defender's willingness to resist but who, ex post, comes to believe otherwise.

55. Note that reciprocal threats are useful for both deterring initial aggression and also for keeping a conflict capped once it erupts.

To put this in a slightly different way, perfect deterrence theory suggests that limited conflicts should be associated with an unexpected response that sends a signal that is strong enough to give the challenger pause. Examples of such mid-stream strategy revisions in international politics are plentiful: during the 1898 Fashoda crisis, France was compelled to back down in the face of Britain's unanticipated reaction to its plan to take control of the Upper Nile; in 1911, in the midst of the Agadir crisis, Britain's unforeseen support of France persuaded Germany to accept limited compensation for ceding control of Morocco to France; and in 1962, the Soviet Union withdrew its missiles from Cuba when the United States unexpectedly blockaded the island. Perfect deterrence theory, therefore, is consistent with Snyder and Diesing's (1977: 397) observation that, during an intense crisis, a 'strategy revision is initiated when a massive input of new information breaks through the barrier of the image and makes a decision maker realize that his diagnosis and expectations were somehow radically wrong and must be corrected'.

Escalation spirals, by contrast, occur *only* when such a strategy revision is not made, i.e. when an Escalatory Limited-Response Equilibrium is in play. Under this equilibrium, after an unexpected response, the challenger believes that the defender is bluffing, that the defender's reaction is but a prelude to eventual capitulation. In consequence, the challenger escalates. More often than not, under an Escalatory Limited-Response Equilibrium, the challenger's expectations will be realized, so that an all-out conflict is avoided. Nonetheless, there will also be times when the challenger's belief that it will be able to out-escalate its opponent will also prove to be wrong. Conflicts like the First World War can, in part, be traced to such mistaken beliefs.

To be sure, perfect deterrence theory is not the only theoretical framework that attempts to explain when and why some conflicts remain limited while others escalate. For example, the 'spiral school' (Jervis, 1976) explains conflict escalation in nearly the same way that perfect deterrence theory does: conflicts that escalate to the highest level are the consequence of a deadly combination of aggressive actions, mistaken beliefs and threats that backfire.

But because they select on the dependent variable (i.e. conflicts that escalate), spiral theorists see only part of the larger picture. In consequence, the main conclusion they reach – that threats intended to deter unintentionally promote instability and conflict – is misleading. It is an undeniable fact that deterrent threats are often successful.

It is interesting to observe, however, that many classical deterrence theorists are similarly guilty of case selection bias. Making much of dramatic instances of *deterrence failures*, such as the instability that ultimately engulfed Europe after the 1938 Munich crisis, classical deterrence theorists like Kagan (1995) conclude that inter-state wars are the consequence of failed appeasement policies. To prevent war, they recommend deterrence

policies that rely on threats that are not only believable but that hurt as well.

From the vantage point of perfect deterrence theory, both theories are correct, but only in part, which is consistent with Jervis's (1976: 84) observation that 'neither theory is confirmed all the time' or why deterrence sometimes succeeds and why conflicts sometimes escalate out of control. Successful deterrence and conflict spirals are events that take place under very different circumstances. By placing both processes, as well as limited conflicts and other outcomes, into a wider theoretical context, perfect deterrence theory achieves a more general understanding of the dynamics of inter-state conflict behavior than either the spiral model or classical deterrence theory.

4. Summary and Conclusions

Classical deterrence theory is logically inconsistent, empirically inaccurate and prescriptively deficient. Perfect deterrence theory, by contrast, makes consistent use of the rationality postulate; it is *prima facie* in accord with the empirical record; and its common-sense policy prescriptions are grounded in strict logic.

As well, the organizing power of each theory is different. Because it makes a fixed and unnecessarily strong assumption about the costs associated with conflict, classical deterrence theory's logical and empirical domain is unduly circumscribed. Indeed, many strategic analysts regard classical deterrence theory as primarily, if not exclusively, a theory of bilateral nuclear relationships; in consequence, conventional or non-nuclear deterrent relationships are oftentimes treated as if they have a separate and distinct dynamic (e.g. Mearsheimer, 1983).

