

# The Deterrence-Versus-Restraint Dilemma in Extended Deterrence: Explaining British Policy in 1914<sup>1</sup>

FRANK C. ZAGARE

*Department of Political Science, University at Buffalo*

AND

D. MARC KILGOUR

*Department of Mathematics, Wilfrid Laurier University, Canada*

This essay explores the deterrence-versus-restraint dilemma in extended deterrence in the context of the Tripartite Crisis Game under incomplete information. This model was developed specifically to capture the mixed motives and contradictory impulses that oftentimes frame extended deterrence encounters. To focus the analysis and to gain tractability, we make specific assumptions about the utilities of the players: Challenger, Defender, and Protégé. Our most significant simplification concerns Defender's type. In particular, we assume that Defender, although not heavily invested in the issues in dispute, is known to prefer conflict to the breakup of its strategic relationship with Protégé. One important result concerns unequivocal commitments. We find that such commitments are efficacious but only when Protégé's threat to sever its relationship with Defender is highly credible. In the absence of this condition, a straddle (or mixed) strategy is optimal for Defender. A straddle strategy, which involves probabilistic support of Protégé, is the mechanism by which Defender attempts to resolve the deterrence-versus-restraint dilemma. Sometimes, the stratagem works and Challenger is deterred and Protégé is restrained. But a straddle strategy will always fail to deter determined Challengers, such as Germany in 1914, that prefer to fight rather than back down during a confrontation. It may even fail to deter hesitant Challengers with an aversion to conflict. We use these insights to explain and evaluate British policy in the runup to World War I.

---

Although we cannot bind ourselves under all circumstances to go to war with France against Germany, we shall also certainly not bind ourselves to Germany not to assist France.

Sir Edward Grey, April 1912

For mainstream deterrence theorists, the robust stability of direct deterrence relationships (at least under parity) is virtually certain. Given a naturally occurring

---

<sup>1</sup>We would like to express our appreciation for the helpful comments and suggestions of Vesna Danilovic, Stephen Quackenbush, and Curtis Signorino.

strategic balance and artificially adjusted levels of armaments, peace is all but assured (Waltz 1964, 1990). Extended deterrence relationships, by contrast, are considered, by theorists and policymakers alike, to be the problematic case (Quester 1989:63). This is not a surprise. After all, almost all major power wars have followed an extended deterrence breakdown (Danilovic 2002:53).

The critical difference between the two deterrence milieu is best explained by the characteristics of the threats that define them. In most direct deterrent situations, threats to resist aggression are inherently credible (Schelling 1966:36). Consequently, when one state directly threatens another's homeland, all that is required to steady a shaky status quo is a threat that is sufficiently capable. Many classical deterrence theorists support the selective proliferation of nuclear weapons for precisely this reason. As they dramatically increase the cost of conflict, nuclear and other weapons of mass destruction render inherently credible threats capable as well.

In extended deterrence, however, threat credibility is oftentimes the missing ingredient. Even when the capability of threats is evident, a Defender of a third party (that is, of a Protégé) may have difficulty establishing its credibility. The US threat to defend Western Europe during the Cold War is perhaps the most pertinent example. Despite strong social, economic, and political ties to the European democracies, some questioned the willingness of the United States to trade New York or Washington for Paris or Bonn.

It is obvious that deterrence, extended or otherwise, is likely to fail when a Defender's threat is insufficiently credible. When a defender fails to make a credible extended deterrent threat, it is said to have "passed the buck." The agreement reached at the Munich Conference of 1938 is the classic case. By failing to stand up to Hitler, it has been argued, both Britain and France all but guaranteed the breakdown of deterrence that is now associated with World War II.

Some theorists, however, claim that extended deterrence can also fail when states project too much credibility. Spiral theorists, for example, argue that an irrevocable commitment to an ally, which is sometimes called "chain-ganging," risks setting in motion an action–reaction process that frequently culminates in war (Christensen and Snyder 1990). Germany's "blank check" to Austria in 1914 is but one of several well-known instances of a relatively minor and potentially avertable conflict that escalated after a Defender made an ironclad commitment to an ally. Given Germany's unconditional support, Austria's motivation to reach a political resolution of its long-simmering conflict with Serbia was dampened. Similarly, Russia's steadfast support of Serbia, as well as France's of Russia, is associated with the escalation spiral that produced World War I.

Clearly, there are situations wherein extended deterrence threats require careful calibration, circumstances under which threats that are either too credible or that are not credible enough are ineffective. For instance, consider the plight of a Defender attempting to forestall a crisis between two other states, one of which is an ally or a Protégé, and the other a dissatisfied Challenger. If the Defender makes too strong a commitment to its ally, it creates a moral hazard: feeling secure, the Protégé may behave irresponsibly, thereby provoking the very confrontation the Defender had hoped to avoid. But if the Defender's commitment is too weak, the Challenger may be emboldened to address its grievance too aggressively and preclude a peaceful resolution of the dispute.

The Defender in this hypothetical example obviously faces a tricky strategic conundrum that arises often in international politics. Glenn Snyder and Paul Diesing (1977:432) call it the "deterrence-versus-restraint" dilemma.<sup>2</sup> Facing this

<sup>2</sup>In later works, Snyder (1984, 1997) refers to it as the "composite security dilemma."

dilemma in 1879, for example, Germany entered into a defensive alliance with Austria–Hungary in part to dissuade a Russian challenge to Austria–Hungary, but also in part to encourage Austria–Hungary to settle its dispute with Russia over the Balkans.<sup>3</sup> The policy was a success. By 1881, Austria–Hungary and Russia were linked (along with Germany) in the Three Emperors’ League.

The British were in a similar position in 1914. But their attempt to deter Germany and restrain the French and (indirectly) the Russians failed to avert a general European war. Currently, the US policy objective in Asia, as codified in the “deliberately vague” 1979 Taiwan Relations Act, is to preserve the status quo by deterring China by promising to defend Taiwan and, simultaneously, to restrain the Taipei government by leaving open the precise set of circumstances under which the treaty is operative (Erlanger 1996). Whether the policy will succeed in the long term is unclear.

### Modeling the Deterrence-Versus-Restraint Dilemma

To analyze the deterrence-versus-restraint dilemma, Snyder (1984, 1997:37) offers two related but theoretically isolated game models: the “alliance game” and the “adversary game.” In the alliance game, a Defender either supports (C) or withholds support (D) from its Protégé. In the adversary game, a Defender either stands firm (D) or conciliates (C) with its Challenger. In general, the choice of (C) in one game implies a choice of (D) in the other, and vice versa. The linkage between the choices in the two games makes it difficult for a Defender to choose optimally in both, that is, to deter the Challenger and restrain the ally simultaneously. Clearly, Snyder views the deterrence-versus-restraint dilemma as a difficult cross-game maximization problem. Snyder’s analysis of the deterrence-versus-restraint dilemma is informal and, as Timothy Crawford (2003:18) observes, his separation of the alliance and adversary games is unduly artificial. As a consequence, Snyder’s conceptual synthesis, while provocative and insightful, remains intuitive and needlessly imprecise. To overcome these and related limitations, we offer a single game form designed specifically to capture some of the tensions implicit in Snyder’s two discrete games. We call this game the *Tripartite Crisis Game* (see Figure 1). The particular assumptions we make about the preferences of the players (see later) are motivated by a desire to bring the deterrence-versus-restraint dilemma into sharp relief. In the end, we hope to gain additional theoretical precision about the interplay of competing constraints on the dynamics of extended deterrence relationships.

