Toward a Unified Theory of Interstate Conflict

FRANK C. ZAGARE
Department of Political Science, University at Buffalo, SUNY, Buffalo, New York, USA

This article seeks to refine the power transition proposition, thereby reducing its permissiveness, by linking it to an axiomatically compatible theory of interstate conflict initiation called perfect deterrence theory. The similarities between the two frameworks are discussed, as are several significant conceptual and terminological differences. The broad conclusions and policy implications of power transition and perfect deterrence theory are shown to be the same. But perfect deterrence theory is much more precise about the conditions associated with the onset, escalation, and resolution of interstate conflict than is power transition theory. These conditions are summarized and discussed; the extensive empirical support for perfect deterrence theory’s principal deductions is also reviewed.

KEYWORDS power transition, perfect deterrence theory, necessary conditions, credibility, interstate conflict, escalation, war

The power transition research program is among the most active and vigorous programs in international relations. Even the untimely death of its originator in 1998 has not dampened the enthusiasm of a large cadre of scholars who continue to extend, refine, test, and apply the framework whose broad outlines A.F.K. Organski first sketched in 1958 in his seminal World Politics.¹

Power transition’s signal theoretical contribution is its identification of a theoretically pertinent condition necessary for the occurrence of major power wars. Specifically, power transition theory asserts that wars between and among the major powers occur only when a previously weaker dissatisfied state achieves parity (or near parity) with its principal rival.

Despite strong empirical support for this proposition, however, complaints about power transition’s descriptive ability abound. At the
Conference on Parity and War held at Claremont University in 1992, for instance, one major international relations researcher complained that the theory was inconsistent with the observation that neither World War I or World War II had started with a direct attack against Great Britain, Germany’s main competitor in 1914 and in 1939. Similarly Copeland (2000, p. 14) issues an indictment of the theory on the grounds that “Germany was in fact preponderant when it took on the system” and precipitated both world wars.

The glaring discrepancy between the impressive empirical backing for the power transition proposition and certain reservations about its theoretical standing among some policymakers and conflict process scholars can be traced to misunderstandings about the meaning of necessary conditions. Such errors, according to Goertz and Starr (2003), are common in the literature of interstate conflict.

Necessary conditions specify a set of antecedents that are required for certain consequences to occur. For example, it is well known that both hydrogen and oxygen are essential to sustain life. Yet a wide variety of circumstances (and venues) remain consistent with the existence of life. Similarly with the power transition proposition. The proposition is not inconsistent with an assortment of processes that could conceivably proceed the onset of a major power conflict.

To put this in a slightly different way, the power transition proposition is as consistent with major powers wars, like World War I and World War II, that occur when extended deterrence breaks down, as it is with wars, like the Franco-Prussian War of 1870, that result from a direct deterrence failure. As well, the central proposition of power transition theory is in fact logically compatible with wars that arise both immediately prior to, and directly after, parity has been achieved.

If blame were to be assigned for these and related misinterpretations of the power transition proposition, more than a modicum could be placed on the way the theory developed. Specifically, Organski’s (1958) original outline of the theory included speculation that war was most likely prior to an actual transition, before parity had been achieved. Twenty years later, after a systematic empirical investigation (Organski and Kugler, 1980, pp. 58–59), this subsidiary conjecture was rejected: Germany was in fact found stronger than Great Britain in 1914 and in 1939.

Defenders of the paradigm have noted, with more than some justification, that theoretical adjustments—especially those that are made in the context of more nuanced observations and more sensitive measurements—are to be expected as a theory is refined and extended, and that such adjustments are part and parcel of normal science. But critics (e.g., DiCicco and Levy, 1999, p. 699) plausibly, though I believe erroneously, claim that such adjustments are a sign that the paradigm is degenerative in the sense of Lakatos (1970), that discrepant information is simply being explained away, and that little or no additional theoretical content is being provided.
As before, the hidden culprit in this dispute is an ambiguity connected with a set-theoretic definition of a necessary condition: if \( X \) is a necessary condition for \( Y \), \( Y \) must be a subset of \( X \), or \( X \) must be a superset of \( Y \). When defined in this way,\(^3\) it is clear that defenders of the paradigm begin with the conviction that the intersection between the power transition condition (\( X \)) and the set of major power wars (\( Y \)) is extensive, while critics believe that the necessary condition is excessively permissive, that is, that the correspondence between \( X \) and \( Y \) is too small to be theoretically or empirically meaningful. In other words, the divergent evaluations of the paradigm are due to a disagreement about the potency of the necessary condition that is central to the power transition proposition. Proponents see the condition as strong and discriminating; critics see it as weak and, therefore, insufficiently diagnostic.

It is not difficult to finger the real problem. As Copeland (2000, p. 13) correctly points out, power transition theory “has no deductively consistent theory of war initiation.”\(^5\) In consequence, the theory remains less than precise, perhaps even inconsistent, about the circumstances under which major power wars occur. These ambiguities notwithstanding, proponents of the paradigm see the glass as half full: since power transition theory has excess empirical content over its main theoretical competitor, balance of power theory, they consider it to be a progressive research program. At the same time, critics see the glass as half empty: with respect to war initiation, the paradigm may be seen as excessively permissive.

Lest the reader be concerned, my purpose here is not to adjudicate this probably irresolvable debate; rather it is to advance the discussion by speaking to the concerns of both camps. More specifically, I propose to refine the power transition proposition, thereby reducing its permissiveness, by linking it to an axiomatically compatible theory of war initiation called perfect deterrence theory (Zagare and Kilgour, 2000; Zagare, 2004). My goal is to provide additional precision about the set of conditions under which major power wars are likely to be avoided. Along the way I hope to dispel some common misperceptions about the nature of interstate conflict.

