

# Deterrence theory: where do we stand?

STEPHEN L. QUACKENBUSH\*

**Abstract.** Although deterrence theory was a central focus in the study of International Relations during the Cold War, attention has shifted away from deterrence since the end of that conflict. Nonetheless, deterrence is a general phenomenon that is not limited to any particular time or space. Moving beyond a simple focus on the US-Soviet relationship, scholars have recently begun further explorations of deterrence, through development of theory, analysis of policy alternatives, and empirical analysis. This article seeks to evaluate where deterrence theory stands today through: (1) a consideration of distinctions between different strands of theory; (2) a discussion of the assumption of rationality in deterrence theory; (3) an examination of three important distinctions in deterrence; (4) an evaluation of the difficult task of testing deterrence theory, and (5) an overview of recent theoretical developments. The primary conclusion is that perfect deterrence theory provides a logically consistent alternative to classical deterrence theory and therefore provides the most appropriate basis for further theoretical development, empirical testing, and application to policy.

**Stephen L. Quackenbush** is Assistant Professor of Political Science at the University of Missouri. His research focuses on deterrence, recurrent conflict, and international conflict. Stephen can be contacted at: {quackenbushs@missouri.edu}.

The bipolar confrontation between the US and Soviet Union during the Cold War prompted statesmen and scholars alike to seek ways to prevent nuclear holocaust. Many argued that this was best done through deterrence, and accordingly much attention was devoted to exploring the dynamics of deterrence, particularly from about 1960–1990. However, the end of the Cold War brought about by the fall of Communism in Eastern Europe and the collapse of the Soviet Union led to a lessened tension between the US and Russia. In response, many believed deterrence to be irrelevant in the post-Cold War world,<sup>1</sup> and attention shifted to topics such as the democratic peace, globalisation, and international terrorism.

Nonetheless, deterrence – the use of a threat (explicit or not) by one party in an attempt to convince another party to maintain the *status quo* – is a general phenomenon that is not limited to any particular time or space.<sup>2</sup> Moving beyond a simple focus on the US-Soviet relationship, scholars have recently begun further

\* The author would like to thank Frank Zagare, Kate Floros, Paul Bellinger, and the anonymous reviewers for their helpful comments and suggestions.

<sup>1</sup> For example, see Charles Krauthammer, 'The Obsolescence of Deterrence', *The Weekly Standard*, 8:13 (9 December 2002).

<sup>2</sup> Frank C. Zagare, 'Deterrence is Dead. Long Live Deterrence', *Conflict Management and Peace Science*, 23 (2006), pp. 115–20.

explorations of deterrence, through development of theory,<sup>3</sup> analysis of policy alternatives,<sup>4</sup> and empirical analysis.<sup>5</sup>

This article seeks to evaluate where deterrence theory stands today, focusing primarily on recent developments. Specifically, what follows is an examination of several important topics in deterrence theory. The article begins with a consideration of the distinction between classical deterrence theory and perfect deterrence theory, and then turns to a discussion of the assumption of rationality in deterrence. Following these initial steps is an examination of three important categorisations of deterrence: unilateral versus mutual deterrence, conventional versus nuclear deterrence, and general versus immediate deterrence. After that, the difficult task of testing deterrence theory and key avenues of recent theoretical developments are examined before turning to final conclusions.

### Theories of deterrence

An evaluation of where deterrence theory stands must start with a consideration of what deterrence theory is. Deterrence theory – often called rational deterrence theory – argues that, in order to deter attacks, a state must persuade potential attackers that: 1) it has an effective military capability; 2) that it could impose unacceptable costs on an attacker, and 3) that the threat would be carried out if attacked.<sup>6</sup> And so deterrence theorists, both early<sup>7</sup> and more recently<sup>8</sup> have attempted to explain deterrence through several key elements: ‘the assumption of a very severe conflict, the assumption of rationality, the concept of a retaliatory

<sup>3</sup> For example, Frank C. Zagare and D. Marc Kilgour, *Perfect Deterrence* (Cambridge: Cambridge University Press, 2000); Patrick M. Morgan, *Deterrence Now* (Cambridge: Cambridge University Press, 2003); Anne E. Sartori, *Deterrence by Diplomacy* (Princeton: Princeton University Press, 2005); Stephen L. Quackenbush, ‘Not only Whether but Whom: Three-party Extended Deterrence’, *Journal of Conflict Resolution*, 50:4 (2006), pp. 562–83.

<sup>4</sup> For example, Stephen J. Cimbala (ed.), *Deterrence and Nuclear Proliferation in the Twenty-first Century* (Westport, CT.: Praeger, 2001); Max G. Manwaring (ed.), *Deterrence in the 21st Century* (London: Frank Cass, 2001); Robert Powell, ‘Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense’, *International Security*, 27 (2003), pp. 86–118; Lawrence Freedman, *Deterrence* (Cambridge: Polity, 2004); Stephen L. Quackenbush, ‘National Missile Defense and Deterrence’, *Political Research Quarterly*, 59:4 (2006), pp. 533–41.

<sup>5</sup> Vesna Danilovic, *When the Stakes Are High* (Ann Arbor: University of Michigan Press, 2002); Stephen L. Quackenbush, ‘General Deterrence and International Conflict: Bridging the Formal/Quantitative Divide’ (Ph.D. diss., University at Buffalo, SUNY, 2003); Curtis S. Signorino and Ahmer Tarar, ‘A Unified Theory and Test of Extended Immediate Deterrence’, *American Journal of Political Science*, 50 (2006), pp. 586–605.

<sup>6</sup> William Kaufmann, ‘The Requirements of Deterrence’, in William Kaufmann (ed.), *Military Policy and National Security* (Princeton: Princeton University Press, 1956).

<sup>7</sup> For example, Bernard Brodie, *Strategy in the Missile Age* (Princeton: Princeton University Press, 1959); Thomas C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960); Thomas C. Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966); Herman Kahn, *On Thermonuclear War* (Princeton: Princeton University Press, 1961); Herman Kahn, *On Escalation: Metaphors and Scenarios* (New York: Praeger, 1965); Glenn Snyder, *Deterrence and Defense* (Princeton: Princeton University Press, 1961).

<sup>8</sup> For example, Richard Ned Lebow, *Between Peace and War: The Nature of International Crises* (Baltimore: Johns Hopkins University Press, 1981); Frank C. Zagare, *The Dynamics of Deterrence* (Chicago: University of Chicago Press, 1987); Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility* (Cambridge: Cambridge University Press, 1990); Zagare and Kilgour, *Perfect Deterrence*.

threat, the concept of unacceptable damage, the notion of credibility, and the notion of deterrence stability.<sup>9</sup>

Morgan argues that all of these efforts are explorations of the same theory.<sup>10</sup> While seemingly disparate in both time and approach, all focus on similar issues and work from similar assumptions. For Morgan, while there may be different deterrence *strategies* there is only one deterrence *theory*.

Morgan's position notwithstanding, the question of whether the distinctions between different variants of deterrence theory are significant enough to consider them separate theories remains open. Zagare argues that much of the deterrence literature can indeed be categorised as a single theory: *classical deterrence theory*.<sup>11</sup> However, because theoretical approaches and assumptions vary widely from one theorist to another, Zagare also divides classical deterrence theory into two sub-groups: *structural deterrence theory* and *decision-theoretic deterrence theory*.<sup>12</sup>

Structural deterrence theory, closely aligned with realism, argues that a balance of power brings peace; if two states are equal in power, each will be deterred since neither will be able to gain an advantage.<sup>13</sup> Furthermore, structural deterrence theory argues that nuclear deterrence is inherently stable. While 'in a conventional world, a country can sensibly attack if it believes that success is probable', with nuclear weapons 'a nation will be deterred from attacking even if it believes that there is only a possibility that its adversary will retaliate'.<sup>14</sup> Thus, the key to deterrence is a second-strike capability, and once this is achieved,<sup>15</sup> deterrence is straightforward since the enormous costs associated with nuclear war make an attack irrational. Accidental war, then, represents the only real threat to deterrence.<sup>16</sup>

Decision-theoretic deterrence theorists, on the other hand, utilise expected utility and game theory to construct models of deterrence.<sup>17</sup> While their approach is quite different, decision-theoretic deterrence theorists take the structural deterrence theorists' idea that nuclear war is irrational to heart by assuming that conflict is always the worst outcome.

<sup>9</sup> Morgan, *Deterrence Now*, p. 8.

<sup>10</sup> Morgan, *Deterrence Now*.

<sup>11</sup> Frank C. Zagare, 'Classical Deterrence Theory: A Critical Assessment', *International Interactions*, 21 (1996), pp. 365–87.

<sup>12</sup> Classical deterrence theory is a large and diverse literature spanning more than five decades. I focus here on the primary arguments, not all the nuances of individual works. For a thorough review of classical deterrence theory, see Zagare, 'Classical Deterrence Theory'.