By contrast, perfect deterrence theory makes no particular assumption about the costs of conflict. It is, therefore, a more general theory, applicable to a much wider range of strategic relationships. In other words, perfect deterrence theory is not simply a divergent theory of nuclear war avoidance. Rather, it is a universal theory of conflict initiation and resolution, applicable to both nuclear and to non-nuclear interactions. As such, it can be used to help explain why crises occur, why some conflicts escalate and others do not, and when and why limited and all-out wars are waged. In fact, perfect deterrence theory's empirical domain is not even restricted to inter-state interactions. As a general theory of strategic interaction, it is potentially applicable to inter-group or inter-personal conflict of interest situations whenever and wherever they occur.

Besides these important differences, there are a number of other features that separate the two approaches to contentious inter-state relationships.

Perfect deterrence theory and classical deterrence theory are built on different axiomatic bases; their treatment of credibility is distinct; their theoretical implications are not the same; they have divergent empirical expectations; and they frequently offer contradictory policy prescriptions. For example, classical deterrence theory supports both an overkill capability and the selective proliferation of nuclear weapons. Perfect deterrence theory opposes these policies. Classical deterrence theory opposes both minimum deterrence deployments and significant arms reductions. Perfect deterrence theory is consistent with both these policies. Finally, classical deterrence theorists prescribe coercive bargaining stances based on increasing war costs and inflexible or reckless bargaining tactics. By contrast, perfect deterrence theory suggests that conditionally cooperative policies based on reciprocity are more efficacious.

What are some of the specific implications of perfect deterrence theory for current foreign policy debates? Assuming that the United States has no future revisionist aspirations in the central state system, perfect deterrence theory suggests that, in the short term, the national missile defense system being deployed by Washington would not seriously undermine strategic stability – even if it does leak.[56] In perfect deterrence theory, the status quo is most likely to survive when a satisfied preponderant power exists.

The longer-term consequences of a missile shield, however, could be less benign. If other states, like China or Russia, respond to an American ABM system either by developing a similar system or, more likely, by expanding their offensive capability in order to blunt its effectiveness, then any purported benefits of a national missile defense may prove to be ephemeral. Worst still, if the abrogation of the ABM Treaty eventually leads to widespread proliferation of Russian or Chinese weapons technology, the inter-state system will only become more dangerous, more likely to break down. In addition, the deployment of even a limited ABM system can only increase the probability that the United States might be tempted proactively to overturn the status quo on the Korean peninsula or elsewhere.

In terms of nuclear weapons policy, perfect deterrence theory is consistent with a minimal deterrent combined with a 'no first use' deployment. While neither approach, of course, can guarantee peace, combined they offer the best chance for avoiding a catastrophe. More generally, the US should continue to build down (but not eliminate) its nuclear arsenal (National Academy of Sciences, 1997). It is far better for the United States to meet China, or other potential rivals, on its way down than on a rival's way up. Ceteris paribus, the more costly nuclear war, the less likely it is that the status quo will survive over time.

56. This is not to say that a missile defense system would augment strategic stability.

The broad foreign policy orientation suggested by perfect deterrence theory is rooted in reciprocity (or conditional cooperation). In practice, this means avoiding inflexible hard-line policies that are not only likely to decrease the satisfaction of other states but are also likely to lead to a negative response that leaves all concerned worse off. For example, a firm pledge not to attack in return for demonstrable disarmament may have averted a costly US war with Iraq in 2003; a similar guarantee might help the US peacefully resolve its current dispute with North Korea. However, reciprocity also means avoiding unconditionally cooperative stances (i.e. appeasement policies) that are patently one-sided. Unilateral concessions are generally invitations for exploitation.