PERFECT DETERRENCE THEORY AND CLASSICAL DETERRENCE THEORY

Perfect deterrence theory is a general theory of conflict initiation, escalation, and resolution, relevant to strategic interactions between both nuclear and nonnuclear states. The theory applies equally to general and immediate deterrence relationships, as well as to situations of direct and immediate deterrence. As a universal theory, its empirical domain includes, but is not limited to, interactions between major powers.\(^6\)

Unlike classical deterrence theory, which is logically inconsistent, empirically inaccurate, and prescriptively deficient (Zagare, 1996a), perfect
deterrence theory 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. One way to think about perfect deterrence theory, therefore, is as a theoretical alternative to classical (or rational) deterrence theory, which it partially subsumes.

Classical deterrence theory has two distinct, yet conceptually compatible, branches. Even though the branches focus on different units of analysis, the assumptions they make, the conclusions they reach, and the policy prescriptions they draw, are essentially the same.

Structural deterrence theorists like Waltz (1979), Mearsheimer (1990), and Gaddis (1986) hold that the post World War II international system is inordinately stable. Tracing this purported stability to the “twin pillars” of bipolarity and high war costs, structuralists contend that the probability of a rational or premeditated war between nuclear powers approaches zero (Waltz, 1990, p. 740). These theorists conclude that any future war between nuclear states will most likely be accidental (or irrational).

Decision-theoretic deterrence theorists like Schelling (1960, 1966), Ellsberg (1959), and Powell (1987) begin where structuralists leave off. Using either formal or informal expected utility and game-theoretic models, they presume war to be irrational. Although this axiom is oftentimes obscured, the instrumental irrationality of war follows logically from the paradigm’s defining assumption—that all-out conflict is the worst possible outcome for all involved parties. Decision-theoretic deterrence theorists, then, take as a given a central deduction of structural deterrence theory.

The assumption that conflict or war is always a mutually worst outcome is clearly problematic. Once war is assumed to be irrational, it is impossible to offer a logically consistent explanation of the stability of the post war international system without violating other tenets of the paradigm (Trachtenberg, 1991, p. 32).

To see why, consider for a moment the *Rudimentary Asymmetric Deterrence Game* summarized in Figure 1. There are two players in this simple game, Challenger and Defender. For expository purposes, assume that Challenger needs to be deterred; that is, that Challenger prefers the outcome labeled *Challenger Wins* to the outcome called *Status Quo*. Also assume (for now) that the outcome labeled *Conflict* is the worst outcome for both players.

If *Conflict* is taken as a mutually worst outcome, then Defender always rationally concedes (at node 2) if Challenger makes a demand (because Defender prefers *Challenger Wins* to *Conflict*). This means that Defender’s threat to resist Challenger’s demand is always irrational to execute. In other words, the threat is never credible (Selten, 1975).

Given that it is always irrational for Defender to carry out its threat, an instrumentally rational Challenger will opt to contest the status quo (at node 1) if it prefers *Challenger Wins* to the *Status Quo*. All of which demonstrates that the defining assumption of decision-theoretic deterrence theory (i.e.,
that conflict is the worst outcome for both players) is inconsistent with the possibility of deterrence success, at least in this simple deterrence game.

To square this circle and explain the so-called long peace, decision-theoretic deterrence theorists have taken one of two tacks. Some theorists, like Brodie (1959, p. 293), Gauthier (1984), and Rhodes (1989) abandon logical consistency and allow for the possibility of irrational threats being executed. Others, like Powell (1987, 2003) and Nalebuff (1986), maintain consistency by introducing a *deus ex machina*, a third player called *Nature*, to carry out threats independent of the preferences of the players. For the first group of theorists, successful deterrence is explained by the possibility of irrational behavior; for the second, it is explained by the possibility of an accidental war.

Neither explanation suffices. Consider first the implication of abandoning logical consistency. It is well-established that any statement can be derived from a logically inconsistent theory.9 In consequence, such frameworks cannot be falsified (Bueno de Mesquita and Morrow, 1999, p. 57). Walt (1999) notwithstanding, then, it is clear that logical consistency is the sine qua non of good theory, that explanations that fail to satisfy this unassailable standard must be set aside.

Now consider the possibility that the long peace is best explained by the autonomous risk of all-out-conflict. As even Powell (1987, p. 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. As well, this explanation is difficult to square with the fact that history provides us with no known example of an inadvertent war (Trachtenberg, 1990/1991, 1991), or with the observation that, during the cold war, the superpowers worked mightily to reduce the probability of an accidental war.
Perfect deterrence theory takes still another approach. Unlike decision-theoretic deterrence theory, where credibility is invariant, perfect deterrence theory makes no fixed assumption about the players’ preferences between Conflict and conceding to an opponent. In perfect deterrence theory, credibility, defined as the extent to which a player is seen to prefer Conflict to acquiescing to its opponent, is measured on a continuum. As well, in perfect deterrence theory, the principal source of the risks run by the players is not some disinterested autonomous force. Rather, all threats are carried out by the players, and only by the players. Even more importantly, only credible (or rational) threats may be executed.

These seemingly small differences between the two paradigms have important theoretical consequences. Because it treats threat credibility as a constant, decision-theoretic deterrence theory is simply incapable of establishing a logical relationship between the credibility of threats and the dynamics of deterrence. But even were credibility allowed to fluctuate, a fundamental problem would still exist: as long as incredible threats are executable, the very possibility of exploring the theoretical relationship between credibility and the operation of deterrence is precluded.