<sup>13</sup> For example, Kaufmann, 'The Requirements of Deterrence'; Bernard Brodie, 'The Anatomy of Deterrence', *World Politics*, 11 (1959), pp. 173–9; Snyder, *Deterrence and Defense*; Michael D. Intriligator and Dagobert L. Brito, 'Can Arms Races Lead to the Outbreak of War?', *Journal of Conflict Resolution*, 28 (1984), pp. 63–84; Michael D. Intriligator and Dagobert L. Brito, 'The Stability of Mutual Deterrence', in Jacek Kugler and Frank C. Zagare (eds), *Exploring the Stability of Deterrence* (Boulder: Lynne Rienner, 1987), pp. 13–39; John J. Mearsheimer, 'Back to the Future: Instability in Europe After the Cold War', *International Security*, 15 (1990), pp. 5–56.

<sup>14</sup> Kenneth N. Waltz, 'The Origins of War in Neorealist Theory', *Journal of Interdisciplinary History*, 18 (1988), p. 626.

<sup>15</sup> Albert Wohlstetter, 'The Delicate Balance of Terror', *Foreign Affairs*, 37 (1959), pp. 211–34.

<sup>16</sup> Accidental war is seen as both a threat to stability and a tactic to be taken advantage of. This tactic and a more detailed discussion of accidental war are discussed below.

<sup>17</sup> For example, Schelling, *Strategy of Conflict*; Schelling, *Arms and Influence*; Steven J. Brams, and D. Marc Kilgour, *Game Theory and National Security* (New York: Basil Blackwell, 1988); Powell, *Nuclear Deterrence Theory*; Powell, 'Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense'.

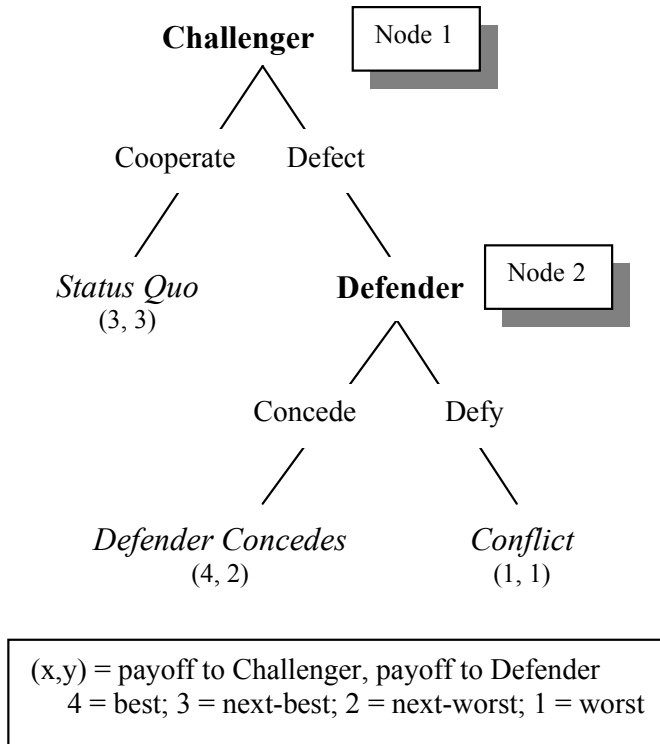


Figure 1. 'Classical' deterrence game.

A typical 'classical' model of deterrence is shown in Figure 1. In this game, there are two actors: a Challenger, who seeks to alter the *status quo*, and a Defender that seeks to deter such challenges. Challenger begins the game by deciding whether to cooperate, in which case the *Status Quo* remains, or attack Defender by choosing to defect. Defender, in turn, has two possible responses to a challenge: to concede, resulting in *Defender Concedes*, or to defy, resulting in *Conflict*. Challenger would most prefer *Defender Concedes*, and Defender likes *Status Quo* the best. However, they both agree that *Conflict* is the worst possible outcome.

Defender faces a quandary: a concession provides a more favourable outcome than the conflict that would follow defiance, but if Challenger knows that Defender will concede, Challenger will always attack and deterrence will always fail. Zagare and Kilgour label this dilemma the 'paradox of mutual deterrence'.<sup>18</sup> Classical deterrence theory offers two primary solutions.

The first proposed solution is for Defender to make an irrevocable commitment to a hard-line strategy. Such irrevocable commitments are made by burning bridges to limit one's options by eliminating the ability to back down even though that would be the preferred alternative.<sup>19</sup> So Defender makes an irrevocable commitment

<sup>18</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>19</sup> Herman Kahn, *Thinking About the Unthinkable* (New York: Horizon Press, 1962); Schelling, *Arms and Influence*.

to defy, and must communicate this commitment to Challenger.<sup>20</sup> Then Challenger is faced with the choice of remaining at the *Status Quo* by cooperating or defecting and starting a *Conflict*. Since choices of cooperate by Challenger and defy by Defender are mutual best responses, this strategy pair denotes a Nash equilibrium with *Status Quo* as the outcome. The problem, however, is that if Challenger does attack, then Defender's only rational response would be to concede, leading to *Defender Concedes*. Although the strategy pair (cooperate, defy) is a Nash equilibrium, it is not sub-game perfect because it involves an irrational choice by Defender off the equilibrium path.<sup>21</sup> The only sub-game perfect equilibrium is the strategy pair (defect, concede), which results in *Defender Concedes*.<sup>22</sup>

The second solution to this dilemma offered by classical deterrence theorists is threats that leave something to chance.<sup>23</sup> These threats allow Defender to circumvent the problem of irrational action by threatening to take action 'that raises the risk that the situation will go out of control and escalate to a catastrophic nuclear exchange'.<sup>24</sup> Thus, rather than relying upon a threat to make an irrational choice for war, Defender can simply make a rational choice to raise the risk of war and leave the question of whether war starts or not to chance. Powell demonstrates that threats that leave something to chance not only lead to successful deterrence, but the resulting equilibrium is perfect as well.<sup>25</sup>

However, the idea of threats that leave something to chance rests upon the possibility of accidental war. There is a lot of speculation about this possibility,<sup>26</sup> with the outbreak of World War I often cited as the prime example of an inadvertent war.<sup>27</sup> However, Trachtenberg conducts an extensive examination of the coming of the First World War and concludes that

when one actually tests these propositions against the empirical evidence, which for the July Crisis is both abundant and accessible, one is struck by how weak most of the arguments turn out to be. The most remarkable thing about all these claims that support the conclusion about events moving 'out of control' in 1914 is how little basis in fact they actually have.<sup>28</sup>

The prime example of accidental war turns out to be not such a good example after all. Rather than support the idea, the outbreak of World War I actually undermines the idea of threats that leave something to chance.

<sup>20</sup> The movie *Dr. Strangelove* highlights the importance of communicating irrevocable commitments nicely.

<sup>21</sup> Nash equilibria only require rational choices along the equilibrium path, whereas sub-game perfect equilibria require every possible decision to be rational. For a discussion of the distinction between the two equilibrium concepts, see James D. Morrow, *Game Theory for Political Scientists* (Princeton: Princeton University Press, 1994).

<sup>22</sup> Note that this latter equilibrium is also a Nash equilibrium.

<sup>23</sup> Schelling, *Strategy of Conflict*.

<sup>24</sup> Powell, 'Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense', p. 90.

<sup>25</sup> Powell, *Nuclear Deterrence Theory*; Powell, 'Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense'.

<sup>26</sup> For example, Schelling, *Arms and Influence*; Paul J. Bracken, *The Command and Control of Nuclear Forces* (New Haven: Yale University Press, 1983); Bruce G. Blair, *The Logic of Accidental Nuclear War* (Washington: Brookings, 1993); Scott D. Sagan, *The Limits of Safety* (Princeton: Princeton University Press, 1993).

<sup>27</sup> For example, Bracken states that the idea of accidental nuclear war would sound unrealistic 'were it not for the history of the outbreak of World War I'. See *Command*, p. 65.

<sup>28</sup> Marc Trachtenberg, *History and Strategy* (Princeton: Princeton University Press, 1991), p. 99. See also Frank C. Zagare, 'Explaining the 1914 War in Europe: An Analytic Narrative', *Journal of Theoretical Politics*, 21 (2009), pp. 63–95.

Schelling suggests that in most crises, one side or another will be willing to run greater risks of mutual assured destruction to achieve its goals. Therefore, credibility is determined by whose interests are more greatly threatened by an ongoing crisis and by who is more willing to take risks to protect their interests. Hence, deterrence becomes a ‘competition in risk taking’ as states employ brinksmanship in what the subtitle to Powell’s 1990 book calls ‘the search for credibility’. But a threat is said to be credible if it is believed.<sup>29</sup> Given that a nuclear attack invites one’s own destruction, the threat to choose to do so is not believable, and is thus not credible. Schelling argues that while this is true, the threat to increase the risk of inadvertent war can in fact be believable.<sup>30</sup> But since historical evidence shows that World War I – the prime example of ‘accidental war’ – arose as a result of conscious decisions, not chance,<sup>31</sup> these threats that leave something to chance seem to be not credible after all.