Perhaps the most difficult foreign policy problem for the United States in the years ahead is to manage its relationship with China. All of the previously advanced prescriptions apply. A rigid confrontational foreign policy toward China will be counterproductive. Additional steps towards deployment of a national missile defense system will continue to antagonize the Chinese, increasing their dissatisfaction and likely prompt China to expand the size of its nuclear arsenal and to finance that growth by exporting dangerous technologies (Tammen et al., 2000). Clearly, the United States will be well served to find a creative compromise on Taiwan, which is obviously a dangerous flash point. At the same time, it should maintain a strong presence in the Pacific, lest it be left with the unpalatable choices associated with all-or-nothing deployments.

Much the same could be said about Russia. While the creation of the NATO–Russia Council in 2002 was both a good sign and a good step, needlessly provoking Russia by expanding NATO to include the Baltic states or, perhaps, Georgia will not serve the cause of peace. In contrast, finding common ground on a range of economic issues, including Russia's trading relationship with a reconstituted Iraq, its entry into the World Trade Organization, and the restructuring of its foreign debt, would help make the world a safer place. Here too, however, reciprocity is the key. American unilateralism – in either direction – would not be constructive. In the long run, neither appeasement nor coercive diplomacy works.

Beyond these specifics, the best hope for peace over the long haul is to promote an international environment in which grievances are addressed and not allowed to fester. To be sure, a peace that rests on credible and capable threats can be a seductive short-term fix. But such a peace is bound to unravel, eventually, as dissatisfied rational agents interact in an imperfect world.

Acknowledgements

I would like express my appreciation for the helpful comments and suggestions of Bruce Bueno de Mesquita, Vesna Danilovic, Paul K. Huth, D. Marc Kilgour, Jacek Kugler, Stephen Quackenbush, John Vasquez and Ann E. Zagare on an earlier version of this manuscript.

APPENDIX

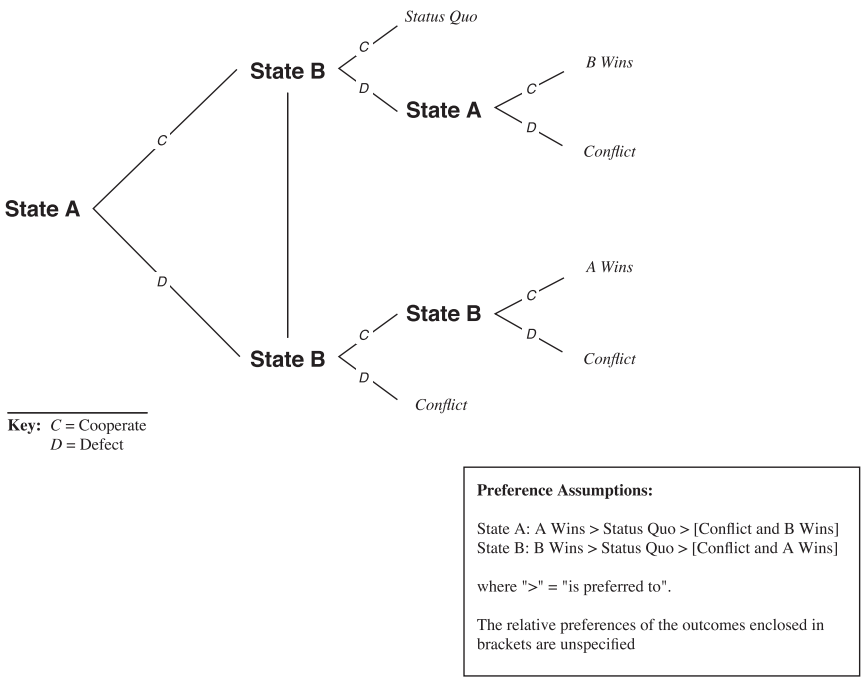

Figure A1. Generalized Mutual Deterrence Game (*Source:* Zagare and Kilgour, 2000)

REFERENCES

Achen, Christopher H. (1987) 'A Darwinian View of Deterrence', in Jacek Kugler and Frank C. Zagare (eds) *Exploring the Stability of Deterrence*. Denver, CO: Lynne Rienner.