Perfect deterrence theory, by contrast, is well situated to explore the logical connection between threat credibility and the dynamics of dyadic interstate relationships. For example, its analysis of the Rudimentary Asymmetric Deterrence Game (under complete information) reveals that for the Status Quo to be stable in this simple game, Defender’s threat must be perfectly credible. In the absence of this condition, deterrence always fails. Similarly, given incomplete information, deterrence success requires that Defender’s threat be “sufficiently” credible to dissuade Challenger from contesting the status quo. When Defender’s threat is insufficiently credible, deterrence rationally fails.

The claim here is not that these findings are exceptional. Just the contrary. The conclusions are straightforward and wholly consistent with intuition and common sense. Nonetheless, these deductions are inconsistent with the micro-foundations of decision-theoretic deterrence theory. This fact alone should raise grave concerns about classical deterrence theory’s descriptive and prescriptive utility.

PERFECT DETERRENCE THEORY AND POWER TRANSITION THEORY

The major deductions of perfect deterrence theory are derived from an examination of four interrelated incomplete information game models. Unlike classical deterrence theory, which holds “the fundamental conflict of interest underlying a crisis as fixed” (Powell, 1985, p. 96), both power transition theory
and perfect deterrence theory treat satisfaction with the status quo as an important strategic variable. At the most basic level, therefore, these two approaches to interstate conflict are entirely consistent.

There are other significant similarities. As discussed below, the broad conclusions of power transition and perfect deterrence theory are congruent. As well, the two paradigms support many of the same policy recommendations.

Nonetheless, there are also a number of nontrivial differences between power transition and perfect deterrence theory. Table 1 summarizes the most salient similarities and several important differences between perfect deterrence theory and power transition.

One such distinction concerns dynamics. Both frameworks are dynamic, but in different ways. Perfect deterrence theory models sequential choices that unfold at a particular point in time while power transition explicitly models power relationships that develop over time.11

The different sense in which the two theories are dynamic does not mean that they compete with one another. In fact, they are theoretical complements. For example, power transition theorists recognize that “decision makers count” (Tammen et al., 2000, p. 96); yet because there is no explicit decision–theoretic component of the theory, power transition theorists are unable to specify exactly how and when policymakers matter. Perfect deterrence theory, however, is far more specific about the conditions associated with successful deterrence, crisis initiation, and conflict escalation. This means that perfect deterrence theory can be used to provide additional precision to power transition’s oftentimes permissive deductions about the

---

**TABLE 1** Perfect Deterrence Theory and Power Transition Compared

<table>
<thead>
<tr>
<th>Similarities:</th>
<th>Power Transition</th>
<th>Perfect Deterrence Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad conclusions</td>
<td>similar</td>
<td>similar</td>
</tr>
<tr>
<td>Policy implications</td>
<td>congruent</td>
<td>congruent</td>
</tr>
<tr>
<td>Status quo</td>
<td>significant</td>
<td>significant</td>
</tr>
<tr>
<td>Axiomatic base</td>
<td>compatible</td>
<td>compatible</td>
</tr>
<tr>
<td>Strategic deterrence</td>
<td>fragile/contingent</td>
<td>fragile/contingent</td>
</tr>
<tr>
<td>Nature of parity</td>
<td>potentially unstable</td>
<td>potentially unstable</td>
</tr>
<tr>
<td>Power asymmetries</td>
<td>potentially very stable</td>
<td>potentially very stable</td>
</tr>
<tr>
<td>Capable threats</td>
<td>necessary but not sufficient</td>
<td>necessary but not sufficient</td>
</tr>
<tr>
<td></td>
<td>for successful mutual deterrence</td>
<td>for successful mutual deterrence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences:</th>
<th>Power Transition</th>
<th>Perfect Deterrence Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamics</td>
<td>temporal shifts of power</td>
<td>sequential choice over time</td>
</tr>
<tr>
<td>Predictions</td>
<td>less than precise</td>
<td>more precise</td>
</tr>
<tr>
<td>Level of analysis</td>
<td>dyadic/structural</td>
<td>dyadic/decision-theoretic</td>
</tr>
<tr>
<td>Definition of capability</td>
<td>attribute or possession</td>
<td>either/or commodity</td>
</tr>
<tr>
<td>Nature of capability</td>
<td>continuous variable</td>
<td>dichotomous variable</td>
</tr>
<tr>
<td>Basis of threat capability</td>
<td>economic/political resources</td>
<td>opponent’s preference for the status quo over conflict</td>
</tr>
<tr>
<td>Definition of credibility</td>
<td>implicit in threat capability</td>
<td>distinct from threat capability</td>
</tr>
</tbody>
</table>
dynamics of conflict initiation and resolution. The contention here is that perfect deterrence theory’s point predictions about strategic relationships at particular stages of the transition process are at once consistent with, but more explicit than, the expectations drawn from power transition theory.

Although the two frameworks are complementary, their nomenclature sometimes differs. In power transition theory, for instance, massive retaliation refers to a particular asymmetric power relationship—clear-cut nuclear preponderance. By contrast, in perfect deterrence theory, this term is used to describe a deterrence relationship in which one or both states forgo tactical (or mid-level) threats in favor of all-or-nothing strategic level threats. The different conventions suggest that care must be taken while moving back and forth from one theory to the other.

Some key terms are also operationalized differently in the two paradigms. Capability is among the most important examples. In power transition theory, capability is defined as an attribute, as something that states possess (to varying degrees)—states either have some, more than a little, or a lot of it. By contrast, in perfect deterrence theory, capability, defined as the ability to hurt, is an all-or-nothing commodity. States either have it or they do not.