As its name suggests, there are three players in the Tripartite Crisis Game: Challenger, Protégé, and Defender. Play begins with Challenger’s choice at node 1. Challenger has two options. If it makes no demand of Protégé, the game ends, extended deterrence succeeds, and the outcome is the *Status Quo* (or *SQ*). But if Challenger confronts Protégé by demanding an adjustment to the *Status Quo*, a crisis occurs. In this contingency, Protégé chooses next at node 2. The choice is critical. If, at node 2, Protégé concedes to Challenger’s demands, the crisis is resolved and the outcome is *Challenger Wins* (*CW*). But if Protégé holds firm, resisting Challenger’s demands, the crisis intensifies and Defender’s dilemma is realized: at node 3 Defender must decide whether to support Protégé. Of course, if Defender supports Protégé and, subsequently (at node 4b), Challenger backs down, the crisis is resolved to Defender’s (and Protégé’s) satisfaction, that is, the outcome is *Challenger Concedes* (*CC*). But if Challenger presses on, *Conflict* (*C*) occurs. Clearly, this is an outcome that Defender prefers to avoid. But it is not the only one.

<sup>3</sup>And, as Langer (1950:180) notes, partly to preclude an Austrian alliance with France.

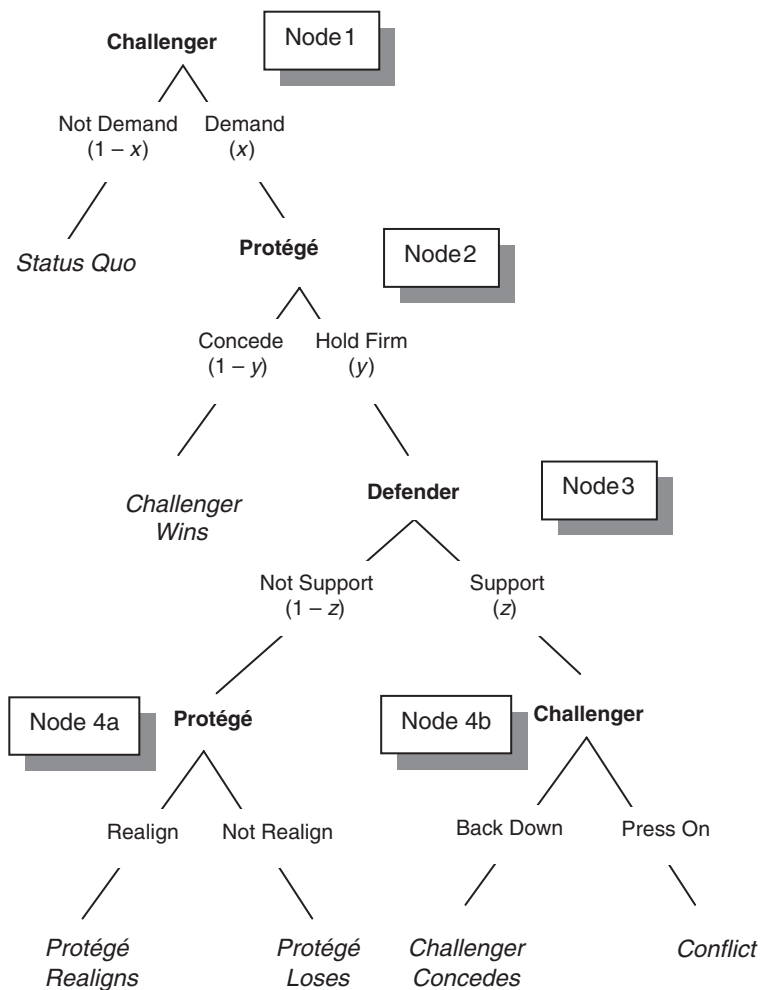


FIG. 1. Tripartite Crisis Game

More specifically, if Defender decides not to support Protégé, it is sure to suffer a costly diplomatic defeat when it becomes apparent that it is an unreliable ally. Just how costly, however, depends on Protégé's (node 4a) reaction.<sup>4</sup> Protégé could simply accept the implications of its abandonment by Defender and accede, finally, to Challenger's demands, as did Austria in 1913 when it was forsaken by Germany during the Balkan crisis. The outcome in this case is *Protégé Loses (PL)*. Or, in addition to accommodating Challenger, Protégé could also sever its relationship with Defender. The outcome associated with this choice, which we call *Protégé Realigns (or PR)*, clearly imposes an additional cost on Defender. Intended or not, it is likely that Defender will view this option as a threat. For example, during the 1908–1909 Bosnian crisis, both German Chancellor Bernard Bülow and Friedrich von Holstein, a key foreign policy advisor at the Wilhelmstrasse, feared that Austria would tilt toward England and France if it were not supported in its dispute with Russia. Under the circumstances, it should be no surprise that “Bülow and Holstein

<sup>4</sup>Note that Protégé's node 4a choices do not include an option to resist Challenger. By not including this option, we implicitly assume that an isolated Protégé is no match for Challenger. Should it be abandoned by Defender, Protégé would be forced to comply to a stronger Challenger's demands.

felt they had no choice but to support Austria in its bid to annex Bosnia and Herzegovina" (Mercer 1996:125; see also Kagan 1995:162–163).

It is important to point out that Protégé is not the only player with a threat in the Tripartite Crisis Game. Defender and Challenger have potent threats as well. Of course, Defender's support of Protégé at node 3 poses a threat to Challenger. But at node 3, Defender can also threaten to withhold support from Protégé should Protégé hold firm at node 2. Finally, at node 4b, Challenger's choice to press its demands and precipitate *Conflict* threatens both Defender and Protégé.

To put all this in a slightly different way, the Tripartite Crisis Game is a three-player game in which the choices of each player can be sanctioned by at least one other player. Also, the choices are dynamically interdependent. Challenger's best choice depends on those of Protégé and Defender; Defender's depends on Protégé's and Challenger's; and Protégé's on Defender's and Challenger's. The fact that none of the players can afford to ignore the choices of any of the others is an important feature of the Tripartite Crisis Game.

### Preferences

Given the variety of circumstances and constraints faced by states, it is well nigh impossible to specify a set of preferences that are universally held. Still, some assumptions are obvious. For instance, a player whose highest-ranked outcome is the status quo cannot properly be considered a Challenger. Similarly, a player who does not place a relatively high value on both the status quo and the welfare of a third state fails to meet the minimal requirements of a Defender. And finally, as a Protégé, by definition, depends on the assistance of another for its survival, it must be the case that a Protégé prefers those outcomes that occur when it is supported by a third party to those that come about when it is left to fend for itself.

Within the confines of the Tripartite Crisis Game, however, there are still a number of critical preference relationships that are not implicit in the names or the roles of the players. In these instances, the specific assumptions we make are guided by both theoretical and pragmatic considerations. Theoretically, our assumptions about certain preference relationships are intended to highlight and intensify the deterrence-versus-restraint dilemma in international politics. Pragmatically, we focus on a special case that makes our analysis tractable. In other words, we do not claim that the particular preference assumptions we make are those of all Challengers, all Defenders, and all Protégés involved in extended deterrence relationships. Rather, they serve as boundary conditions that delimit the empirical circumstances to which our model applies.

Table 1 summarizes the specific assumptions we make about the player's preferences and assigns hypothetical cardinal utility values to each outcome. Later, we use these hypothetical values to illustrate the theoretical implications of our model and its assumptions. For now we consider only preferences—which are rank ordered in each column from most to least preferred. We begin with Challenger.