Classical deterrence theory and its associated ideas such as mutual assured destruction and brinksmanship represent the conventional wisdom about deterrence. Zagare and Kilgour offer an alternative to classical deterrence theory, which they call *perfect deterrence theory*.<sup>32</sup> In particular, they depart from classical deterrence theory in their view of credibility. Zagare and Kilgour argue that threats are believable, and thus credible, when they are rational to carry out.<sup>33</sup> Connecting credibility with rationality in this way is consistent with the treatment of credibility in game theory.<sup>34</sup>

This connection between credibility and rationality can be seen by re-examining the game in Figure 1. If Defender prefers *Defender Concedes* to *Conflict*, and this is known to Challenger, Challenger has no reason whatsoever to believe that Defender will carry out her threat; thus Defender’s threat is not credible. On the other hand, if Defender prefers *Conflict* to *Defender Concedes*, and this is known to Challenger, Challenger would believe Defender’s threat; that is, Defender’s threat would be credible. The simple deterrence game with a credible threat by Defender is shown in Figure 2.

In this game, Defender will choose to defy at node 2 because a concession would result in her least preferred outcome. Knowing this, Challenger will cooperate at node 1 because the outcome that results (*Status Quo*) is preferred to the outcome that results from defection (*Conflict*). Thus, if Defender has a credible threat, *Status Quo* is the sole equilibrium outcome.

Accordingly, the solution to the paradox of mutual deterrence is the presence of credible threats.<sup>35</sup> When both sides possess credible and capable threats,

<sup>29</sup> Schelling, *Arms and Influence*; Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974); Robert Jervis, ‘Introduction: Approach and Assumptions’, in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein (eds), *Psychology and Deterrence* (Baltimore: Johns Hopkins University Press, 1985), pp. 1–12.

<sup>30</sup> Schelling, *Strategy of Conflict*.

<sup>31</sup> Trachtenberg, *History and Strategy*; Zagare, ‘Explaining the 1914 War in Europe’.

<sup>32</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>33</sup> Also see Lebow, *Between Peace and War*; Richard K. Betts, *Nuclear Blackmail and Nuclear Balance* (Washington: Brookings, 1987); Richard Smoke, *National Security and the Nuclear Dilemma* (Reading: Addison-Wesley, 1987).

<sup>34</sup> For example, Reinhard Selten, ‘A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games’, *International Journal of Game Theory*, 4 (1975), pp. 25–55; Eric Rasmusen, *Games and Information* (Oxford: Basil Blackwell, 1989); Robert Gibbons, *Game Theory for Applied Economists* (Princeton: Princeton University Press, 1992).

<sup>35</sup> Zagare and Kilgour, *Perfect Deterrence*.

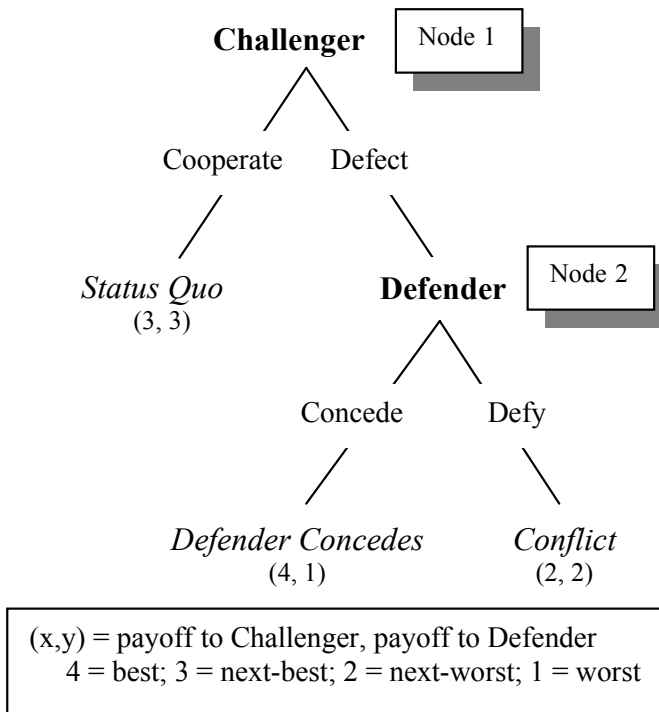


Figure 2. Simple deterrence game with a credible threat.

*Status Quo* emerges as the rational outcome. Furthermore, the resulting equilibrium is sub-game perfect, and it does not rest upon untenable assumptions of the possibility for accidental war.

Zagare and Kilgour label their theory 'perfect' because of their insistence on the use of perfect equilibria. This insistence stems from the observation that 'perfectness rules out threats that are not credible'.<sup>36</sup> However, Powell has also employed perfect equilibria, but his work certainly falls within classical, rather than perfect, deterrence theory.<sup>37</sup> The real hallmark of perfect deterrence theory is the insistence that credibility varies, and that credibility is determined by a state's preference between conflict and backing down.<sup>38</sup>

While Morgan argues that there is only one deterrence theory, it is clear that there are at least two: classical and perfect.<sup>39</sup> While classical deterrence theory is

<sup>36</sup> Rasmusen, *Games and Information*, p. 87.

<sup>37</sup> Powell, *Nuclear Deterrence Theory*; Powell, 'Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense'.

<sup>38</sup> I focus here on complete information conditions, where each actor's preferences are common knowledge. Zagare and Kilgour also examine situations of incomplete information, where each actor forms beliefs about its opponent's preference between conflict and backing down. In that case, a state's credibility results from its opponent's belief about its type. While the details are different, the implications are quite similar. See Zagare and Kilgour, *Perfect Deterrence*.

<sup>39</sup> Morgan, *Deterrence Now*. However, this does not mean that there are only two. For example, Arquilla and Davis develop what could be considered a separate theory of deterrence in a series of studies for RAND. See Paul K. Davis and John Arquilla, *Thinking About Opponent Behavior in Crisis and Conflict*, N-3322-JS (Santa Monica: RAND, 1991); Paul K. Davis and John Arquilla, *Detering or Coercing Opponents in Crisis*, R-4111-JS (Santa Monica: RAND, 1991); John Arquilla

rooted in the basic assumption that the high costs of nuclear war make conflict the worst outcome for everyone, perfect deterrence theory is rooted in the assumption that different states have different preferences. While some may indeed prefer backing down to fighting, others prefer to fight and only these latter states have credible threats. Although there are other differences between the two theories,<sup>40</sup> this one assumption makes a tremendous impact on the predictions and explanations offered.

### Assumption of rationality in deterrence theory

But what about the assumption of rationality? Both classical and perfect deterrence theory are rooted in the assumption of rationality, which can be defined as

gaining as much information as possible about the situation and one's options for dealing with it, calculating the relative costs and benefits of those options as well as their relative chances of success and risks of disaster, then selecting – in light of what the rational opponent would do – the course of action that promised the greatest gain or, if there would be no gain, the smallest loss.<sup>41</sup>

Morgan goes to great lengths discussing limits of rational behaviour. These limits arise because decision makers lack sufficient time to analyse all alternatives in a crisis situation, they lack information about the opponent and the consequences of decisions, and/or they are affected by emotions or cognitive limitations. These limitations to rational decision-making processes have been previously well established.<sup>42</sup>

Furthermore, deterrence can fail even when states appear rational or succeed despite irrationality. Thus, Morgan laments that 'rationality is an inconsistent guide to how deterrence turns out'.<sup>43</sup> For these reasons, Morgan concludes that the assumption of rationality is unnecessary for the development of deterrence theory and that an alternative theory needs to move away from reliance upon it.

But Morgan and others fail to account for the distinction between *procedural* and *instrumental* rationality. Morgan defines procedural rationality, and it is these procedures that he and others focus upon in determining whether or not states are 'really' rational.<sup>44</sup> However, the assumption used in rational deterrence theory – as in rational choice theory more broadly – is instrumental rationality.<sup>45</sup>

and Paul K. Davis, *Extended Deterrence, Compellence, and the 'Old World Order'*, N-3482-JS (Santa Monica: RAND, 1992).

<sup>40</sup> Frank C. Zagare, 'Reconciling Rationality with Deterrence: A Re-examination of the Logical Foundations of Deterrence Theory', *Journal of Theoretical Politics*, 16 (2004), pp. 107–41.

<sup>41</sup> Morgan, *Deterrence Now*, p. 12.

<sup>42</sup> See, for example, Graham T. Allison, *Essence of Decision* (Boston: Little, Brown, 1971); Lebow, *Between Peace and War*; Robert Jervis, Richard Ned Lebow, and Janice Gross Stein (eds), *Psychology and Deterrence* (Baltimore: Johns Hopkins University Press, 1985).

<sup>43</sup> Morgan, *Deterrence Now*, p. 64.

<sup>44</sup> For example, Richard Ned Lebow and Janice Gross Stein, 'Rational Deterrence Theory: I Think, Therefore I Deter', *World Politics*, 41 (1989), pp. 208–24; Richard Ned Lebow and Janice Gross Stein, 'Deterrence: The Elusive Dependent Variable', *World Politics*, 42 (1990), pp. 336–69; Edward Rhodes, *Power and MADness: The Logic of Nuclear Coercion* (New York: Columbia University Press, 1989).