To be somewhat more specific, in power transition capability is determined by the bundle of economic and political resources controlled by a state. In perfect deterrence theory, the cost/benefit calculations of a state’s rival determines whether its threat is capable or not. In perfect deterrence theory, therefore, a state with a large military arsenal and a strong economy would probably, but not necessarily, possess a capable threat, as it would in power transition theory. In perfect deterrence theory, if the threat is not sufficiently hurtful that an opponent would prefer the status quo to fighting, it would not be considered capable. For example, power transition theorists consider the U.S. threat against North Vietnam to have been capable. But from the vantage point of perfect deterrence theory, the U.S. threat lacked capability. Even after it was executed, the threat was not costly enough to persuade the North Vietnamese to give up the fight.

Another key difference, perhaps the most important, concerns the definition of credibility. Power transition theorists tend to equate credibility with capability; states with capable threats are presumed to have credible threats, and vice versa. Such is not the case in perfect deterrence theory where these concepts are conceptually distinct. Thus, in perfect deterrence theory, there are four, rather than two, kinds of threats: 1) capable but not credible, 2) credible but not capable, 3) neither capable nor credible, or 4) both capable or credible.

The proliferation of threat categories in perfect deterrence theory is theoretically enabling. By introducing a second dimension along which threats may be distinguished, and by refining the strategic environment framing interstate disputes, perfect deterrence theory is able to be more...
discriminating about the conditions associated with crisis initiation and dispute escalation.

NECESSITY, SUFFICIENCY, AND THE OCCURRENCE OF MAJOR POWER WAR

Power Transition is a theory of the necessary conditions of major power war. As noted, critics charge, either implicitly or explicitly, that the power transition proposition is excessively permissive, while defenders assert that the condition is an important diagnostic of interstate conflict. Perfect deterrence theory speaks to the concerns of both groups by characterizing the causal relationship between capability, credibility, and the likelihood of peaceful power transitions.

Capability

There is considerable opinion among classical deterrence theorists that threat capability constitutes a sufficient condition for deterrence success (Levy, 1988). Existential deterrence theorists like Bundy (1983), for example, claim that the mere existence of nuclear weapons produces an existential fear of conflict among decisionmakers that all but eliminates the possibility of an all-out war. Others, taking a less extreme position, nonetheless presume that the high costs associated with nuclear and other weapons of mass destruction have rendered moot the necessity of making threats that are even minimally believable. Quinlan (2000/2001, p. 142) is a case in point. 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.”

Notice the underlying congruence between Quinlan’s argument and the resolution of the paradox of deterrence offered by decision–theoretic deterrence theorists: the issues at stake in a conflict are presumed to be of little or no consequence. Threat credibility is considered virtually irrelevant. Rather, the fear of an inadvertent war is called upon to explain why, during the Cold War, leaders in Washington and Moscow were loath to precipitate crises and were excessively cautious when managing them.

Proponents of this hypothesis often point to the July crisis of 1914 as a compelling example. In the run up to World War I, policymakers in Vienna, Paris, Berlin, and St. Petersburg, it is said, simply lost control of events when rigid, preestablished, mobilization plans took on lives of their own. What followed was a war “that no one wanted.”

Trachtenberg (1990/1991), however, shows that this argument does not stand up to intense examination, and that there is little basis in fact for the
claim that World War I was in any sense accidental. Since there is no other potentially viable example of an inadvertent conflict, resolutions of the paradox of deterrence that rest on the autonomous probability of war are based on a nonexistent empirical process.

All of which is not to imply that threat capability is irrelevant, or that the costs of war are of no strategic import. In perfect deterrence theory, increasing the cost of conflict can, under certain conditions, have important stabilizing consequences. Specifically, in perfect deterrence theory, there is a minimum cost threshold below which deterrence cannot succeed. This is the threshold that separates threats that are capable from those that are not. In other words, in perfect deterrence theory capable threats constitute a necessary condition for deterrence success. But they are not sufficient mechanisms for stabilizing the status quo—as the previous discussion of the Rudimentary Asymmetric Deterrence Game clearly demonstrates.13

Credibility

The alert reader may have noticed, however, that the discussion of the Rudimentary Asymmetric Deterrence Game also assumed that Defender’s threat lacked credibility—since Defender’s preference was for Challenger Wins over Conflict. Wouldn’t deterrence success be assured if this preference were reversed and Defender’s threat taken as credible? In the Rudimentary Asymmetric Deterrence Game the answer is most assuredly “yes.” Nonetheless, it is not generally true that a credible (and capable) threat is sufficient for deterrence to work. For example, in the Generalized Mutual Deterrence Game (see Fig. 2), two outcomes are consistent with rationality when the two players (now called States A and B) are afforded threats that are both credible and capable (Zagare and Kilgour, 2000, pp. 77–78).14 One outcome is associated with the status quo (i.e., with peaceful transitions) while the other is associated with conflict (i.e., unsuccessful deterrence). Clearly, mutually credible threats are not sufficient to ensure that deterrence succeeds—even when a necessary condition (i.e., threat capability) is satisfied.

It is interesting to observe, however, that a credible retaliatory threat is also not necessary for deterrence to succeed. To illustrate this, consider for now the Unilateral Deterrence Game given by Figure 3. Once again, the players in this game are called Challenger and Defender. Assume, for the moment, that Defender’s threat is capable; that is, that Challenger prefers the Status Quo to Conflict. Also assume that both players lack credible threats; that is, each prefers to back down rather than endure Conflict. Surprisingly, under these conditions, the Status Quo is stable and deterrence works (Zagare and Kilgour, 2000, pp. 139–142).