As Table 1 reveals, we assume that there are three outcomes that Challenger prefers to the *Status Quo*. All three—*Challenger Wins*, *Protégé Realigns*, and *Protégé Loses*—involve Protégé's isolation and forced capitulation to Challenger's demands. We further assume (somewhat arbitrarily) that Challenger prefers a concession that is immediate and relatively costless (that is, *Challenger Wins*) to a delayed concession that, by needlessly embarrassing Defender, could imply future costs.<sup>5</sup> Also, we assume that, *ceteris paribus*, Challenger prefers to disrupt the relationship between Defender and Protégé, that is, its preference is for *Protégé Realigns* over *Protégé Loses*.

<sup>5</sup>Both Germany and Austria seemed oblivious to these costs during the Bosnian crisis of 1908–1909. By contrast, France and Britain were very sensitive to them in 1914.

TABLE I. Initial Preference Assumptions for Tripartite Crisis Game

| Challenger:                          |               | Protégé:                                   |                | Defender:                            |               |
|--------------------------------------|---------------|--|----------------|--------------------------------------|---------------|
| Preferences                          | Utilities     | Preferences                                | Utilities      | Preferences                          | Utilities     |
| <i>Challenger Wins</i>               | $c_{CW} = 10$ | <i>Status Quo</i>                          | $e_{SQ} = 10$  | <i>Status Quo</i>                    | $d_{SQ} = 10$ |
| <i>Protégé Realigns</i>              | $c_{PR} = 8$  | <i>Challenger Concedes</i>                 | $e_{CC} = 8$   | <i>Challenger Concedes</i>           | $d_{CC} = 8$  |
| <i>Protégé Loses</i>                 | $c_{PL} = 6$  | <i>Conflict</i>                            | $e_C = 4$      | <i>Challenger Wins</i>               | $d_{CW} = 4$  |
| <i>Status Quo</i>                    | $c_{SQ} = 4$  | <i>Challenger Wins</i>                     | $e_{CW} = 2$   | <i>Protégé Loses</i>                 | $d_{PL} = 2$  |
| <i>Conflict</i><br>(when determined) | $c_{C+} = 1$  | <i>Protégé Realigns</i><br>(when disloyal) | $e_{PR+} = 1$  | <i>Conflict</i><br>(when staunch)    | $d_{C+} = 1$  |
| <i>Challenger Concedes</i>           | $c_{CC} = 0$  | <i>Protégé Loses</i>                       | $e_{PL} = 0$   | <i>Protégé Realigns</i>              | $d_{PR} = 0$  |
| <i>Conflict</i><br>(when hesitant)   | $c_{C-} = -1$ | <i>Protégé Realigns</i><br>(when loyal)    | $e_{PR-} = -1$ | <i>Conflict</i><br>(when perfidious) | $d_{C-} = -1$ |

Of the remaining three outcomes, we assume that Challenger prefers the *Status Quo* to either *Conflict* or *Challenger Concedes*. We remain silent, however, about Challenger's preferences about its two lowest-ranked outcomes that decide its type: when Challenger is *determined*, it prefers *Conflict* to *Challenger Concedes*; a *hesitant* Challenger has the opposite preferences.

Let us turn next to Protégé's preference. In our model, Protégé's three most preferred outcomes are *Status Quo*, *Challenger Concedes*, and *Conflict*, in that order. None of these outcomes involves a sure concession to Challenger. But the *Status Quo* involves no loss or evident costs. By contrast, when *Challenger Concedes*, Protégé's dependence on Defender is revealed; also, a future obligation to Defender may be implied. Finally, *Conflict* carries with it the possibility of a defeat and an eventual concession. As an immediate concession involves certified costs, we further assume that Protégé always prefers *Conflict* to *Challenger Wins*. Protégé's two worst outcomes follow abandonment by Defender. When Protégé is *loyal*, it prefers *Protégé Loses* to *Protégé Realigns*. A Protégé with the opposite preference is called *disloyal*.

Finally, consider Defender's preferences. As they are strategic partners, Defender and Protégé have some common interests. Specifically, we assume that, like Protégé, Defender's two highest-ranked outcomes are *Status Quo* and *Challenger Concedes*, respectively. But this is as far as it goes. After this, Defender's and Protégé's interests diverge.

At *Challenger Wins*, Protégé capitulates to Challenger, but at a minimum cost to Defender. Hence, we assume that Defender prefers *Challenger Wins* to the remaining four outcomes.<sup>6</sup> We further assume that Defender always prefers to maintain its relationship with Protégé, that is, it prefers *Protégé Loses* to *Protégé Realigns*. Yet to be established, however, is Defender's preference for *Conflict* relative to *Protégé Loses* and *Protégé Realigns*.

One could reasonably assume that Defender is strongly committed to Protégé, preferring *Conflict* to *Protégé Loses* and *Protégé Realigns*. But extended deterrence relationships defined by this boundary condition are theoretically uninteresting. A Defender with this preference always supports Protégé at node 3. Anticipating support, Protégé always holds firm at node 2. As a consequence, Challenger always chooses not to make a demand at node 1. Put differently, whenever Defender prefers *Conflict* to *Protégé Loses*, extended deterrence always succeeds and the outcome is always *Status Quo*. To avoid a trivial analysis, therefore, we assume that Defender prefers *Protégé Loses* to *Conflict*. Finally, Defender's preference for its two

<sup>6</sup>Thus, we implicitly assume that the stakes are of little interest to Defender. Bismarck's famous comment about Germany's interest in the Balkans immediately comes to mind.

lowest-ranked alternatives determines its type. A Defender that prefers *Conflict* to *Protégé Realigns* is called *staunch*. The preferences of a *perfidious* Defender are the other way around.

To summarize briefly: the version of Tripartite Crisis Game we consider presently is defined by the game tree given by Figure 1 and the preference assumptions listed in Table 1. The model was specifically designed to incorporate the key elements of Snyder's theoretically distinct "alliance" and "adversary" games and to capture the essence of the deterrence-versus-restraint dilemma. Thus, Defender's node 3 choice, which impacts and is impacted by Challenger's choices at nodes 1 and 4b and by Protégé's choices at nodes 2 and 4b, will oftentimes put Defender on the spot. By pledging strong support for Protégé, Defender may make it less likely that Challenger will make a demand at node 1 but at the expense of also increasing the probability that Protégé will hold firm at node 2 should deterrence breakdown. Conversely, half-hearted support will in fact make Protégé less rigid but it will also embolden Challenger.

Defender is not the only player with a dilemma in this game. Protégé's threat to realign (at node 4a) will tend to elicit Defender's support at node 3. But when a Challenger is motivated to disrupt a strategic partnership, as is the Challenger in the Tripartite Crisis Game, Protégé's threat may also prompt a crisis (Zagare and Kilgour 2003). Finally, note that Challenger's node 1 choice is equally complex. Its decision whether to contest the *Status Quo* depends on its expectations about the interdependent choices of each of its opponents.

### Information and Utility Assumptions

The decision-making environment of the Tripartite Crisis Game is theoretically rich but conceptually complex. To gain tractability, we begin our present analysis with two sets of simplifying assumptions.

We begin with several assumptions about the information structure of the game. First, we assume that Defender is staunch and that this information about Defender's type is common knowledge. We also assume that all remaining information concerning the Tripartite Crisis Game is common knowledge with the exception that the players are uncertain about Challenger's and Protégé's types.<sup>7</sup> In other words, we assume that the players are aware of the game in extensive form as given by the game tree of Figure 1 and the utility assumptions summarized in Table 1 and discussed below.

As the Tripartite Crisis Game is intrinsically more interesting when Defender is known to be staunch than when Defender is known to be perfidious, the latter assumption is easy to defend. Against a perfidious Defender, Protégé's node 4a threat to realign is unduly diminished and Challenger's node 4b threat to press on is unnecessarily enhanced. Thus, the assumption that Defender's preference for *Conflict* over *Protégé Realigns* is common knowledge equalizes the impact of Challenger's and Protégé's threats that are directed against Defender. As a consequence, the deterrence-versus-restraint dilemma of Defender is intensified.