<sup>45</sup> Frank C. Zagare, 'Rationality and Deterrence', *World Politics*, 42 (1990), pp. 238–60; Stephen L. Quackenbush, 'The Rationality of Rational Choice Theory', *International Interactions*, 30 (2004), pp. 87–107.



An instrumentally rational actor is one who, when confronted with ‘two alternatives which give rise to outcomes [...] will choose the one which yields the more preferred outcome’.<sup>46</sup> Thus, Morgan is quite correct to state that ‘what is rational for an actor depends on the actor’s preferences’. However, his further claim that ‘knowing preferences is not necessarily enough [...] Rationality is not simply acting out one’s preferences or objectives – it is arriving at that action by choosing in a specified way’<sup>47</sup> is not in accord with instrumental rationality. If an actor chooses according to her preferences, then she is instrumentally rational, regardless of which procedures are used. And since these preferences are subjective in nature, emotions, cognitive limitations, and the like may shape preferences but do not make an actor irrational.<sup>48</sup>

Of course it can be quite difficult to determine an actor’s subjectively held preferences with certainty. Accordingly, determining the rationality of any action can be quite difficult as well. To determine that an action is irrational, one must prove that the actor *deliberately* chose contrary to his *subjective* preferences given the situation *as he saw it*. On the other hand, determining that an action is rational requires proof that the actor chose according to her preferences given the situation *as she saw it*. I would argue that either task is immensely difficult if not impossible. And this is precisely why it is *assumed*. It is thus no surprise to a rational deterrence theorist that rationality is an inconsistent guide to how deterrence turns out because since: 1) all actions are rational (by assumption) and, 2) deterrence sometimes succeeds and sometimes fails (by observation), this point is obvious.

One final point regarding rationality deserves consideration. Morgan correctly points out that, in seeking solutions to the paradox of mutual deterrence discussed in the previous section, classical deterrence theorists such as Schelling and Kahn move away from rationality in discussions of strategic ploys such as ‘feigning irrationality’. These contradictions introduce logical inconsistency in to classical deterrence theory – the very thing that rational choice theory is used to avoid.<sup>49</sup> But perfect deterrence theory demands consistent use of the rationality assumption and thus provides a logically consistent alternative.

### Categorisations of deterrence

I now turn to a consideration of three common categorisations of deterrence, starting with the distinction between unilateral and mutual deterrence.

#### *Unilateral vs. mutual deterrence*

Deterrence deals with states’ attempts to prevent others from challenging the *status quo*. The need for deterrence to prevent these challenges depends upon whether the

<sup>46</sup> R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley, 1957), p. 50.

<sup>47</sup> Morgan, *Deterrence Now*, p. 65.

<sup>48</sup> R. Harrison Wagner, ‘Rationality and Misperception in Deterrence Theory’, *Journal of Theoretical Politics*, 4 (1992), pp. 115–41.

<sup>49</sup> Bruce Bueno de Mesquita and James D. Morrow, ‘Sorting Through the Wealth of Notions’, *International Security*, 24 (Autumn 1999), pp. 56–73.

state desires alterations to the *status quo*. Deterrence is not needed to prevent a state that is satisfied with the *status quo* from initiating a challenge. Deterrence is only relevant when states seek alterations in the *status quo*. If neither state in a dyad seeks to alter the *status quo*, then there is no deterrence. If one state seeks to alter the *status quo* but the other does not, there is unilateral deterrence. And finally, a situation where both sides of a dyad seek alterations to the *status quo* is one of mutual deterrence.

Perfect deterrence theory models both mutual and unilateral deterrence situations and demonstrates that deterrence is more stable in unilateral than mutual deterrence.<sup>50</sup> Senese and Quackenbush apply this logic to an analysis of settlements to militarised interstate disputes.<sup>51</sup> They argue that imposed settlements lead to unilateral deterrence whereas negotiated settlements and disputes ending without a settlement require mutual deterrence to maintain peace. Accordingly, they expect imposed settlements to be the most stable, and their empirical analysis strongly supports this conclusion.

Given strong empirical support for the theoretical prediction that unilateral deterrence is more stable than mutual deterrence, it would seem to make sense for states to desire an 'escape from mutual deterrence'.<sup>52</sup> Since power is satisfying<sup>53</sup> a state may be able to move from mutual to unilateral deterrence by increasing its power relative to its counterpart. Morgan argues that this implies 'that a challenger is irrational enough to let the deterrer achieve unilateral deterrence – the deterrer is rational in trying to escape mutual deterrence while the challenger is irrational enough to let this happen'.<sup>54</sup>

But a move from mutual to unilateral deterrence does not have such implications. Power is not the same thing as satisfaction.<sup>55</sup> Therefore, if the deterrer wants to escape mutual deterrence, then he only needs to no longer desire an alteration of the *status quo* – no increases in power are necessary. And even if the deterrer does increase in power dramatically, this would not at all imply irrational action by the challenger, since such an increase does not depend on any action by the challenger whatsoever.

Therefore, Morgan's concerns that attempts to escape from mutual deterrence could cause instability are not logically supported. Unilateral deterrence is more stable than mutual deterrence, all else being equal. If a state wants to escape from mutual deterrence, it only needs to decide that it is satisfied with the *status quo* and therefore no longer needs to be deterred.

While the distinction between unilateral and mutual deterrence rests on the interests of the states involved, the next distinction is based upon the weapons possessed by these states.

<sup>50</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>51</sup> Paul D. Senese and Stephen L. Quackenbush, 'Sowing the Seeds of Conflict: The Effect of Dispute Settlements on Durations of Peace', *Journal of Politics*, 65 (2003), pp. 696–717.

<sup>52</sup> Morgan, *Deterrence Now*, p. 51.

<sup>53</sup> John R. Oneal, Indra de Soysa, and Yong-Hee Park, 'But Power and Wealth Are Satisfying: A Reply to Lemke and Reed', *Journal of Conflict Resolution*, 42 (1998), pp. 517–20.

<sup>54</sup> Morgan, *Deterrence Now*, p. 53.

<sup>55</sup> Douglas Lemke and William Reed, 'Power Is Not Satisfaction: A Comment on de Soysa, Oneal, and Park', *Journal of Conflict Resolution*, 42 (1998), pp. 511–16.

*Nuclear and conventional deterrence*

Deterrence theory was brought to prominence in academic and policymaking circles because of the threat of nuclear holocaust during the Cold War. Therefore, many scholars have focused on the dynamics of nuclear deterrence.<sup>56</sup> Others have focused specifically on conventional deterrence.<sup>57</sup> Support for this analytical distinction rests on the notion that ‘nuclear weapons dissuade states from going to war much more surely than conventional weapons do’.<sup>58</sup>

However, this expectation regarding the stabilising effects of nuclear weapons is contradicted both empirically and logically. Empirically, although classical deterrence theorists claim that nuclear weapons are inherently stabilising, there is evidence that non-nuclear opponents of nuclear powers do not appear cautious or constrained in their hostile activity.<sup>59</sup> Furthermore, the possession of nuclear weapons does not appear to impede escalatory behaviour by non-nuclear opponents.<sup>60</sup>

Logical problems stem from classical deterrence theorists’ claim that nuclear warfare entails such high costs that it is the worst possible outcome for all sides. It is precisely the assumption that conflict is the worst possible outcome that leads to the paradox of mutual deterrence discussed previously. If conflict – nuclear or conventional – indeed is always the worst possible outcome, deterrence *cannot ever* be successful. Any state that is attacked will always capitulate rather than bring about its own worst outcome, and knowing this, challengers will always attack.

One might think that policymakers would either have to have nerves of steel or brains of lead to launch a massive nuclear attack against a similarly armed adversary and expect that the adversary would do nothing as it faces total destruction. After all, what leader would *not* retaliate in response to a nuclear attack? In other words, it is hard to imagine that a state would prefer backing down after being attacked by nuclear weapons to retaliation and all-out nuclear war. This highlights the problem with the use of Chicken as a model of deterrence, for it stipulates that backing down is indeed preferred to fighting back. And classical deterrence theory assumes not only that some states hold this preference, but that all do.

The only logically consistent resolution to this paradox is the existence of mutually credible threats, where each state prefers to fight rather than back down.<sup>61</sup> Accordingly, if classical deterrence theory is correct that nuclear weapons

<sup>56</sup> For a review of the literature on nuclear deterrence, see Frank Harvey and Patrick James, ‘Nuclear Deterrence Theory: The Record of Aggregate Testing and an Alternative Research Agenda’, *Conflict Management and Peace Science*, 12 (1992), pp. 17–45.