The reason is manifest: because Challenger’s threat lacks credibility, Challenger is unable to deter Defender from resisting a demand for a change in the Status Quo. In consequence, deterrence succeeds.
To summarize briefly: in perfect deterrence theory capable threats emerge as necessary but not sufficient conditions for deterrence success. This result squares well with power transition theory. Specifically, even
under parity, when threats are most likely to be capable all around, conflict may not be avoidable. Still, war is not inevitable. As well, credible threats, even when coupled with capable threats, are neither necessary nor sufficient for stabilizing the status quo. Again, this result is entirely consistent with power transition’s main conclusion that parity relationships, nuclear or otherwise, are tenuously stable, at best, and patently unstable, at worst.

SOME GENERAL OBSERVATIONS

At this point one may well ask what, beyond theoretical convergence and the refinement of the power transition proposition, does perfect deterrence theory add to power transition’s explanatory power? The short answer is that perfect deterrence theory provides specificity to a number of questions about the genesis of interstate conflicts. Before explicating these specifics, however, some general observations drawn from the theory are laid out.

Deterrence is Rarely Certain

Classical deterrence theorists, by definition, hold that bipolar nuclear relationships are inordinately stable. Waltz (1990, p. 740), for instance, argues that “the probability of major war among states having nuclear weapons approaches zero.” Intriligator and Brito (1981, p. 256) reach exactly the same conclusion.

However, in perfect deterrence theory, deterrence is rarely certain. To be sure, deterrence equilibria—under which no player contests the status quo—exist in all of the incomplete information game models that form the theory’s core. Nonetheless, it is also the case that, for the most part, the deterrence equilibria coexist with other equilibria, some of which are associated with all-out conflict. To put this in a slightly different way, in perfect deterrence theory it is almost always true that the conditions that are most conducive to peace are also consistent with war.

This observation is important for two reasons. First, it dovetails with power transition’s claim that parity conditions constitute a necessary, but not a sufficient, condition for major power conflicts. And second, it suggests that deterrence is not as certain or as robust a relationship as classical deterrence theorists argue. In perfect deterrence theory, as in power transition, decisionmakers matter. Oftentimes, the strategies they adopt, and the equilibria they settle on, determines whether war or peace follows. Neither outcome can be ruled in, or out, by the dictates of rationality.

There is, however, one important exception to the conclusion that deterrence is always tenuous. In the Unilateral Deterrence Game (see Figure 3), a deterrence equilibrium (appropriately termed the Certain Deterrence Equilibrium) uniquely exists whenever Defender’s credibility is sufficiently high.
But the existence of this equilibrium doesn’t obviate the general conclusion that deterrence may be less than sure.

To explain why, note first that for a Certain Deterrence Equilibrium to exist, one of the players (i.e., Defender) must be completely satisfied with the prevailing order. Not only is this a special circumstance, but the assumption of differentiated actors also contradicts the standard realist assumption of like units (Waltz, 1979). This means that the circumstances that give rise to a highly stable status quo are inconsistent with a core axiom of classical deterrence theory. And whenever that axiom is realized, deterrence becomes less than certain. All of which suggests that a highly stable status quo is an exception rather than a rule.

As well, even if one interprets the existence of a unique deterrence equilibrium in the Unilateral Deterrence Game as suggestive of the potential robustness of deterrence relationships, one should keep in mind that the suggestion applies only to direct deterrence relationships and ignores the possibility that war might erupt when extended deterrence breaks down. In other words, even when a satisfied defender is able to deter a frontal attack, it may be unable to prevent an indirect assault on its interests. In perfect deterrence theory, as in the real world, extended deterrence is the more problematic case (Quester, 1989, p. 63). Danilovic’s (2002, p. 53) observation that, with the exception of the Franco–Prussian war, all major power wars since the Congress of Vienna have involved a failure of extended deterrence confirms the deduction.

The bottom line, therefore, is that the conclusion that the careful calibration of deterrence threats can all but eliminate the possibility of a rational conflict cannot be sustained, either theoretically or empirically. Very special circumstances must prevail for deterrence to be highly probable. Still, it is noteworthy that these circumstances are entirely consistent with the power transition hypothesis that parity, nuclear or otherwise, is a necessary—but not sufficient—condition for major power wars.

Deterrence is Not an Anachronism

There are some who argue that deterrence is an anachronism, an unwanted vestige of a world order that no longer exists. Some critics of the strategy of deterrence charge that if the end of the Cold War did not mark deterrence’s demise, the September 11 attacks on the United States surely did. Others claim that, even in its heyday, policies based on deterrence were inherently deficient. Spiral theorists (Jervis, 1976), for example, hold that wars are the consequence of aggressive actions, mistaken beliefs, and threats that backfire.

Such arguments, however, are seriously flawed. Premature pronouncements of the irrelevance of deterrence tend to flow from an excessively narrow view of the theory’s empirical domain. Deterrence, however, is a
universal phenomenon that is central to strategic interactions in diverse settings: in the family, in the workplace, in the marketplace, in domestic politics, and in relations between and among interstate actors. There is no compelling theoretical or empirical reason why deterrence relationships are restricted to any particular kind of weapons technology or any specific type of strategic relationship.