Second, we make what we consider to be several plausible, albeit arbitrary, assumptions about the players' utilities. In our opinion, the specific assumptions we make will not materially affect our conclusions. To be sure, a more general analysis is likely to yield additional strategic possibilities. But, based on our previous research (Zagare and Kilgour 2000), we are confident that these possibilities will be complex yet minor variations on those uncovered in the numerical example to be offered here.

<sup>7</sup>Players are assumed to know their own type.

Before proceeding, however, we must first pause to explain the notation for the utility values given for Challenger, Protégé, and Defender in the second, fourth, and sixth columns of Table 1, respectively. In general, we denote the utility of outcome  $O$  to Challenger by  $c_O$ , to Protégé by  $e_O$ , and to Defender by  $d_O$ . For example, Challenger's utility for the *Status Quo* is  $c_{SQ}$ . But, to account for the different player types, we assign two different values for the utility of one critical outcome for each player as follows:

- $c_{C+}$  is a determined Challenger's utility for Conflict
- $c_{C-}$  is a hesitant Challenger's utility for Conflict
- $e_{PR+}$  is a disloyal Protégé's utility for Protégé Realigns
- $e_{PR-}$  is a loyal Protégé's utility for Protégé Realigns
- $d_{C+}$  is a staunch Defender's utility for Conflict
- $d_{C-}$  is a perfidious Defender's utility for Conflict

The hypothetical utilities given in Table 1 range from a high of 10 to a low of  $-1$ . Thus, in our numerical example, both Protégé ( $e_{SQ} = 10$ ) and Defender ( $d_{SQ} = 10$ ) place a high value on maintaining the *Status Quo* relative to the other possible outcomes of the game. Challenger, by contrast, is comparatively dissatisfied with the existing order ( $c_{SQ} = 4$ ). Similarly, both types of Challengers ( $c_{C+} = 1$ ,  $c_{C-} = -1$ ) and both types of Defenders ( $d_{C+} = 1$ ,  $d_{C-} = -1$ ) have relatively low utilities for *Conflict*. It is entirely understandable, however, that Protégé's evaluation of *Conflict* ( $e_C = 4$ ) is comparatively higher than that of both Challenger and Defender. After all, it is Protégé who stands to lose the most if it is forced to accommodate Challenger.

### Analysis

In this section, we offer a general discussion of the strategic properties of the Tripartite Crisis Game under *incomplete information*. More specifically, we identify and describe all *perfect Bayesian equilibria* of the game, given the utility and information assumptions outlined earlier.<sup>8</sup> Our analysis is informal; many fine points are omitted. A comprehensive Appendix, containing a formal treatment and all relevant technical details, is available from the authors on request.

To start, note that in the Tripartite Crisis Game, the choices of Protégé and Challenger at nodes 4a and 4b, respectively, are strictly determined by each player's type. Specifically, at node 4a, a disloyal Protégé always chooses to realign and a loyal Protégé always choose not to; similarly, at node 4b, a determined Challenger always presses on, while a hesitant Challenger always backs down. Notice also that if it should happen that Protégé is faced with a decision at node 2, it will have observed Challenger's prior action choice at node 1. And should the game reach node 3, Defender will have observed previous action choices by Challenger and Protégé. Protégé can use its observations to update its belief about Challenger's type before making its choice at node 2. Similarly, Defender will have an opportunity to update its beliefs about Challenger's and Protégé's types before its node 3 decision.

Given these considerations, it follows that a perfect Bayesian equilibrium of the Tripartite Crisis Game will consist of a seven-tuple of probabilities  $[x_D, x_H, r; y_D, y_L, q; z]$  where:

- $x_D$  = the probability that a determined Challenger makes a demand at node 1
- $x_H$  = the probability that a hesitant Challenger makes a demand at node 1

---

<sup>8</sup>A perfect Bayesian equilibrium is the standard solution concept of a game with incomplete information. A perfect Bayesian equilibrium consists of a plan of action (that is, a strategy) for each player, plus the player's beliefs about (that is, subjective probabilities concerning) other players' types, such that each player (1) always acts to maximize its expected utility given its current beliefs, and (2) always updates those beliefs rationally (that is, according to the Bayes's Rule), given the actions it observes during the play of the game.



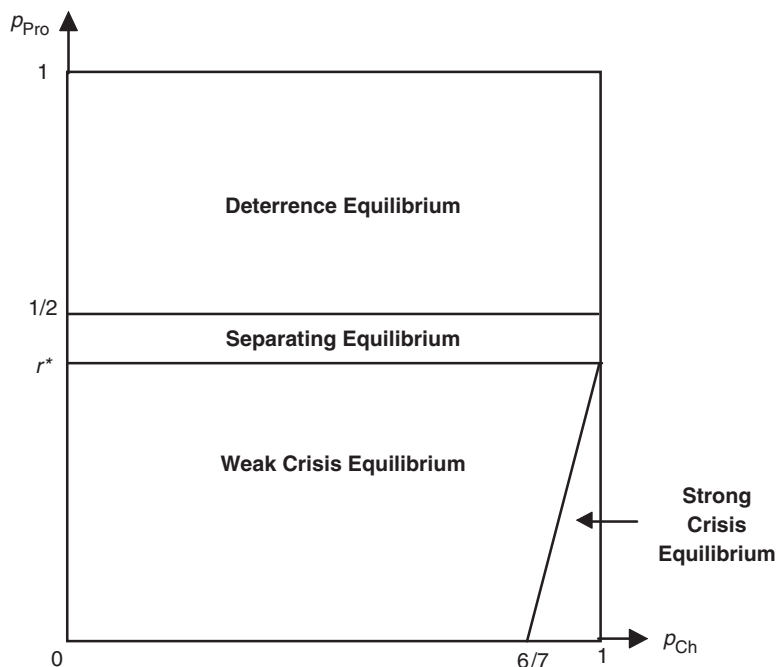


FIG. 2. Location of Plausible Equilibria in the Tripartite Crisis Game with Incomplete Information When Defender Is Staunch

- $r$  = the updated probability that Challenger is determined, given that Challenger makes a demand at node 1
- $y_D$  = the probability that a disloyal Protégé will choose to hold firm at node 2
- $y_L$  = the probability that a loyal Protégé will choose to hold firm at node 2
- $q$  = the updated probability that Protégé is disloyal, given that Protégé holds firm at node 2
- $z$  = the probability that (a staunch) Defender will choose to support Protégé at node 3.

Five of the probabilities ( $x_D$ ,  $x_H$ ,  $y_D$ ,  $y_L$ , and  $z$ ) are strategic variables that specify Challenger's, Protégé's, and Defender's choices at nodes 1, 2, and 3, respectively, contingent on their type. (Because Defender is known to be staunch, its node 3 choice is represented by a single variable,  $z$ ). The remaining variables ( $r$  and  $q$ ) are the a posteriori probabilities that are updated after certain action choices by Challenger (at node 1) and Protégé (at node 2) are observed.

As it turns out, there are nine distinct nontransitional perfect Bayesian equilibria in the numerical example of the Tripartite Crisis Game that we are examining here.<sup>9</sup> Not all these equilibria are associated with beliefs that are plausible. We focus on the four that are: Deterrence, Separating, Weak Crisis, and Strong Crisis equilibria. Table 2 summarizes their technical properties. Figure 2 contains a graphical summary of the existence conditions for the plausible equilibria. When only plausible equilibria are considered, each perfect Bayesian equilibrium of the Tripartite Crisis Game is unique.<sup>10</sup>

<sup>9</sup>Hereafter, we drop the qualification.