Betts, Richard K. (1987) *Nuclear Blackmail and Nuclear Balance*. Washington, DC: Brookings.

Brodie, Bernard (1959) *Strategy in the Missile Age*. Princeton, NJ: Princeton University Press.

Bueno de Mesquita, Bruce (1981) *The War Trap*. New Haven, CT: Yale University Press.

Bueno de Mesquita, Bruce (1985) 'Toward a Scientific Understanding of International Conflict: A Personal View', *International Studies Quarterly* 29: 121–36.

Bueno de Mesquita, Bruce and James D. Morrow (1999) 'Sorting Through the Wealth of Notions', *International Security* 24: 56–73.
Bueno de Mesquita, Bruce, James D. Morrow and Ethan R. Zorick (1997) 'Capabilities, Perception and Escalation', *American Political Science Review* 91: 15–27.
Bundy, McGeorge (1983) 'The Bishops and the Bomb', *New York Review of Books* (16 June): 3–8.
Cashman, Greg (1993) *What Causes War? An Introduction to Theories of International Conflict*. New York: Lexington Books.
Daalder, Ivo H. (1991) *The Nature and Practice of Flexible Response: NATO Strategy and Theater Nuclear Forces Since 1967*. New York: Columbia University Press.
Danilovic, Vesna (2001) 'The Sources of Threat Credibility in Extended Deterrence', *Journal of Conflict Resolution* 45: 34–69.
Danilovic, Vesna (2002) *When the Stakes Are High: Deterrence and Conflict among Major Powers*. Ann Arbor, MI: University of Michigan Press.
Ellsberg, Daniel (1959) 'The Theory and Practice of Blackmail', Lecture at the Lowell Institute, Boston, MA, 10 March. Reprinted in Oran R. Young, ed. (1975). *Bargaining: Formal Theories of Negotiation*. Urbana, IL: University of Illinois Press.
Fearon, James D. (1994a) 'Signaling Versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model', *Journal of Conflict Resolution* 38: 236–69.
Fearon, James D. (1994b) 'Domestic Political Audiences and the Escalation of International Disputes', *American Political Science Review* 88: 577–92.
Freedman, Lawrence (1989) *The Evolution of Nuclear Strategy*, 2nd edn. New York: St Martin's.
Friedman, Thomas L. (2001) 'Who's Crazy Here?', *The New York Times*, 15 May: A27.
Gacek, Christopher M. (1994) *The Logic of Force: The Dilemma of Limited War in American Foreign Policy*. New York: Columbia University Press.
Gaddis, John Lewis (1986) 'The Long Peace: Elements of Stability in the Postwar International System', *International Security* 10: 99–142.
Gaddis, John Lewis (1997) *We Now Know: Rethinking Cold War History*. New York: Oxford University Press.
Gauthier, David (1984) 'Deterrence, Maximization, and Rationality', *Ethics* 94: 474–95.
Gibbons, Robert (1992) *Game Theory for Applied Economists*. Princeton, NJ: Princeton University Press.
Geller, Daniel S. and J. David Singer (1998) *Nations at War: A Scientific Study of International Conflict*. Cambridge: Cambridge University Press.
Glaser, Charles (1989) 'Why Do Strategists Disagree about the Requirements of Strategic Nuclear Deterrence?', In Lynn Eden and Steven E. Miller (eds) *Nuclear Arguments*. Ithaca, NY: Cornell University Press.
Gray, Colin S. (1974) 'The Urge to Compete: Rationales for Arms Racing', *World Politics* 26: 207–33.
Harsanyi, John C. (1974) 'Communications', *American Political Science Review* 68: 1694–5.
Harvey, Frank P. (1998) 'Rigor Mortis, or Rigor, More Tests: Necessity, Sufficiency, and Deterrence Logic', *International Studies Quarterly* 42: 675–707.
Howard, Nigel (1971) *Paradoxes of Rationality: Theory of Metagames and Political Behavior*. Cambridge, MA: MIT Press.
Huth, Paul K. (1988) *Extended Deterrence and the Prevention of War*. New Haven, CT: Yale University Press.
Huth, Paul K. (1999) 'Deterrence and International Conflict: Empirical Findings and Theoretical Debates', *Annual Review of Political Science* 2: 61–84.
Intriligator, Michael D. and Dagobert L. Brito (1981) 'Nuclear Proliferation and the Probability of Nuclear War', *Public Choice* 37: 247–60.