But even if deterrence is defined narrowly as a relationship between nuclear rivals, as it frequently is, the conceptual importance of deterrence would not be eliminated. As more and more states acquire nuclear or other weapons of mass destruction, even the narrowest conceptualizations of the meaning of deterrence becomes more and more pertinent. The relationship between India and Pakistan, for example, is clearly a rivalry whose dynamic is framed by a nuclear context. And, it is more than likely that nuclear weapons will play a central role in the relationship between the United States and China if and when the interests of these two nations seriously clash sometime in the future.

Nonetheless, in the wake of the terrorist attacks in 2001, the United States government has taken the position that deterrence policies are inadequate. In consequence it has substituted preemptive action, rather than retaliatory threats, as the privileged strategic response to threats from non-state actors and hostile subnational groups.

To be sure, it may be difficult to deter an opponent who is not easily identified or who is immune to punishment. Threats that cannot be executed, or that do not hurt enough, lack capability and are, consequently, ineffective deterrents. But as Snyder (1961) demonstrated long ago, deterrence is not exclusively rooted in punishment. Deterrence can also be achieved through denial. A policy of homeland defense, therefore, is not only a potentially effective prophylactic against terrorism, it is also consistent with a strategy of deterrence broadly defined.

But this is not the end of the story. Perfect deterrence theory, and power transition as well, suggests another avenue for mitigating terrorist threats, namely, manipulating the status quo. In perfect deterrence theory, as in life, challenges to the status quo become less likely, *ceteris paribus*, as its value increases.

To some the suggestion that the root causes of terrorism should be addressed in order to lessen its occurrence smacks of appeasement and the failed policies of the British and French prior to World War II. But before accepting this line of argument, several points are in order.

First, there was a time in which the term “appeasement” was considered a time-honored and respectable diplomatic stratagem and not the pejorative, politically charged, concept it is today. It was only after Munich in 1938 that the term fell into disrepute (Gilpin, 1981, p. 193). That this is so is unfortunate. The failure of the British and French notwithstanding, the Munich analogy is misleading. Those who advance it are clearly guilty of
case selection bias; they tend to ignore the fact that political accommodation or conciliation, to suggest more neutral terms, lies at the heart of many effective foreign policy initiatives. For instance, the British were able to successfully navigate a potentially dangerous power transition with the United States around the turn of the twentieth century by recognizing U.S. interests in South and Central America and in Alaska (Rock, 2000, ch. 1). Critics of accommodation also tend to ignore the fact that after Munich Hitler was likely neither appeasable nor deterrable. So if one approach to stabilizing the status quo is to be rejected because of this dramatic failure, so should the other. Finally, those who reject efforts to conciliate an opponent also tend to ignore the very large empirical literature that supports the proposition that satisfaction with the status quo makes war considerably less likely.

**PERFECT DETERRENCE THEORY: MAJOR PROPOSITIONS AND POLICY PRESCRIPTIONS**

As noted above, classical deterrence theory has two main branches: one is structural; the second decision-theoretic. Though distinct, the two branches are axiomatically compatible and synergistic. Much the same can be said about the connection between power transition and perfect deterrence theory. While power transition’s focus is on interstate power dynamics, perfect deterrence theory’s micro-foundations are rooted in rational choice. Thus, it should not be surprising that many of perfect deterrence theory’s major propositions and its most important policy prescriptions are entirely consistent with those of power transition. For instance, in both paradigms, capability asymmetries that favor a satisfied defender are stabilizing; asymmetries that favor a dissatisfied challenger are destabilizing; the conditions associated with peace and stability coexist with those associated with conflict and war.

One important theoretical insight that is unique to perfect deterrence theory, however, concerns the relationship between conflict costs and deterrence success. Both structural deterrence theorists and decision-theoretic deterrence theorists see a strictly positive and monotonic relationship between these two variables. For example, Mearsheimer (1990, p. 19) argues that “the more horrible the prospect of war, the less likely it is to occur.” On the basis of this belief classical deterrence theorists support an overkill capability and oppose “significant” arms reductions. An overkill capability, it is argued, increases war costs and, hence, reduces its probability. Conversely, since meaningful cuts in military capability diminish these costs, they raise war’s likelihood.

By contrast, in perfect deterrence theory, the relationship between conflict costs and the probability of deterrence success is decidedly non-monotonic.
As noted earlier, there is a minimum cost threshold below which deterrence cannot succeed. But, in both the Generalized Mutual Deterrence Game and the Unilateral Deterrence Game, the likelihood that the status quo survives rational play increases as the cost of conflict increases—but only up to a certain point. Once this latter threshold is reached, however, further increases in the cost of conflict do not increase the probability of direct deterrence success. Thus, rather than an overkill capability, the logic of perfect deterrence theory supports a policy of minimum deterrence, which rests on a threat that is costly enough to deter an opponent, but not so costly that the threat itself is rendered incredible.

Equally important, however, is the fact that there are circumstances under which an increase in the costs of conflict may actually undermine the status quo. Specifically, in situations of extended deterrence wherein one state hopes to prevent a challenge to another, an increase in the cost of conflict may render deterrence success less likely since, ceteris paribus, more costly threats are likely less credible. In other words, in international affairs, as in most things human, moderation may be the key to success.

Another important difference between classical deterrence theory and perfect deterrence theory lies in their prescribed negotiating stances. In the decision-theoretic strand of classical deterrence theory, success in crisis bargaining stems not only from stratagems that increase the costs of conflict but also from tactical maneuvers that involve seizing the initiative or making a commitment to a hard-line position. (Ellsberg, 1959; Schelling, 1960, 1966; Hovi, 1998). But in perfect deterrence theory, coercive policies tend to lead to breakdowns of deterrence and, frequently, to all-out conflict. Much more effective, however, are reciprocal strategies (like tit-for-tat). For example, in both the Generalized Mutual Deterrence Game and the Unilateral Deterrence Game, successful deterrence is associated with conditionally cooperative strategies that require states to be prepared not only to cooperate if the other cooperates, but also to withhold cooperation if the other intends not to cooperate. In other words, in perfect deterrence theory, reciprocity rather than unilateral action is seen as the diplomatic stance most likely to lead to peaceful relationships.