<sup>57</sup> For example, John J. Mearsheimer, *Conventional Deterrence* (Ithaca: Cornell University Press, 1983); Johnathan Shimshoni, *Israel and Conventional Deterrence* (Ithaca: Cornell University Press, 1988); George H. Quester, *Deterrence Before Hiroshima* (New Brunswick: Transaction, 1986); Peter Karsten, Peter D. Howell, and Artis Frances Allen, *Military Threats: A Systematic Historical Analysis of the Determinants of Success* (Westport, CT.: Greenwood, 1984); Barry Wolf, *When the Weak Attack the Strong: Failures of Deterrence*, N-3261-A (Santa Monica: RAND, 2001).

<sup>58</sup> Waltz, ‘Origins of War’, p. 625.

<sup>59</sup> A. F. K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980).

<sup>60</sup> Daniel S. Geller, ‘Nuclear Weapons, Deterrence, and Crisis Escalation’, *Journal of Conflict Resolution*, 34 (1990), pp. 291–310; T. V. Paul, *Asymmetric Conflicts* (Cambridge: Cambridge University Press, 1994).

<sup>61</sup> Zagare and Kilgour, *Perfect Deterrence*.

make conflict the worst possible outcome – and thus, less preferable than backing down – then they are destabilising, not stabilising. If classical deterrence theory is not correct in this assumption, then the fundamental idea that deterrence is a game of Chicken falls apart. In any case, Zagare and Kilgour demonstrate that, beyond a certain threshold, the ability to impose additional costs has no effect on the stability of deterrence. Thus, as Morgan argues, ‘nuclear deterrence [...] is given too much credit for the long peace’.<sup>62</sup>

To draw conclusions regarding the distinction between nuclear and conventional deterrence, one can focus on three themes. First, the concept of deterrence is broad and is not limited to either nuclear or conventional conflicts. Secondly, states have an interest in deterring both conventional and nuclear conflicts. And finally, classical deterrence theory’s claims about nuclear deterrence – which are the basis of the analytic distinction between nuclear and conventional deterrence – are contradicted both logically and empirically. Put together, these points indicate that deterrence theory should focus on general explanations of the dynamics of deterrence, rather than limited explanations of nuclear or conventional deterrence.

One final categorisation to be reviewed here is the distinction between general and immediate deterrence.

#### *General vs. immediate deterrence*

Morgan highlights the importance of the distinction between general deterrence and immediate deterrence.<sup>63</sup> Immediate deterrence ‘concerns the relationship between opposing states where at least one side is seriously considering an attack while the other is mounting a threat of retaliation in order to prevent it.’ Conversely, general deterrence ‘relates to opponents who maintain armed forces to regulate their relationship even though neither is anywhere near mounting an attack’.<sup>64</sup> Thus, general deterrence has less to do with ‘crisis decision making’ than with everyday decision-making in relationships involving conflicts of interest.

General deterrence is much broader than immediate deterrence. For example, consider a case of immediate deterrence such as the Cuban Missile Crisis. Studies of immediate deterrence seek to understand how escalation can be controlled within the context of a crisis. That is, once a state has already challenged the *status quo*, how can a defending state deter the challenger from taking further action and thus avoid all-out war? Successful immediate deterrence entails a challenger’s backing down following the defender’s threat to retaliate, whereas the failure of immediate deterrence results in the challenger’s attacking despite the defender’s retaliatory threat.

The need for immediate deterrence indicates that general deterrence has previously failed.<sup>65</sup> If general deterrence always succeeds, crises and wars do not

<sup>62</sup> Morgan, *Deterrence Now*, p. 133.

<sup>63</sup> Patrick M. Morgan, *Deterrence: A Conceptual Analysis*, 2nd edition (Beverly Hills: Sage, 1983), and Morgan, *Deterrence Now*.

<sup>64</sup> Morgan, *Deterrence*, p. 30.

<sup>65</sup> Vesna Danilovic, ‘Conceptual and Selection Bias Issues in Deterrence’, *Journal of Conflict Resolution*, 45 (2001), pp. 97–125.

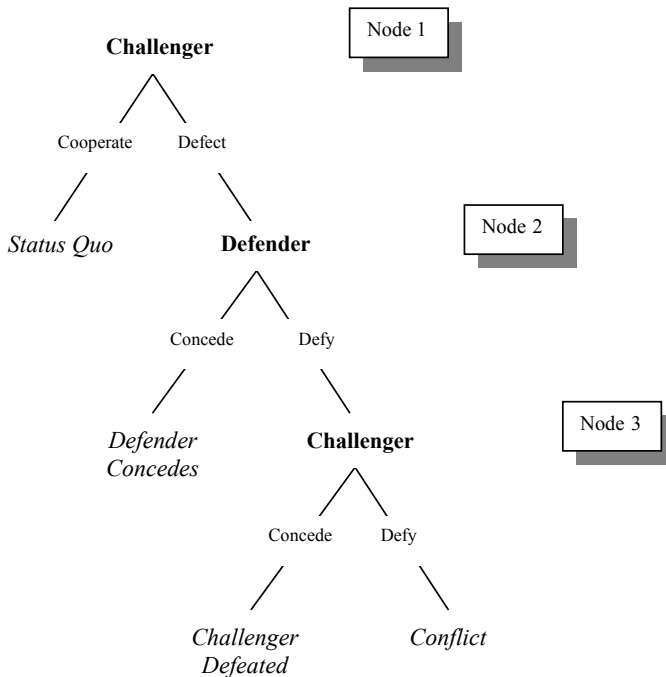


Figure 3. *Unilateral deterrence game.*

occur. Stated differently, the successful operation of general deterrence precludes the existence of immediate deterrence. It would seem appropriate, then, to focus on the origins of international crises before examining management of those crises. Since general deterrence necessarily precedes immediate deterrence, the analysis of general deterrence is more important for a general understanding of international conflict than the analysis of immediate deterrence.<sup>66</sup> Furthermore, as the literature on selection bias makes clear, examination of immediate deterrence without consideration of the origins of immediate deterrence cases (that is, the failure of general deterrence) can produce misleading empirical results.<sup>67</sup>

Accordingly, what is really needed in deterrence theory is a theory of general deterrence because 'general deterrence is a situation much more typical of international politics' and immediate deterrence is 'a type of situation that seldom exists'.<sup>68</sup> But Morgan argues that 'at best we have fragments of a theory of general deterrence'.<sup>69</sup>

However, perfect deterrence theory is a theory of general deterrence. This can easily be seen by examining the Unilateral Deterrence Game shown in Figure 3.<sup>70</sup> There are two players, Challenger and Defender. At node 1, Challenger can choose

<sup>66</sup> Morgan, *Deterrence Now*.

<sup>67</sup> See William Reed, 'A Unified Statistical Model of Conflict Onset and Escalation', *American Journal of Political Science*, 44 (2000), pp. 84–93, and especially James D. Fearon, 'Selection Effects and Deterrence', *International Interactions*, 28 (2002), pp. 5–30.

<sup>68</sup> Morgan, *Deterrence*, pp. 42 and 45.

<sup>69</sup> Morgan, *Deterrence Now*, p. 86.

<sup>70</sup> Zagare and Kilgour, *Perfect Deterrence*, ch. 5.

whether to cooperate or defect. If Challenger cooperates, the *Status Quo* persists. If Challenger chooses to defect, Defender has an opportunity to respond. At node 2, Defender can choose whether to concede, resulting in *Defender Concedes*, or defy, giving Challenger the next choice. At node 3, Challenger can also choose whether to concede or defy. If Challenger concedes, the outcome is *Challenger Defeated*. However, if Challenger chooses to defy, the outcome is *Conflict*.

General deterrence fails as a result of some challenge to the *status quo* (that is, Challenger chooses to defect at node 1). So if no such challenge occurs (Challenger cooperates at node 1), then general deterrence succeeds and the *status quo* remains unchanged. Immediate deterrence deals with attempts by defender to get the challenger to back down (Defender's choice of how to respond at node 2 and Challenger's choice of whether to back down at node 3). Clearly the Unilateral Deterrence Game is a model of general deterrence.

In addition to the model of unilateral deterrence reviewed here, perfect deterrence theory incorporates models of mutual and extended deterrence, which are also focused within the realm of general deterrence. Thus, perfect deterrence theory provides a theory of general deterrence. Unfortunately, however, the 'empirical study of general deterrence is less extensive and less well developed than is the body of work on immediate deterrence'.<sup>71</sup> I now turn to an examination of the difficult task of testing deterrence theory.

### Testing deterrence theory

Theories are useful to the extent that they help to understand and explain reality. To determine the usefulness of deterrence theory, empirical testing is needed.<sup>72</sup> Two empirical approaches have been used in the deterrence literature: case-studies and quantitative analysis.<sup>73</sup> The case-study literature has focused on in-depth analyses of particular deterrence situations, but these studies have generally endeavoured to criticise, rather than test, rational deterrence theory.<sup>74</sup>

Furthermore, an unfortunate divide exists between formal theories and the quantitative analysis of deterrence. There are several reasons for this disconnect: 1) while deterrence theory has typically focused on general deterrence, quantitative studies have focused almost exclusively on immediate deterrence; 2) quantitative studies that have been done have generally not tested rational deterrence theory, but rather have tested independently developed hypotheses and, 3) when studies

<sup>71</sup> Paul K. Huth, 'Deterrence and International Conflict: Empirical Findings and Theoretical Debates', *Annual Review of Political Science*, 2 (1999), p. 63.