Intriligator, Michael D. and Dagobert L. Brito (1984) 'Can Arms Races Lead to the Outbreak of War?', *Journal of Conflict Resolution* 28: 63–84.
Jervis, Robert (1976) *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
Jervis, Robert (1978) 'Cooperation Under the Security Dilemma', *World Politics* 30: 167–214.
Jervis, Robert (1979) 'Deterrence Theory Revisited', *World Politics* 31: 289–324.
Jervis, Robert (1985) 'Introduction', in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein (eds) *Psychology and Deterrence*. Baltimore, MD: The Johns Hopkins University Press.
Jervis, Robert (1988) 'The Utility of Nuclear Weapons', *International Security* 13: 80–90.
Kagan, Donald (1995) *On the Origins of War and the Preservation of Peace*. New York: Doubleday.
Kahn, Herman (1962) *Thinking About the Unthinkable*. New York: Horizon Press.
Kilgour, D. Marc and Frank C. Zagare (1991) 'Credibility, Uncertainty, and Deterrence', *American Journal of Political Science* 35: 303–34.
Krauthammer, Charles (2001) 'Dense on Missile Defense', *The Washington Post*, 11 May: A45.
Krauthammer, Charles (2002) 'The Terrible Logic of Nukes', *Time*, 2 September: 84.
Kydd, Andrew (1997) 'Game Theory and the Spiral Model', *World Politics* 49: 371–400.
Lebow, Richard Ned (1981) *Between Peace and War: The Nature of International Crisis*. Baltimore, MD: The Johns Hopkins University Press.
Lebow, Richard Ned (1984) 'Windows of Opportunity: Do States Jump Through Them?', *International Security* 9: 147–86.
Legro, Jeffrey W. and Andrew Moravsik (1999) 'Is Anybody Still a Realist?', *International Security* 24: 5–55.
Levy, Jack S. (1988) 'When Do Deterrent Threats Work?', *British Journal of Political Science* 18: 485–512.
Martin, Lisa (1999) 'The Contributions of Rational Choice: A Defense of Pluralism', *International Security* 24: 74–83.
Mearsheimer, John J. (1983) *Conventional Deterrence*. Ithaca, NY: Cornell University Press.
Mearsheimer, John J. (1990) 'Back to the Future: Instability in Europe After the Cold War', *International Security* 15: 5–56.
Mearsheimer, John J. (2001) *The Tragedy of Great Power Politics*. New York: Norton.
Morgenthau, Hans J. (1973) *Politics Among Nations*, 5th edn. New York: Knopf.
Morrow, James D. (1994) *Game Theory for Political Scientists*. Princeton, NJ: Princeton University Press.
Morrow, James D. (2000) 'The Ongoing Game-Theoretic Revolution', in Manus I. Midlarsky (ed.) *Handbook of War Studies II*. Ann Arbor, MI: University of Michigan Press.
Nalebuff, Barry (1986) 'Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation', *Conflict Management and Peace Science* 9: 19–30.
Nash, John (1951) 'Non-cooperative Games', *Annals of Mathematics* 54: 286–95.
National Academy of Sciences (1997) *The Future of U.S. Nuclear Weapons Policy*. Washington, D.C.: National Academy Press.
Niou, Emerson M.S. and Peter C. Ordeshook (1999) 'The Return of the Luddites', *International Security* 24: 84–96.
O'Neill, Barry (1992) 'Are Game Models of Deterrence Biassed Towards Arms-Building? Wagner on Rationality and Misperception', *Journal of Theoretical Politics* 4(4): 459–77.
Organski, A.F.K. and Jacek Kugler (1980) *The War Ledger*. Chicago: University of Chicago Press.
Powell, Robert (1985) 'The Theoretical Foundations of Strategic Nuclear Deterrence', *Political Science Quarterly* 100: 75–96.
Powell, Robert (1987) 'Crisis Bargaining, Escalation, and MAD', *American Political Science Review* 81: 717–35.