<sup>10</sup>For further details, see the Appendix that is available from the authors at fczagare@buffalo.edu or mkilgour@wlu.ca.

TABLE 2. (Plausible) Perfect Bayesian Equilibria for the Tripartite Crisis Game

|                                  | Challenger |               |                  | Protégé |   |                         | Defender                |
|----------------------------------|------------|---------------|------------------|---------|---|-------------------------|-------------------------|
|                                  | $x_D$      | $x_H$         | $r$              | $y_D$   | $y_L$                                   | $q$                     |                         |
| <i>Deterrence Equilibrium</i>    | 0          | 0             | <i>Arbitrary</i> | 1       | 1                                       | $p_{Pro}$               | 1                       |
| <i>Separating Equilibrium</i>    | 1          | 0             | 1                | 1       | $\frac{p_{Pro}}{1 - p_{Pro}}$           | 1/2                     | 1/2                     |
| <i>Weak Crisis Equilibrium</i>   | 1          | $0 < x_H < 1$ | $p_{Ch} < r < 1$ | 1       | $0 < y_L < 1$                           | $p_{Pro} < q < 1$       | $0 < z < 1$             |
| <i>Strong Crisis Equilibrium</i> | 1          | 1             | $p_{Ch}$         | 1       | $\frac{p_{Pro}(1 - q)}{q(1 - p_{Pro})}$ | $\frac{7p_{Ch} - 6}{2}$ | $\frac{1}{4 - 2p_{Ch}}$ |

To describe the equilibria, a few additional assumptions are required. Assume now that Challenger is determined with probability  $p_{Ch}$  (where  $0 < p_{Ch} < 1$ ), that Protégé is disloyal with probability  $p_{Pro}$  (where  $0 < p_{Pro} < 1$ ), and that the values of  $p_{Ch}$  and  $p_{Pro}$  are known to all players. These assumptions allow us to model the consequences of the players’ uncertainty about Challenger’s and Protégé’s types. We continue to assume that the realized value of  $p_{Ch}$  ( $p_{Pro}$ ) is known to Challenger (Protégé), that is, that each player knows its own type, and that Defender’s type (that is, staunch) is common knowledge.

Recall that in the Tripartite Crisis Game, Protégé’s choice at node 4a and Challenger’s choice at node 4b represent threats. Recall also that a disloyal Protégé always realigns at node 4a, and that a determined Challenger always presses on at node 4b. It is quite natural, therefore, to interpret the two belief variables ( $p_{Pro}$  and  $p_{Ch}$ ) as measures of the credibility of Protégé’s and Challenger’s end-game threats, respectively (Zagare and Kilgour 2000): the higher the  $p_{Pro}$ , the more likely/credible Protégé’s threat is to realign; the higher the  $p_{Ch}$ , the more likely/credible Challenger’s threat is to press on.

*Deterrence Equilibrium*

A Deterrence Equilibrium is any equilibrium under which Challenger never makes a demand at node 1, that is, where ( $x_D = x_H = 0$ ). There are four distinct Deterrence equilibria in the Tripartite Crisis Game. But we restrict our attention to the only one that is based on beliefs we consider plausible. Here, we refer to it as **the Deterrence Equilibrium**.

As Table 2 shows, under the Deterrence Equilibrium, both loyal and disloyal Protégés always intend to realign at node 4a ( $y_D = y_L = 1$ ), and Defender always intends to support Protégé at node 3 ( $z = 1$ ). Thus, it should not be very surprising that Challenger never issues a demand at node 1. Whenever the Deterrence Equilibrium is in play, extended deterrence success is assured.

Because there is no overt behavioral marker, deterrence success is, in some sense, impossible to observe. This explains why, under the Deterrence Equilibrium, there is no opportunity for Protégé or Defender to update their beliefs about Challenger’s type. This is why Protégé’s and Defender’s updated belief ( $r$ ) about Challenger’s type is the same as their initial belief, and why Defender’s a posteriori belief ( $q$ ) is always equivalent to its a priori belief ( $p_{Pro}$ ) about Challenger’s type. Clearly, Defender does not face a dilemma whenever the Deterrence Equilibrium is in play. To explain exactly when the dilemma is avoided, we now consider Figure 2, which summarizes graphically the existence conditions of all plausible perfect Bayesian equilibria in the Tripartite Crisis Game.

Along the horizontal and vertical axes of Figure 2 are graphed, respectively, the two belief variables:  $p_{Ch}$  and  $p_{Pro}$ . As mentioned above, the belief variables can be

interpreted as measures of the credibility of Challenger's ( $p_{Ch}$ ) and Protégé's ( $p_{Pro}$ ) threats. Also graphed are several constants that are convenient thresholds for categorizing and interpreting the equilibria. Here, we refer to these constants only in passing; the interested reader should acquire the Appendix, mentioned previously, for more detail.

As Figure 2 reveals, the Deterrence Equilibrium exists, uniquely, whenever  $p_{Pro}$  is relatively large, that is, whenever Defender believes it likely that Protégé is disloyal and, hence, likely to realign (at node 4a) should it withhold support (at node 3). (In the context of our numerical example, a strong Deterrence Equilibrium will exist as long as  $p_{Pro} \geq 1/2$ ).

Notice that the existence of the Deterrence Equilibrium does not depend on the value of  $p_{Ch}$ , the probability that Challenger is determined and intends to press on at node 4b. In other words, the credibility of Challenger's end-game threat is irrelevant to the existence of the Deterrence Equilibrium. Only Protégé's threat to realign matters. In the Tripartite Crisis Game, if this threat is sufficiently credible, the Deterrence Equilibrium will be the only plausible equilibrium that exists.

### *Separating Equilibrium*

The second distinct equilibrium form in the Tripartite Crisis Game is the Separating Equilibrium. This equilibrium separates Challengers by type: determined Challengers always contest the *Status Quo*; hesitant Challengers never do. Thus, under a Separating Equilibrium, Challenger's type is revealed by its node 1 choice. As a consequence, the updated probability that Challenger is determined given that it makes a demand,  $r$ , always equals 1.

As Figure 2 indicates, the Separating Equilibrium exists when Protégé's threat to realign falls just below the level necessary to support the Deterrence Equilibrium. (In our hypothetical example, when  $p_{Pro} < 1/2$ ). At the same time, Protégé's threat must be credible enough that a hesitant Challenger will always be deterred (that is, when  $p_{Pro} > r^*$ ). To achieve this result, a loyal Protégé must be prepared to realign at node 4a with a probability that is sufficient to induce Defender's likely support at node 3. In the final analysis, it is the strong likelihood that Defender will stand with Protégé that is key to the existence of a Separating Equilibrium.

As with the Deterrence Equilibrium, the existence of the Separating Equilibrium does not depend on the credibility of Challenger's threat. But this is not unexpected. Under a Separating Equilibrium, Protégé's threat to realign is credible enough that only determined Challengers contest the *Status Quo*. This means that when a Separating Equilibrium is in play, Protégé and Defender will know for sure what Challenger will do at node 4b should they be faced with a decision in the Tripartite Crisis Game. As a consequence, Protégé's node 2 choice will be fully determined by its estimate of Defender's likely behavior at node 3, and Defender's node 3 choice will rest completely on its estimate of Protégé's behavior at node 4a. Stated differently, given that Challenger makes a demand at node 1, the only choices of consequence under a Separating Equilibrium are those of Protégé and Defender.