<sup>72</sup> Frank P. Harvey, 'Rigor Mortis or Rigor, More Tests: Necessity, Sufficiency, and Deterrence Logic', *International Studies Quarterly*, 42 (1998), pp. 675–707.

<sup>73</sup> Jack Levy, 'When Do Deterrent Threats Work?', *British Journal of Political Science*, 18 (1988), pp. 485–512, and Huth, 'Deterrence and International Conflict', provide reviews of the empirical literature on deterrence.

<sup>74</sup> For example, George and Smoke, *Deterrence in American Foreign Policy*; John Steinbrunner, 'Beyond Rational Deterrence', *World Politics*, 28 (1976), pp. 223–45; Lebow, *Between Peace and War*; Richard Ned Lebow, 'Windows of Opportunity: Do States Jump Through Them?', *International Security*, 9 (1984), pp. 147–86; Robert Jervis, 'Deterrence and Perception', *International Security*, 7 (Winter 1982/83), pp. 14–7; Jervis, Lebow, and Stein, *Psychology and Deterrence*; Lebow and Stein, 'Deterrence: The Elusive Dependent Variable'.

have harkened back to rational deterrence theory, they have typically examined classical deterrence theory, a formal framework characterised by logical inconsistency and empirical inaccuracy.

An example of this continuing divide is Danilovic's recent quantitative study of extended deterrence.<sup>75</sup> Danilovic argues that a state's inherent credibility – determined in large part by its regional interests – is a far more important predictor of deterrence outcomes than attempts to shore up credibility through the use of commitment strategies recommended by Schelling and other classical deterrence theorists.<sup>76</sup> However, she does not explicitly tie her quantitative analysis to any theory of deterrence. Thus, it is left to the reader to bridge this divide.

Fortunately, through her focus on credibility, Danilovic directly addresses a key distinction between perfect and classical deterrence theory.<sup>77</sup> As discussed above, the reason that classical deterrence theorists focus on commitment tactics as ways to shore up credibility is that they assume that all threats are inherently incredible. Zagare and Kilgour, on the other hand, argue that states' threats are only credible to the extent that the state actually prefers conflict to backing down.<sup>78</sup> Accordingly, Danilovic's analysis can be seen as an attempt to determine what factors (relative power, regional interests, and democracy) lead states to prefer conflict over backing down.

Since rational deterrence theory is primarily expressed through game-theoretic models, empirical tests of the theory need to be able to test predictions that result from such models. There are two primary avenues available for such tests: evaluation of equilibrium predictions and evaluation of relationship predictions.<sup>79</sup> Evaluation of equilibrium predictions entails comparison of the outcome predicted in equilibrium at each observation with the outcome actually observed. Bennett and Stam's test of Bueno de Mesquita and Lalman's International Interaction Game is an example of such evaluations of equilibrium predictions.<sup>80</sup> Once one relates Danilovic's argument about inherent credibility to Zagare and Kilgour's arguments about credibility, it becomes clear that her analysis is at least indirectly an evaluation of perfect deterrence theory's relationship predictions.<sup>81</sup>

Selection of deterrence cases has been a subject of much disagreement, primarily between quantitative and case study researchers. An important part of the debate in *World Politics* in 1989–1990 dealt with the proper identification of immediate deterrence cases.<sup>82</sup> Unfortunately, however, the existing deterrence literature provides little guidance on the selection of general deterrence cases. This is particularly important because of the importance of general deterrence, as discussed above.

<sup>75</sup> Danilovic, *When the Stakes Are High*.

<sup>76</sup> Schelling, *Strategy of Conflict*; Schelling, *Arms and Influence*.

<sup>77</sup> Danilovic, *When the Stakes Are High*.

<sup>78</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>79</sup> Rebecca B. Morton, *Methods and Models* (Cambridge: Cambridge University Press, 1999).

<sup>80</sup> D. Scott Bennett and Allan C. Stam, 'A Universal Test of an Expected Utility Theory of War', *International Studies Quarterly*, 44 (2000), pp. 451–80; Bruce Bueno de Mesquita and David Lalman, *War and Reason: Domestic and International Imperatives* (New Haven: Yale University Press, 1992).

<sup>81</sup> Danilovic, *When the Stakes Are High*; Zagare and Kilgour, *Perfect Deterrence*.

<sup>82</sup> Christopher H. Achen and Duncan Snidal, 'Rational Deterrence Theory and Comparative Case Studies', *World Politics*, 41 (1989), pp. 143–69; Robert Jervis, 'Rational Deterrence: Theory and Evidence', *World Politics*, 41 (1989), pp. 183–207; Lebow and Stein, 'Rational Deterrence Theory'; Lebow and Stein, 'Deterrence: The Elusive Dependent Variable'; Paul K. Huth and Bruce Russett, 'Testing Deterrence Theory: Rigor Makes a Difference', *World Politics*, 42 (1990), pp. 466–501.

Morgan sees the maintenance of armed forces as indicative of general deterrence behaviour.<sup>83</sup> One can safely assume that every state wishes to deter attacks against itself – this is the basic rationale for the maintenance of armed forces. This is essentially identical to the assumption within the alliance portfolio literature that every state has a defence pact with itself.<sup>84</sup> Therefore, the difficult part of general deterrence case selection is not determining who makes deterrent threats (everyone does), but rather what states the threats are directed against. General deterrent threats are directed against any state that might consider an attack, but it can be difficult to identify exactly which states might do so. The key to general deterrence case selection then is to identify which states may consider attacking which other states. I argue that the states that might consider an attack on a particular state are those that have the opportunity to fight that state. Fortunately, Quackenbush has developed the concept of politically active dyads, which capture opportunity as a necessary condition for international conflict.<sup>85</sup> Therefore, politically active dyads can be used as a case selection mechanism for studies of general deterrence.

Another concern that has often driven debates between those employing case-study<sup>86</sup> and quantitative<sup>87</sup> methodologies is the determination of deterrence success and failure. However, game-theoretic models of deterrence do not make predictions regarding the success or failure of deterrence, *per se*; rather, the particular outcome of an interaction is predicted.<sup>88</sup> Furthermore, classification of certain cases as successful deterrence runs into a variety of conceptual and selection bias issues.<sup>89</sup>

Quackenbush tests the equilibrium outcome predictions of perfect deterrence theory.<sup>90</sup> Since the equilibrium predictions of perfect deterrence theory rely on the actors' preferences, this requires explicit specification of utility functions for the actors involved. To do this, he employs a modification of the technique – using relative power as a measure for probability of victory and measuring utilities based on similarity of alliance portfolios – pioneered by Bueno de Mesquita and Lalman.<sup>91</sup> Quackenbush uses binary and multinomial logit methods to examine the

<sup>83</sup> Morgan, *Deterrence*; also see Paul K. Huth, *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988).

<sup>84</sup> Bruce Bueno de Mesquita, 'Measuring Systemic Polarity', *Journal of Conflict Resolution*, 19 (1975), pp. 187–215; Bruce Bueno de Mesquita, 'Systemic Polarization and the Occurrence and Duration of War', *Journal of Conflict Resolution*, 22 (1978), pp. 241–67; Curtis S. Signorino and Jeffrey M. Ritter, 'Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions', *International Studies Quarterly*, 43 (March 1999), pp. 115–44. In other words, every state threatens to respond if attacked, even though it may not actually carry out the threat if attacked (just as defence pacts are not always honoured when tested).

<sup>85</sup> Stephen L. Quackenbush, 'Identifying Opportunity for Conflict: Politically Active Dyads', *Conflict Management and Peace Science*, 23(2006), pp. 37–51.

<sup>86</sup> Lebow and Stein, 'Rational Deterrence Theory'; Lebow and Stein, 'Deterrence: The Elusive Dependent Variable'.

<sup>87</sup> Huth and Russett, 'Testing Deterrence Theory'.

<sup>88</sup> For example, in the Unilateral Deterrence Game (Figure 3), the only rational outcome is *Status Quo* if Challenger prefers *Status Quo* to *Defender Concedes*. Although Challenger's decision to not challenge the *status quo* in this case is not really 'successful deterrence,' it is predicted by perfect deterrence theory.

<sup>89</sup> Danilovic, 'Conceptual and Selection Bias Issues in Deterrence'.

<sup>90</sup> Stephen L. Quackenbush, 'General Deterrence and International Conflict: Testing Perfect Deterrence Theory', *International Interactions*, 36 (2010), pp. 60–85.

<sup>91</sup> Bueno de Mesquita and Lalman, *War and Reason*.



prediction of militarised interstate disputes and of particular game outcomes, and the results indicate that the predictions of perfect deterrence theory are generally supported by the empirical record.