Powell, Robert (1999) 'The Modeling Enterprise and Security Studies', *International Security* 24: 97–106.
Quackenbush, Stephen L. (2003) 'General Deterrence and International Conflict: Bridging the Formal/Quantitative Divide', Ph.D. thesis, University at Buffalo, SUNY, Buffalo, NY.
Quackenbush, Stephen L. and Frank C. Zagare (2001) 'Territorial Disputes and General Deterrence', Paper presented at the annual meeting of the American Political Science Association, San Francisco, CA, 30 August–2 September.
Quester, George (1966) *Deterrence Before Hiroshima*. New York: Wiley.
Quinlan, Michael (2000/2001) 'How Robust is India–Pakistan Deterrence?', *Survival* 42: 141–54.
Rapoport, Anatol (1968) 'Editor's Introduction', in Carl von Clausewitz *On War*. Harmondsworth, England: Penguin Books.
Rapoport, Anatol (1992) 'Comments on ''Rationality and Misperceptions in Deterrence Theory'' ', *Journal of Theoretical Politics* 4: 479–84.
Schelling, Thomas C. (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schelling, Thomas C. (1966) *Arms and Influence*. New Haven, CT: Yale University Press.
Scoville, Herbert, Jr (1981) *MX: Prescription for Disaster*. Cambridge, MA: MIT Press.
Selten, Reinhard (1975) 'A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory* 4: 25–55.
Senese, Paul D. and Stephen L. Quackenbush (2003) 'Sowing the Seeds of Conflict: The Effect of Dispute Settlements on Durations of Peace', *Journal of Politics* 65: 696–717.
Singh, Jaswant (1998) 'Against Nuclear Apartheid', *Foreign Affairs* 77: 41–52.
Smoke, Richard (1987) *National Security and the Nuclear Dilemma*. Reading, MA: Addison-Wesley.
Snyder, Glenn H. (1972) 'Crisis Bargaining', in Charles F. Hermann (ed.) *International Crises: Insights from Behavioral Research*. New York: Free Press.
Snyder, Glenn H. and Paul Diesing (1977) *Conflict Among Nations: Bargaining, Decision Making and System Structure in International Crises*. Princeton, NJ: Princeton University Press.
Tammen, Ronald L., Jacek Kugler, Douglas Lemke, Allan C. Stam III, Mark Abdollahian, Carole Alsharabati, Brian Efird and A.F.K. Organski (2000) *Power Transitions: Strategies for the 21st Century*. New York: Chatham House.
Trachtenberg, Marc (1990/1991) 'The Meaning of Mobilization in 1914', *International Security* 15: 120–50.
Trachtenberg, Marc (1991) *History and Strategy*. Princeton, NJ: Princeton University Press.
Van Evera, Stephen (1984) 'The Cult of the Offensive and the Origins of the First World War', *International Security* 9: 58–107.
Van Evera, Stephen (1990/91) 'Primed for Peace: Europe After the Cold War', *International Security* 15: 7–57.
Van Gelder, Timothy J. (1989) 'Credible Threats and Usable Weapons: Some Dilemmas of Deterrence', *Philosophy and Public Affairs* 18: 158–83.
Walt, Stephen M. (1999a) 'Rigor or Rigor Mortis? Rational Choice and Security Studies', *International Security* 23: 5–48.
Walt, Stephen M. (1999b) 'A Model Disagreement', *International Security* 24: 115–30.
Waltz, Kenneth N. (1964) 'The Stability of the Bipolar World', *Daedalus* 93: 881–909.
Waltz, Kenneth N. (1979) *Theory of International Politics*. Reading, MA: Addison-Wesley.
Waltz, Kenneth N. (1981) 'The Spread of Nuclear Weapons: More May Be Better', Adelphi Paper No. 171. London: International Institute for Strategic Studies.
Waltz, Kenneth N. (1990) 'Nuclear Myths and Political Realities', *American Political Sciences Review* 84: 731–45.
Waltz, Kenneth N. (1993) 'The Emerging World Structure of International Politics', *International Security* 18: 44–79.