### *Weak and Strong Crisis Equilibria*

Such is clearly not the case, however, when Protégé's credibility falls below the minimum level needed to support a Separating Equilibrium (that is, when  $p_{Pro} < r^*$ ). At the lowest levels of Protégé's credibility, Challenger's credibility interacts with Protégé's to determine which of two plausible perfect Bayesian equilibria will exist.

When Protégé's credibility is low, either a Weak or a Strong Crisis Equilibrium will exist. Under either form, disloyal Protégés always hold firm at node 2 while loyal Protégés bluff by holding firm probabilistically. Also, Defender hedges its bet

by sometimes, but not always, backing Protégé at node 3. Finally, determined Challengers always contest the *Status Quo* at node 1 under either the Weak or the Strong Crisis Equilibrium.

It is the behavior of a hesitant Challenger that distinguishes the two crisis equilibria. Under the Weak form, a hesitant Challenger probes the limits of Protégé's and Defender's resolve with specified probabilities. But under the Strong Crisis Equilibrium, a hesitant Challenger always demands an adjustment of the *Status Quo*.

As one might expect, and as Figure 2 shows, the Strong Crisis Equilibrium exists only at the highest level of Challenger credibility. In effect, the high likelihood that Challenger will press on at node 4b reduces both the probability that Protégé will hold firm at node 2 and the probability that Defender will support Protégé at node 3. In effect, Challenger is able to deter its opponents with a high probability so that even a hesitant Challenger is motivated to instigate a crisis.

Notice from Figure 2 that the threshold separating the Weak and Strong Crisis Equilibria is positive, sloping upward and to the right. This slope reflects the fact that higher and higher levels of Challenger credibility are needed to support a Strong Crisis Equilibrium as Protégé's credibility increases. Conversely, as the credibility of Protégé's threat declines, lower and lower levels of Challenger credibility are required to deter both Protégé and Defender.

### Discussion

Our conclusion that extended deterrence success is most likely when Protégé's threat to realign is most credible appears to run counter to Crawford's (2003:1–2) argument that “pivotal deterrence tends to work when the adversaries have bad alignment options, or none at all, and it tends to fail when they have good alignment options.”<sup>11</sup> These starkly divergent assessments of the strategic impact of alignment patterns and the possibility of extended deterrence success raise some interesting questions. First, how can we account for the contradictory hypotheses? Second, can the two points of view be reconciled? And, third, is Crawford's case study of the July crisis of 1914 supportive of our response to the first two questions? In what follows, we attempt to answer these three questions serially.

One possible explanation for the difference between Crawford's empirical generalization and the logical argument we derive from an analysis of the Tripartite Crisis Game is that the theoretical domain of the two studies is different. As discussed earlier, our conclusions apply only when the boundary conditions established by the particular preference assumptions we make are satisfied. By contrast, Crawford's focus is somewhat broader. Specifically, his definition of “pivotal deterrence” includes three distinct triangular dilemmas or scenarios. Only one of these, the straddle strategy, corresponds to the assumptions that frame our analysis of the Tripartite Crisis Game. It should be clear, however, that domain differences cannot account fully for the divergent assessments. Because the situations of extended deterrence we model are a subset of Crawford's broader conceptualization of pivotal deterrence relationships, his empirical generalizations should subsume ours. As a consequence, we must look elsewhere for an explanation of the theoretical contradiction.

Another possibility is that the Tripartite Crisis Game model is too austere to generate the fine-grained conclusions that Crawford derives from his carefully crafted analysis of the British attempt to restrain France and to deter Germany in 1914, the only one of Crawford's four detailed case studies that he associates with the straddle strategy. Clearly, the July crisis was not technically a three-player game. Just before World War I, there were five major players whose policies were, to varying degrees, relevant to the outcome of the crisis. Perhaps, Crawford's

<sup>11</sup>Snyder (1984) makes a similar argument.

conclusions are drawn in part from a case analysis that is too nuanced to be captured by the strategic structure of the Tripartite Crisis Game.

We do not find this explanation for the differences between our conclusions and Crawford's particularly compelling either. In principle, the analysis of the Tripartite Crisis Game model should be able to take account of the policies of both Austria-Hungary and Russia, albeit indirectly, through the utility functions of Germany (that is, Challenger) and France (that is, Protégé). But even were this not the case, we believe that our three-person game model should remain relevant to an analysis of British policy during the run up to World War I. As Crawford (2003:99) correctly notes, "if Britain's army cut a meager figure on the Western front, it cut no figure at all in the East: Russia would not gain any immediate benefits from British allegiance in a short war, and Austria would not suffer any immediate costs. It is not surprising therefore that British pivotal policy had no discernable effect on the behavior of those powers in the July crisis."

If the discrepancy between Crawford's and our conclusions cannot be traced to either domain differences or modeling simplifications, how can we explain it? The short answer is that we believe that Crawford's empirical generalization itself is the culprit: his explanation is incomplete, perhaps even misleading. To develop this point, we begin by observing we believe that the Tripartite Crisis Game is an appropriate model for examining the strategic dynamic just before the First World War. While the fit between any model and actual event will never be exact, we believe that in this instance it is close enough. Three of the real-world actors involved in the July crisis, Germany, Great Britain, and France, played roles that correspond, broadly, to the players in our game. Like the Defender of our model, Britain's interests were not directly at risk. Like the Challenger of our model, Germany was clearly dissatisfied with what it perceived to be a deteriorating status quo. And like the Protégé of our model, France's choice would be critical in determining whether there would be an acute crisis that potentially placed Defender in harm's way.

The similarities between the Tripartite Crisis Game and the July crisis do not end here. Because the British had no real interest in the Balkans and were not particularly invested in protecting Serbian sovereignty, they sought a resolution to the crisis that avoided the costs associated with both conflict and strategic isolation. To achieve this end, Britain had to restrain France and deter Germany simultaneously. Sir Edward Grey's straddle strategy, which is consistent with several of the perfect Bayesian equilibria of our model, was designed to minimize the probability of a general European war. Thus, we accept Crawford's (2003:78, 82) point that "Grey's approach to the July crisis expressed a coherent strategic logic . . . [and that it] . . . did not fail *because* it was ambiguous." We part company, however, with his further contention that "it failed because the leverage that usually flows from such a policy was enfeebled by France's and Germany's strong alignment options (Russia and Austria)."

To sustain this argument, Crawford would have to show that (1) a Defender's straddle strategy was successful in a number of confrontations that were similar to the July crisis, except for the fact that either Protégé or Challenger, or both, lacked strong alignment options, and (2) that the availability of alignment options was the decisive difference in 1914. To be sure, as Crawford (2003:84) points out, before 1914, Grey's "entente policy yielded a track record of successes." Yet, it remains to be shown that the critical differences between Grey's successes and the breakdown of deterrence in 1914 had anything to do with France's and Germany's alignment options.<sup>12</sup>

---

<sup>12</sup>It is important to note here that we are speaking of the expansion of a localized Austro-Serbian war that possibly included Russia into a wider war that involved France, Germany, and possibly Britain.

It is difficult to sustain the claim that the availability of strong alignment options was the critical difference between prior successes and the eventual failure of Grey's policy. During World War I, Austria–Hungary added little to German capability, except initially on the Eastern Front. By way of contrast, without Great Britain, France (even with Italy) was no match for Germany on the Western Front (Kugler and Domke 1986:60–63). In other words, abstracting away some of the particulars of the July crisis, it is highly doubtful that Germany would have backed away from a war with France in 1914 simply because it lacked support from Austria. It is similarly dubious that, when push came to shove, France would have behaved much differently in 1914 were it not for its alliance with Russia. If Grey's straddle strategy did not fail in 1914 because it was ambiguous or because France and Germany had strong alignment options, why did it fail? Our analysis of the equilibrium structure of the Tripartite Crisis Game provides some important clues.