Perfect deterrence theory has also been applied to the analysis of conflict settlements by Senese and Quackenbush.<sup>92</sup> Their analysis demonstrates that perfect deterrence theory – through a comparison of the unilateral and generalised mutual deterrence games – makes predictions about the stability of different types of conflict settlements, and these predictions are strongly supported by the empirical record. This is an insight that was not gained through other analyses of the same subject.

Most attempts to test deterrence theory have utilised large-N, quantitative methods. Such analysis is useful because of the generalisable nature of conclusions derived from it. However, case studies would also be useful for future research. Since quantitative analyses demonstrate that perfect deterrence theory is generally supported by the historical record,<sup>93</sup> there is evidence that the theory can usefully be applied to particular cases. This can be done through the application of the insights of perfect deterrence theory to detailed case studies of particular historical episodes in order to understand these intrinsically interesting events more completely. For example, one study applies perfect deterrence theory to an analysis of the war in Kosovo to answer questions – such as why NATO's threat of bombing was unable to deter Serbia, and later why Serbia escalated ethnic cleansing once the bombing started – to which others have struggled for answers.<sup>94</sup> Another study explains the July 1914 crisis leading to World War I through an analytic narrative based on perfect deterrence theory.<sup>95</sup> The analysis demonstrates that although general war was not sought by any of the actors, the war was no accident.

### Theoretical extensions

Although Zagare and Kilgour resolve the paradox of mutual deterrence and develop a logically consistent theory of deterrence, perfect deterrence theory, further questions still remain.<sup>96</sup> Recent efforts to extend our understanding of deterrence have focused on three primary areas: three-actor games, bargaining, and identifying credibility.

#### *Three-actor games*

Although direct deterrence deals with two states, extended deterrence deals with a minimum of three. Werner outlines the basic logic of extended deterrence:

<sup>92</sup> Senese and Quackenbush, 'Sowing the Seeds of Conflict'.

<sup>93</sup> Danilovic, *When the Stakes Are High*; Quackenbush, 'General Deterrence and International Conflict'; Senese and Quackenbush, 'Sowing the Seeds of Conflict'.

<sup>94</sup> Stephen L. Quackenbush and Frank C. Zagare, 'Game Theory: Modeling Interstate Conflict', in Jennifer Sterling-Folker (ed.), *Making Sense of International Relations Theory* (Boulder, CO.: Lynne Rienner, 2006), pp. 98–114.

<sup>95</sup> Zagare, 'Explaining the 1914 War in Europe'.

<sup>96</sup> Zagare and Kilgour, *Perfect Deterrence*.

The attacker must decide whether or not to attack the target, which then must decide whether or not to resist. After observing both the attacker's and the target's actions, the third party must decide whether or not to come to the aid of the target if the target has in fact been attacked.<sup>97</sup>

However, standard practice has been to include only two players in models of extended deterrence: the attacker (or challenger) and the defender. Indeed, virtually all analyses of extended deterrence have focused on only two players.<sup>98</sup> Thus, although Werner clearly highlights the presence and the choices of the target (as do others), the target's role in the strategic calculation involved are generally ignored. Indeed, one article that strives to determine the 'role of the pawn in extended deterrence' does not model the pawn (that is, target or *protégé*) as one of the players.<sup>99</sup>

Therefore, an important development in deterrence theory has been to explicitly model deterrence situations wherein the choices of the target are strategically salient. Zagare and Kilgour do so through a Tripartite Crisis Game, which focuses on *Protégé's* potential to realign with a more reliable partner should Defender not support her following a challenge.<sup>100</sup> This highlights the importance of the 'deterrence versus restraint' dilemma in extended deterrence wherein Defender has competing interests in having a strong enough commitment to support *Protégé* so as to deter Challenger, but not making such a strong commitment that *Protégé* would be encouraged to act recklessly.

Zagare and Kilgour extend this analysis to explain British policy in the July crisis leading to the outbreak of World War I.<sup>101</sup> Zagare and Kilgour demonstrate why half-hearted extended deterrent signals are so often observed in international politics in general, and in Britain's attempts to deter Germany in 1914 in particular. In so doing, they demonstrate that their analysis more fully captures the dynamics of the deterrence versus restraint dilemma than Crawford's model of 'pivotal deterrence', which unfortunately is not based on the axiomatic base of perfect deterrence theory.<sup>102</sup>

Quackenbush also extends deterrence theory to consider the strategic interactions of all three actors in extended deterrence.<sup>103</sup> He develops a Three-party Extended Deterrence Game, a formal model that matches the basic informal logic of extended deterrence. This model is the first to explicitly model the simultaneous conduct of extended and direct deterrence identified by Snyder.<sup>104</sup> Furthermore,

<sup>97</sup> Suzanne Werner, 'Deterring Intervention: The Stakes of War and Third-Party Involvement', *American Journal of Political Science*, 44 (October 2000), p. 721.

<sup>98</sup> For example, Danilovic, *When the Stakes Are High*; Huth, *Extended Deterrence*; D. Marc Kilgour and Frank C. Zagare, 'Uncertainty and the Role of the Pawn in Extended Deterrence', *Synthese*, 100 (1994), pp. 379–412; Bruce M. Russett, 'The Calculus of Deterrence', *Journal of Conflict Resolution*, 7 (1963), pp. 97–109; Signorino and Tarar, 'Unified Theory and Test'; Werner, 'Deterring Intervention'; Frank C. Zagare and D. Marc Kilgour, 'Modeling "Massive Retaliation"', *Conflict Management and Peace Science*, 13 (1993), pp. 61–86.

<sup>99</sup> Kilgour and Zagare, 'Uncertainty and the Role of the Pawn'.

<sup>100</sup> Frank C. Zagare and D. Marc Kilgour, 'Alignment Patterns, Crisis Bargaining, and Extended Deterrence: A Game-Theoretic Analysis', *International Studies Quarterly*, 47 (2003), pp. 587–615.

<sup>101</sup> Frank C. Zagare and D. Marc Kilgour, 'The Deterrence-Versus-Restraint Dilemma in Extended Deterrence: Explaining British Policy in 1914', *International Studies Review*, 8 (2006), pp. 623–41.

<sup>102</sup> Timothy W. Crawford, *Pivotal Deterrence: Third-Party Statecraft and the Pursuit of Peace* (Ithaca: Cornell University Press, 2003).

<sup>103</sup> Quackenbush, 'Not only Whether but Whom'.

<sup>104</sup> Snyder, *Deterrence and Defense*.

the inclusion of all three players in the analysis allows examination of other issues such as the target of a challenge and alliance reliability.

Quackenbush's analysis indicates that potential challengers look to attack members of unreliable alliances.<sup>105</sup> Furthermore, deterrence is more likely to succeed the more highly Defender and *Protégé* value a war (bilateral or multilateral) with Challenger, and the less highly they value a concession to Challenger. In addition, Challenger's increasing utility for the *status quo* also increases the probability of successful deterrence. And finally, when deterrence fails, the likely target of an attack by Challenger is the more reliable of the allies – in order to keep the unreliable side on the sidelines and prevent a multilateral war.

By including all three actors within game-theoretic models, these studies have significantly improved our understanding of the dynamics of extended deterrence. Certainly, further developments in three-actor models are warranted, and are being developed.<sup>106</sup> As Quackenbush demonstrates, standard two-party extended deterrence games rely upon the oft-hidden assumption that states always respond to direct challenges.<sup>107</sup> But since this assumption directly contradicts theories of direct deterrence, analysing all three actors in extended deterrence appears to be a more consistent way of approaching the problem.

### *Bargaining and deterrence*

Traditional analyses of International Relations have modelled decisions as stark choices between two, or sometimes three, alternatives. This is illustrated in the Unilateral Deterrence Game in Figure 3, which presents choices between 'Cooperate' and 'Defect' or 'Concede' and 'Defy.' However, states often have a wide variety of options available to them; for example, rather than just choosing between 'attack' or 'not attack', a state could do nothing, partially mobilise troops, fully mobilise, embark on a show of force, launch airstrikes, or even initiate a full-on invasion. This of course provides a much more nuanced menu of options than traditional models would suggest.

The bargaining model of war has been a recent focus of research on International Relations.<sup>108</sup> Although models vary, they incorporate these nuanced options within the game-theoretic model, making the stakes under consideration endogenous to the model itself. Thus, states choose not only *whether* to make a demand, attack, etc., but also *how much* to demand, attack, etc.

<sup>105</sup> Quackenbush, 'Not only Whether but Whom'.

<sup>106</sup> For example, Amy Yuen, 'The Strategic Dynamics of Third-Party Intervention in Interstate Conflict' (Ph.D. diss., Emory University, 2007).

<sup>107</sup> Quackenbush, 'Not only Whether but Whom'.

<sup>108</sup> For example, R. Harrison Wagner, 'Bargaining and War', *American Journal of Political Science*, 44 (2000), pp. 469–84; Darren Filson and Suzanne Werner, 'A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War', *American Journal of Political Science*, 46 (2002), pp. 819–38; Robert Powell, 'Bargaining Theory and International Conflict', *Annual Review of Political Science*, 5 (2002), pp. 1–30; Dan Reiter, 'Exploring the Bargaining Model of War', *Perspectives on Politics*, 1 (2003), pp. 27–43.