**EMPIRICAL EXPECTATIONS**

To this point, the theoretical differences between classical deterrence theory and perfect deterrence theory have been noted. The theoretical affinities between power transition and perfect deterrence theory have also been highlighted. Yet little evidence has been offered to support the view that perfect deterrence theory provides a more nuanced understanding of interstate conflict initiation, escalation, and resolution than does power transition. And without empirical support, theoretical insight is a devalued currency.
It is the case, however, that many of perfect deterrence theory’s theoretical expectations are supported empirically. For example, given perfect deterrence theory’s claim that a capable threat constitutes a necessary condition for deterrence success, one would expect conflict initiators, in general, to be more capable than those they attack. As it turns out, there is strong empirical support of this proposition which holds, regardless of the belligerents’ alignment status or the level of initiated conflict (Bueno de Mesquita, 1981, pp. 154–156).20,21

If perfect deterrence theory has any theoretical traction one would also expect that a significant positive relationship between satisfaction with the status quo and peace. As mentioned before, empirical support for this underappreciated proposition is both extensive and strong.

Another important regularity that flows from perfect deterrence theory concerns the most efficacious political/diplomatic stances. Recall that classical deterrence theory prescribes coercive policies based on increasing war costs and inflexible bargaining tactics. In contrast, perfect deterrence theory recommends a conditionally cooperative approach to interstate politics. In terms of an overall foreign policy orientation, then, the two theories offer diametrically opposed prescriptions and have contradictory empirical expectations.

The empirics on this score are unambiguous. There is clearly an abundance of evidence that suggests that the manipulative bargaining tactics offered by decision-theoretic deterrence theorists are neither descriptive of actual behavior, nor effective when they are adopted. In an important study of extended deterrence, for example, Danilovic (2002) finds that coercive bargaining tactics are poor predictors of deterrence success. Similarly, Huth (1999, p. 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.” And in their now classic analysis of crisis bargaining, Snyder and Diesing (1977, pp. 489–490) observe that “strongly coercive tactics such as physically ‘irrevocable commitments’ or severe committing threats are rarely used.” Clearly, the prescriptions of classical deterrence theory are empirically dubious and for the most part inconsistent with standard diplomatic practices.22

On the other hand, perfect deterrence theory’s expectations about interstate behavior are strongly supported. Levy (1989) notwithstanding, the empirical literature suggesting that reciprocity operates as a behavioral norm is about as close to an empirical law as exists in international politics. (See, inter alia, Cashman, 1993, ch. 6 and Sullivan, 2002, ch. 9). As well, Huth’s (1988) findings that both tit-for-tat strategies and firm-but-flexible negotiating styles are associated with an increased probability of deterrence success are consistent with perfect deterrence theory’s predictions.

In perfect deterrence theory deterrence relationships oftentimes break down. When they do, it is generally the case that conflicts do not escalate to
the highest level. In each of the four models that constitute perfect deterrence theory, war is a rare event. Indeed, one-sided victories are the most common outcome type if and when deterrence fails. Thus, it should not be surprising to learn that “do nothing” or “take no military action” is also the modal category of all defenders involved in real-world militarized interstate disputes (Hart and Ray, 1996). Once again, theory and fact converge.

Perfect deterrence theory also offers a logic to explain the impact of dispute settlements on recurring conflict. Since, in perfect deterrence theory, unilateral deterrence situations are, ceteris paribus, more stable than mutual deterrence relationships, imposed settlements—which presumably create a one-sided post dispute relationship—should last longer than negotiated settlements, which imply a post dispute situation of mutual deterrence. A recent empirical study “strongly supports” this implication of perfect deterrence theory. In an analyses of 2,536 interstate conflict settlements between 1816 and 1992, Senese and Quackenbush (2003) find that disputes whose settlement is imposed last significantly longer, on average, than either unsettled disputes or disputes that are resolved through negotiation. The results hold even when they are controlled for regime type, relative capability, power shifts, contiguity, decisive outcomes, and war. Senese and Quackenbush’s (2003, p. 715) conclusion is that, in their highly-focused test of one of perfect deterrence theory’s relationship predictions, the theory “performs quite well.”

In a more direct test, Quackenbush (2003) uses binary and multinomial logit methods to examine some of perfect deterrence theory’s equilibrium predictions. Quackenbush’s innovative study indicates that the expectations of the theory are generally supported by the empirical record. According to Quackenbush (2003, ch. 6), his far-reaching test demonstrates both “the accuracy of the theory’s predictions” and its “heuristic and organizing power.”

**SUMMARY AND CONCLUSIONS**

The power transition paradigm represents an active and progressive research program in international security studies. Its identification of a theoretically significant and empirically-supported necessary condition for the occurrence of major power war has advanced our understanding of the most important interstate conflict processes. Nonetheless, the theory remains less than explicit on a number of central questions associated with war’s outbreak. As a synergistic theory of interstate conflict initiation, escalation and resolution, perfect deterrence theory provides additional precision to the insights of power transition.