We begin by noting what might seem to be an essential difference between the Tripartite Crisis Game model and the situation in Europe during July 1914. Recall that in the model, Protégé makes a realignment decision at node 4a that is a critical determinant of Defender's node 3 decision. Depending on Protégé's choice, one of two outcomes is possible. The relationship of each of these outcomes with *Conflict* is significant for Defender. As a staunch Defender prefers *Protégé Loses* to *Conflict*, but prefers *Conflict* to *Protégé Realigns*, Defender is more likely, *ceteris paribus*, to support Protégé the more likely it is that Protégé is disloyal and intends to realign at node 4a. There appears to have been no analogous choice for France during the July crisis (Kiesling 2003:227–228).<sup>13</sup>

Nonetheless, on the eve of war in 1914, Britain, like the Defender in our model, had to evaluate the consequences of abandoning its ally. Were Britain to stand aside, French sovereignty was likely to be compromised, one way or another. But it is not possible to say that Britain would have fought just to save France. Writing in 1906, Grey remarked, "if there is a war between France and Germany, it will be very difficult for us to stay out of it . . . . On the other hand the prospect of a European war and of our being involved in it is horrible" (quoted in Harris 2003:271).<sup>14</sup>

Notice that Grey did not say that it was impossible for Britain to remain neutral, only that it would be difficult to do so. From the British point of view, neither alternative was especially attractive. Thus, caught between a rock and a hard place, Grey adopted what amounted to a mixed strategy, leaving open exactly what the government might or might not do as well as under what circumstances. There is no dispute about this. As Donald Kagan (1995:211) observes, "not only could Britain's friends and enemies not be sure what the British would do until the last minute, the British themselves did not know."

Reasoning backwards, however, we know that on August 4, the British issued an ultimatum demanding that Germany desist from violating Belgium's neutrality. And previously, on July 30, Grey had personally rejected Germany's offer not to annex any part of France or Belgium in return for British neutrality. "Poor little Belgium" was the likely key to Britain's choice. As Harris (2003:299, n. 119) speculates, it is entirely possible that "had the Germans avoided Belgium, the British might, in effect, have ditched the *entente*."

Of course, until it occurred, Grey did not know for sure that Germany would invade Belgium. Thus, in late July, there was considerable uncertainty in London about the precise consequences of British neutrality. In the Tripartite Crisis Game, the uncertain implications of Defender standing aside are associated with uncertainty about Protégé's type. In the real world, however, this uncertainty might spring from another source. Such was the case in 1914 where  $p_{Def}$ , interpreted

<sup>13</sup>This is not to say that Germany was uninterested in splitting the Entente (Fischer 1974:23).

<sup>14</sup>According to J. Paul Harris (2003:271), "Grey's reading of these matters would not change, in essentials, between February 1906 and August 1914."

loosely as British uncertainty about the implications of abandoning France, was middling, at best.

The inordinate uncertainty that framed British decision making in July 1914 is consistent with the existence of a Separating, a Weak Crisis, or even a Strong Crisis Equilibrium. Under any of these equilibria, Defender's optimal strategy, as was Britain's, is mixed, which is why we do not contest Crawford's assessment of the strategic appropriateness of Grey's straddle strategy. In other words, given the operative constraints, British policy was sound. But it was not fail-safe. As our analysis of the Tripartite Crisis Game reveals, a mixed (that is, a straddle) strategy never deters a determined Challenger when Defender is known to be staunch. Under any of these equilibria, determined Challengers always contest the *Status Quo* at node 1 and always choose *Conflict* at node 4b. That Germany was determined in 1914 is undeniable (Fischer 1974:23). By ignoring the British ultimatum, it revealed both its preference for war over backing down and its type. Early in July, the Kaiser had boasted, "this time I shall not chicken out" (Kagan 1995:190). On August 4, he was true to his word.

Against a determined Germany, Grey's policy, while rational, was destined to fail over the long haul. But it did not fail because of the particular alignment options available to Germany and France. Rather, Grey's policy failed because after Sarajevo, German leaders, but especially the Kaiser, had come to prefer a general European war to backing down (that is, to *Challenger Concedes*).<sup>15</sup> Given the facts on the ground on August 4, Holger Herwig (2003:186–187) concludes that Germany's preference for general war is inexplicable, perhaps even criminal. Nonetheless, he also finds it real and incontrovertible.

To put all this in a slightly different way, the breakdown of deterrence that culminated with the invasion of Belgium had less to do with British policy than it did with German intentions—intentions that may be difficult to explain or justify. Also, it is dubious that these intentions were conditioned in any meaningful way by Germany's alliance with Austria–Hungary. Rather, facing an uncertain future, German leaders took "a calculated risk." Ultimately, their "all-out gamble" failed (Fischer 1974:55).

Of course, Britain did not know for sure that Germany was determined. Nonetheless, it is reasonable to ask whether another approach might have prevented a world war. Indeed, a persistent criticism of Grey's policy was that it lacked clarity. For example, in memoirs published after the war, the Russian Foreign Minister Serge Sazonov (1928:40) remarked:

I cannot refrain from expressing the opinion that if in 1914 Sir Edward Grey had, as I insistently requested him, made a timely and equally unambiguous announcement of the solidarity of Great Britain with France and Russia, he might have saved humanity from that terrible cataclysm, the consequences of which endangered the very existence of European civilization.

Similarly, Snyder (1984:482) speculates that if "early in the crisis, he [Grey] had declared his firm support of Russia and France and warned Germany unequivocally . . . the war might have been averted." More generally, James Fearon (1997:71) wonders "why we sometimes observe halfhearted signals when convincing ones are possible?"

The facile response would be simply to agree with this line of criticism. After all, there exists a plausible Deterrence Equilibrium in the Tripartite Crisis Game under

<sup>15</sup>It is possible that German leaders, in fact, preferred general war to the status quo, in which case Fischer (1974) and several others are correct: World War I was "inevitable" and Germany could not have been deterred in 1914 (see also Snyder and Diesing's 1977:93–95 discussion of German preferences). Germany's behavior is consistent with either interpretation of its preferences. But a preference for war over the status quo cannot be established empirically.

which even determined Challengers are deterred. Significantly, under this form of equilibrium, Defender always supports Protégé and Challenger is always deterred. In principle, therefore, the alternative policy recommended by Grey's critics is, under certain conditions, consistent with rationality.

We do not believe that such conditions existed in July and early August 1914. Recall that the existence of the Deterrence Equilibrium is based on the premise that Defender is known to be staunch. As we show elsewhere (Zagare and Kilgour 2003), when Defender is known to be of the perfidious type, deterrence is unlikely whenever Challenger is determined. It is obvious that until August 3, if Britain was staunch, it was barely staunch. Reflecting the small margin that may have type cast Britain was the strong division that existed within the British cabinet. Many members of H.H. Asquith's Liberal party, including the Prime Minister himself, initially opposed British participation in the war under any conditions. And until German troops marched into Belgium, British public opinion also was decidedly in favor of remaining neutral. It is small wonder, then, that Grey's policy was equivocal and that his strategy was mixed.

Given the state of public opinion in Britain, it seems safe to conclude that until German troops finally entered Belgium on August 4, Grey could not rationally, that is, credibly, threaten to intervene on France's behalf. After all, as Vesna Danilovic (2002) demonstrates, when states fight to protect an ally, they generally do so only when the stakes are high. In 1914, a critical majority of British and German policymakers simply did not believe that the stakes were, in fact, high enough to engage the British. Indeed, German Chancellor Bethmann-Hollweg was surprised that Britain went to war for "just a scrap of paper."