Wohlforth, William C. (2000) 'Correspondence: Brother, Can You Spare a Paradigm? (Or, Was Anybody Still a Realist?)', *International Security* 25: 182–4.
Young, Oran R., ed. (1975) *Bargaining: Formal Theories of Negotiation*. Urbana, IL: University of Illinois Press.
Zagare, Frank C. (1987) *The Dynamics of Deterrence*. Chicago: University of Chicago Press.
Zagare, Frank C. (1990) 'Rationality and Deterrence', *World Politics* 42: 238–60.
Zagare, Frank C. (1996a) 'Classical Deterrence Theory: A Critical Assessment', *International Interactions* 21: 365–87.
Zagare, Frank C. (1996b) 'The Rites of Passage: Parity, Nuclear Deterrence and Power Transitions', in Jacek Kugler and Douglas Lemke (eds) *Parity and War: Evaluations and Extensions of 'The War Ledger'*. Ann Arbor, MI: University of Michigan Press.
Zagare, Frank C. (1999) 'All Mortis, No Rigor', *International Security* 24: 107–14.
Zagare, Frank C. and D. Marc Kilgour (1993a) 'Asymmetric Deterrence', *International Studies Quarterly* 37: 1–27.
Zagare, Frank C. and D. Marc Kilgour (1993b) 'Modeling ''Massive Retaliation'' ', *Conflict Management and Peace Science* 13: 61–86.
Zagare, Frank C. and D. Marc Kilgour (1995) 'Assessing Competing Defense Postures: The Strategic Implications of ''Flexible Response'' ', *World Politics* 47: 373–417.
Zagare, Frank C. and D. Marc Kilgour (1998) 'Deterrence Theory and the Spiral Model Revisited', *Journal of Theoretical Politics* 10: 59–87.
Zagare, Frank C. and D. Marc Kilgour (2000) *Perfect Deterrence*. Cambridge: Cambridge University Press.

FRANK C. ZAGARE is Professor and Chair, Department of Political Science, State University of New York at Buffalo. He is co-author of *Perfect Deterrence* (Cambridge, 2000), author of *The Dynamics of Deterrence* (Chicago, 1987) and *Game Theory: Concepts and Applications* (Sage, 1984), editor of *Modeling International Conflict* (Gordon and Breach, 1990) and co-editor of *Exploring the Stability of Deterrence* (Lynne Rienner, 1987). His work has appeared in the *American Journal of Political Science*, *World Politics*, *International Studies Quarterly*, *Journal of Peace Research*, *Conflict Management and Peace Science*, *Theory and Decision*, *Journal of Conflict Resolution*, *International Interactions* and *Synthese*. ADDRESS: Department of Political Science, University at Buffalo, SUNY, 520 Park Hall, Buffalo, NY 14260, USA. [email: fczagare@buffalo.edu]