Perfect deterrence theory lays out the implications of different threat characteristics for peaceful transitions in a variety of deterrence venues. Successful deterrence absolutely requires a capable threat. But the relationship between
capability and deterrence success is not monotonic. There is usually a point beyond which an increase in the cost of conflict does not increase the probability of successful deterrence. Indeed, to the extent that they are seen as less credible, more costly threats may even undermine the status quo.

By contrast, credible threats are neither necessary nor sufficient for peace. It is theoretically possible for conflict to be avoided when all threats are incredible, and for deterrence to fail when threats are credible all-around. The relationship between deterrence success and credible threats, therefore, is not necessarily straightforward. In general, however, greater credibility is associated with a higher probability of successful deterrence.

Not all deterrence situations are equally likely to succeed. For example, the status quo is potentially very stable in asymmetric or one-sided deterrence situations. Deterrence, however, is more tenuous when deterrence is mutual. Extended deterrence relationships are the most problematic. In all cases, though, satisfaction with the status quo is a critical strategic variable. As intuition suggests, in perfect deterrence theory, highly dissatisfied states are the most difficult to deter. In terms of specific strategies, reciprocal strategies offer the best prospects for peace. In contrast, unilateral, hard-line policies tend to result in conflict, either limited or all-out.

The view of deterrence that emerges from the theoretical confines of perfect deterrence theory has a number of important policy implications. For instance, perfect deterrence theory’s deductive framework supports both a policy of minimum deterrence and significant arms control initiatives while it opposes both an overkill capability and even the selective proliferation of weapons of mass destruction. As well, it prescribes conditionally cooperative diplomatic initiatives based on reciprocity. The descriptive hypotheses and policy prescriptions of the theory are supported by a number of systematic empirical studies.

CONTRIBUTORS

Frank C. Zagare is UB Distinguished Professor of Political Science at the University at Buffalo, SUNY. His theoretical work has focused on deterrence, crises, conflict escalation, and bargaining and negotiation. Dr. Zagare’s books include Perfect Deterrence (with D. Marc Kilgour), The Dynamics of Deterrence, and Game Theory: Concepts and Applications.

NOTES

1. For recent applications and extensions see, inter alia, Kugler and Lemke (2000) and Tammen et al. (2000).
2. Papers from this conference are collected in Kugler and Lemke (1996).
3. For reasons that will be apparent shortly, it is significant that DiCicco and Levy’s criticism of the power transition research program is confined to issues that revolve around the timing and initiation of conflict. They also see much that is progressive about the paradigm.

4. According to Goertz and Starr (2003), a set-theoretic definition is one among the several complementary logics that can be associated with the notion of a necessary condition.

5. See footnote 3.

6. In principle, it could also be used to explore economic rivalries and strategic relationships between non-state actors such as groups or individuals. Nonetheless, perfect deterrence theory’s primary focus has been on interstate security relationships.

7. For a more detailed discussion of the differences between perfect deterrence theory and classical deterrence theory, see Zagare (2004).

8. It is not difficult to show that a mutually worst outcome can never be part of a pure strategy equilibrium in any game with strict preference rankings over outcomes. For a more nuanced discussion of this point, see Zagare (2004, p. 112).

9. Waltz’s (1979) structural realism is a good example. It is generally acknowledged that Waltz’s theory is inconsistent. [See, for instance, Walt (1999, p. 17).] It should be no surprise, then, that Christensen and Snyder (1990, p. 138) find that “any foreign policy and its opposite can sometimes be deduced from Waltz’s theory.”

10. The Rudimentary Asymmetric Deterrence Game is a basic model, used primarily to define terms and illustrate concepts. The extensive-form representation of two core models of direct deterrence models are provided below. The most elaborate model, the Asymmetric Escalation Game, is not specifically discussed in this article. It particularly applies to situations of extended deterrence.

11. For explicit attempts to model the transition process, see Kim and Morrow (1992), Powell (1996a), and Alsharabati (1997).


13. Recall that in the analysis of the Rudimentary Asymmetric Deterrence Game Conflict was assumed to be the worst outcome for each player. Per force, this assumption renders Defender’s threat capable since Challenger is worse off if Defender executes the threat than Challenger would be if it simply decided not to contest the Status Quo. What is significant here is the fact that, given this assumption, the Status Quo is unstable (i.e., is inconsistent with rationality) in the Rudimentary Asymmetric Deterrence Game. It follows, therefore, that a capable threat is not a sufficient condition for deterrence success.

14. In the Generalized Mutual Deterrence Game each of two states, A and B, is trying to deter the other. In the Rudimentary Asymmetric Deterrence Game and the Unilateral Deterrence Game (see Fig. 3), one player (Defender) is trying to deter the other (Challenger) but not the other way around.

15. Perfect Deterrence Theory, therefore, is consistent with the argument (e.g., Trachtenberg, 1990/1991, p. 143) that, with different leaders or with different policies, conflicts like World War I can be avoided.

16. Legro and Moravcik, (1999, p. 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.

17. There is never a unique deterrence equilibrium in the Generalized Mutual Deterrence Game (see Figure 2).

18. For expository purposes I here use the term “deterrence” in the more traditional sense as a threat-based relationship. There is no good theoretical reason, however, to define it so narrowly. When broadly construed, attempts to stabilize the status quo through political reconciliation are part and parcel of “deterrence.”


20. This regularity was first offered in support of Bueno de Mesquita’s expected utility theory. This does not, however, diminish its significance for perfect deterrence theory.

21. As well, Harvey’s (1998, p. 691) empirical analysis “indirectly supports [perfect deterrence theory’s] claim about the crucial role of capabilities” in deterrence relationships.

22. See also Press (2005).

23. See also Quackenbush (2005).
REFERENCES