Werner developed the first model of deterrence to incorporate these endogenous stakes.<sup>109</sup> In her Modified Extended Deterrence Game, the attacker can choose the magnitude of the threat. By modelling the stakes of war as endogenous in this way, Werner is able to identify circumstances under which extended deterrence fails because the attacker deters the third party from intervening.

Langlois and Langlois have also examined the relationship between bargaining and deterrence in analyses of direct, asymmetric deterrence.<sup>110</sup> This allows the rivals to bargain over the terms of an alternative *status quo*. It also shows how war can erupt between two rational states, even when there is complete information, contrary to traditional arguments.<sup>111</sup>

Further analyses integrating such bargaining models within studies of deterrence are likely to continue to be a fruitful avenue for research. Nonetheless, it seems important to recognise the importance of domestic politics, and its influence on the demands that states make. For example, Bueno de Mesquita and Lalman focus on two variants of their International Interaction Game: in the Realpolitik variant, the magnitude of the demand is purely endogenous to the game, whereas in the Domestic variant, the magnitude of the demand is assumed to be determined by domestic political processes, and is thus exogenous to the game.<sup>112</sup> Furthermore, they find that the Domestic variant does a much better job of explaining international interactions than the Realpolitik variant.

### *Identifying credibility*

A third important focus of recent theoretical developments has been on identifying credibility. Perfect deterrence theory demonstrates that credibility is an important determinant of deterrence success. In the presence of complete information, a state's credibility is simply its preference between fighting and backing down; it either has a credible threat (that is, prefers conflict) or does not (that is, prefers backing down). However, states seldom possess such full information about their opponents.

When there is incomplete information and a state's preference between fighting and backing down is not known to its opponent, the situation is more complicated. In this case, a state's credibility is the opponent's estimate of the probability that it prefers fighting to backing down; thus, credibility varies continuously between 0 and 1.<sup>113</sup> Forming accurate estimates of the opponent's credibility is complicated by the reality that states have incentives to misrepresent their private information.<sup>114</sup> Thus, states have incentives to claim that they prefer to fight rather than back down

<sup>109</sup> Werner, 'Deterring Intervention'.

<sup>110</sup> Catherine C. Langlois and Jean-Pierre P. Langlois, 'Bargaining and the Failure of Asymmetric Deterrence: Trading off the Risk of War for the Promise of a Better Deal', *Conflict Management and Peace Science*, 23 (2006), pp. 159–80; Catherine C. Langlois and Jean-Pierre P. Langlois, 'When Fully Informed States Make Good the Threat of War: Rational Escalation and the Failure of Bargaining', *British Journal of Political Science*, 36 (2006), pp. 645–69.

<sup>111</sup> James D. Fearon, 'Rationalist Explanations for War', *International Organization*, 49 (1995), pp. 379–414.

<sup>112</sup> Bueno de Mesquita and Lalman, *War and Reason*.

<sup>113</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>114</sup> Fearon, 'Rationalist Explanations for War'.

regardless of their actual preference between the two. Thus, how can states credibly communicate their commitment to fight in the face of these difficulties?

Two recent books have sought to address this question. Sartori develops a reputational theory of diplomacy whereby defenders acquire reputations for honesty or bluffing depending on whether or not they carry through on their threats to defend against the challenger.<sup>115</sup> She finds that a defender's reputation for honesty enhances their threat credibility and therefore increases the likelihood of successful deterrence; on the other hand, a reputation for bluffing undermines threat credibility.

In contrast, Press proposes a 'current calculus' model, which evaluates credibility based on the balance of power and interests at stake, and argues that it does a better job of gauging state's perceptions of their opponents' credibility than the 'past actions' model, which argues that credibility depends on one's record for keeping or breaking commitments.<sup>116</sup> His focus on assessing the interests at stake to determine credibility is quite reasonable given that credibility deals with a state's preference between backing down and conflict.

However, it is important to distinguish between credibility and capability; whereas capability is a necessary condition for deterrence success, credibility is not.<sup>117</sup> But assessments of the balance of power appear to have more to do with a state's preference between conflict and the *status quo* than a state's preference between conflict and backing down. Therefore, it appears that Press is conflating credibility and capability. For example, he argues that Britain's threat to defend Poland in 1939 was not credible, in large part because Germany (especially Hitler) misjudged the balance of power. It seems clear that, although they could not be certain, Germany believed that British intervention was likely, and thus the British threat was credible. But as Weinberg states, 'in August 1939 Hitler preferred to attack Poland as a preliminary to attacking France and England, but he was quite willing to face war with the Western Powers earlier if that was their choice'.<sup>118</sup> In other words, Germany preferred conflict with France and Britain to the *status quo*; despite being credible, the British threat was not capable.

## Conclusion

This article sought to evaluate the current standing of deterrence theory. This was done through an examination of several important topics in deterrence theory. Contrary to Morgan,<sup>119</sup> there are two main theories of deterrence: classical deterrence theory and perfect deterrence theory. While both are rational-choice theories, they differ in several respects, particularly regarding their treatment of credibility. Classical deterrence theory assumes that conflict is the worst possible

<sup>115</sup> Sartori, *Deterrence by Diplomacy*.

<sup>116</sup> Daryl G. Press, 'The Credibility of Power: Assessing Threats during the "Appeasement" Crises of the 1930s', *International Security*, 29 (2004), pp. 136–69; Daryl G. Press, *Calculating Credibility: How Leaders Assess Military Threats* (Ithaca: Cornell University Press, 2005).

<sup>117</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>118</sup> Gerhard L. Weinberg, *Germany, Hitler, and World War II* (Cambridge: Cambridge University Press, 1995), p. 91.

<sup>119</sup> Morgan, *Deterrence Now*.

outcome for both sides, meaning that no retaliatory threats are credible. However, this assumption leads directly to the paradox of mutual deterrence, wherein classical theory is unable to explain deterrence success.

The paradox of mutual deterrence is solved by perfect deterrence theory, which argues that the credibility of a state's threat depends upon its preference between backing down and conflict. Therefore, scholars need to move away from the assumption that conflict is the worst possible outcome. By doing so, Zagare and Kilgour are able to develop a logically consistent theory of general deterrence that usefully explains the dynamics of deterrence in mutual, unilateral, and extended deterrence situations.<sup>120</sup>

Although the distinction between classical and perfect deterrence theories is stark, this distinction – or at least the challenge of perfect deterrence theory – is often ignored. For example, although Morgan argues that there is only one deterrence theory, in his list of 'standard works on deterrence theory' any reference to perfect deterrence theory is conspicuously absent.<sup>121</sup> Similarly, although Sartori allows the defender's preference between war and backing down to vary in her game-theoretic model, she makes no reference to the work of Zagare or perfect deterrence theory – which is where the idea of variable threat credibility comes from – but instead relates her work to classical deterrence theorists such as Schelling and Powell.<sup>122</sup> More examples could be provided, but the theme is consistent.

The contrast between the two theories is not simply a matter of differences over assumptions and abstract theoretical issues. The policy recommendations of perfect deterrence theory and classical deterrence theory are diametrically opposed on many issues.<sup>123</sup> For example, classical deterrence theory argues that national missile defence undermines the stability of deterrence,<sup>124</sup> whereas an application of perfect deterrence theory demonstrates that national missile defence can enhance deterrence stability.<sup>125</sup> However, policy discussions in academia and government are generally based on classical deterrence theory. Given its strong empirical support, coupled with logical and empirical limitations of classical deterrence theory, perfect deterrence theory provides a much better basis for analysing various aspects of national security policy.

Clearly, classical deterrence theory constitutes the 'conventional wisdom' regarding deterrence. Nonetheless, classical deterrence theory is badly flawed. The (unnecessarily restrictive) assumption that conflict is always the worst possible outcome needs to be discarded. It has not proven useful for developing logically consistent and empirically accurate theory. By contrast, perfect deterrence theory provides a logically consistent alternative to understand the dynamics of deterrence. Therefore, perfect deterrence theory provides the most appropriate basis for further theoretical development, empirical testing, and application to policy.

<sup>120</sup> Zagare and Kilgour, *Perfect Deterrence*.

<sup>121</sup> Morgan, *Deterrence Now*, p. 8.

<sup>122</sup> Sartori, *Deterrence by Diplomacy*; in particular, see Frank C. Zagare, 'Toward a Reformulation of the Theory of Mutual Deterrence', *International Studies Quarterly*, 29 (1985), pp. 155–69 and Zagare, *Dynamics of Deterrence*, which introduce the ideas that eventually develop into perfect deterrence theory.

<sup>123</sup> Zagare, 'Reconciling Rationality with Deterrence', especially Table 1.

<sup>124</sup> For example, Powell, 'Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense'.

<sup>125</sup> Quackenbush, 'National Missile Defense and Deterrence'.