That Bethmann-Hollweg was surprised by the British response also helps to shed some theoretical light on Germany's choice in 1914. For most of the nineteenth century, Britain had stood apart from continental politics and intrigues, eschewing, as Palmerston's dictum goes, both eternal allies and perpetual enemies. But in the post-Bismarck era, Germany's foreign policy (*Weltpolitik*) had turned in a direction that potentially threatened British interests. Under these conditions, Britain could no longer maintain its splendid isolation. As a consequence, it began to pull closer to France and, later, Russia. Britain's *entente* (straddle) strategy did not change when Grey became Foreign Secretary in 1905. In fact, Grey's implementation of the policy was even more artful than that of his immediate predecessor, Lord Lansdowne. And, until 1914, the policy seemed to work. Kagan (1995:173), for example, attributes the resolution of the second Moroccan crisis in 1911 to "Grey's policy of restraint on France and firmness toward Germany."

Nonetheless, although a war did not occur among the great powers, general deterrence broke down several times during Grey's tenure. In 1905 and again in 1911, for example, Germany clashed with France over control of Morocco and, in each instance, a hesitant Germany backed down. Both Paul Huth (1988:24) and Danilovic (2002:66) code these two crises as instances of immediate deterrence success. But they are also clear instances of a general deterrence failure.

It is impossible to say whether any single equilibrium form conditioned politics among the major European powers from 1905 until the onset of World War I. Assuming that Germany was hesitant until 1914, a period of peace punctuated by acute crises is consistent with the existence of a Weak Crisis Equilibrium.<sup>16</sup> It is also possible, however, that play fluctuated between either a Weak or a Strong Crisis Equilibrium and a Separating or, perhaps, even the Deterrence Equilibrium. In any event, it also seems safe to say that some crisis equilibrium was in play in 1905, in 1911, and again in 1914. For the sake of discussion, let us assume that it was a Weak Crisis Equilibrium. Grey's policy is clearly consistent with this assumption.

---

<sup>16</sup>Under a Strong Crisis Equilibrium, crises would have occurred constantly.



Under a Weak Crisis Equilibrium, a straddle strategy leaves Challenger guessing about Defender's likely response in a confrontation. Indeed, this is its purpose. If the policy is successful, Challenger oftentimes hesitates before provoking a crisis. In equilibrium, however, this will not happen all the time. Every now and then, even a hesitant Challenger may take a chance, as Germany did on at least three occasions at the beginning of the twentieth century. Expecting success, Challenger gambles. Sometimes, expectations are satisfied and the gamble works. Unfortunately for all involved, this was not the case on August 4, 1914, the day the lamps began to dim throughout Europe.

### Summary and Conclusion

The focus of this essay is the deterrence-versus-restraint dilemma in extended deterrence relationships. Germany faced this dilemma in 1879; Great Britain in 1914. Today the United States faces it in its relationship with China and Taiwan and, to a lesser extent, in the Middle East and on the Korean peninsula. In studying this type of dilemma, we have first described and then explored the Tripartite Crisis Game under incomplete information. This game model was developed specifically to capture the mixed motives and contradictory impulses that oftentimes frame extended deterrence relationships.

The Tripartite Crisis Game is strategically intricate. The choices of the players—Challenger, Defender, and Protégé—are dynamically interdependent. Also, with six distinct outcomes, it can take on many forms. To simplify our analysis, we made several nontrivial assumptions. First, we made a number of assumptions about the utilities of the players and explained the theoretical motivations behind them. With the exception of two critical preference relationships, one each for Challenger and Protégé, we also assumed that the players' preferences and utilities were common knowledge. Most specifically, we assumed that Defender is known to prefer *Conflict* to *Protégé Realigns*, the outcome we associate with the breakup of the strategic association between Defender and Protégé. Defenders with this preference are called staunch. The assumption that Defender is known to be staunch has immediate strategic implications. The more likely it is that Protégé is loyal and prefers not to realign, the less likely it is that Defender will support Protégé in a dispute with Challenger, and conversely. Thus, it should be no surprise that most of the strategic action in the Tripartite Crisis Game can be traced to the credibility level of Protégé's threat to realign should it be abandoned by Defender in a crisis.

At the highest levels of Protégé credibility, the status quo is completely stable and deterrence success is certain. Whenever Protégé's threat is highly credible, Defender's support is assured and Protégé always intends to resist Challenger's demands. Anticipating the certainty of conflict, Challenger is deterred. Our finding that extended deterrence is most likely to be successful when Protégé's alignment options are attractive runs counter to Crawford's (2003:1) argument to the contrary.

Deterrence starts to unravel in the Tripartite Crisis Game at lower levels of Protégé's credibility. Below a minimum threshold associated with the existence of the Deterrence Equilibrium, determined Challengers, or Challengers who prefer to fight rather than concede during a crisis, are never deterred. In contrast, depending on the extent to which Protégé's threat to realign is believed, hesitant Challengers, or those Challengers who intend to back down, may sometimes be deterred.

As noted, at middling levels of credibility (under a Separating Equilibrium), even hesitant Challengers are always deterred. But at the lowest levels, a different dynamic takes over. There are two distinct situations that are determined by the credibility of Challenger's, and not Protégé's, threat. One situation—which occurs when Challenger's credibility is low to moderately high—is associated with the

existence of a Weak Crisis Equilibrium. The second—which exists only at the highest levels of Challenger credibility—is associated with the existence of a Strong Crisis Equilibrium.

That Challenger's credibility becomes critical when Protégé's is low is consistent with intuition. Recall that we have assumed that Defender is staunch. Thus, when Protégé's credibility is relatively high, Defender's behavior is largely determined by the high probability that it will lose a valued strategic partner should it abandon Protégé. As a consequence, Defender intends to support Protégé with higher and higher probabilities, *ceteris paribus*, as the credibility of Protégé's threat increases. But when Protégé's credibility is low, the chances of this dénouement diminish and the importance of Challenger's threat takes on added significance. Of course, because there is always some chance that Protégé will defect, Defender never completely forsakes Protégé. Even at the lowest levels of Protégé credibility, Defender intends to support Protégé—sometimes. Such is the mechanism by which Defender attempts to resolve the deterrence-versus-restraint dilemma. This stratagem will never work, however, when Challenger's credibility is at the highest levels and a Strong Crisis Equilibrium is in play. But a straddle strategy may succeed—every now and then—against a hesitant Challenger under a Weak Crisis Equilibrium when Challenger's credibility is low to moderately high.

It is sometimes argued that credibility is the magic ingredient in deterrence (Freedman 1989:96), that threats work best when threats are clearly communicated (Lebow 1981:85), or that uncertainty decreases the chances of successful deterrence (Huth 1988:3). In the context of extended deterrence, however, clearly communicating a threat and committing irrevocably to it is only rational for Defender when Protégé's threat to realign is itself extremely credible. In the absence of this condition, however, a steadfast commitment to hold firm is inconsistent with rationality, all of which helps to explain why half-hearted signals are so often observed in international politics.

## References

- CHRISTENSEN, THOMAS J., AND JACK SNYDER. (1990) Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity. *International Organization* 44:137–168.
- CRAWFORD, TIMOTHY W. (2003) *Pivotal Deterrence: Third-Party Statecraft and the Pursuit of Peace*. Ithaca: Cornell University Press.
- DANILOVIC, VESNA. (2002) *When the Stakes Are High: Deterrence and Conflict among Major Powers*. Ann Arbor: University of Michigan Press.
- ERLANGER, STEVEN. (1996) "Ambiguity" on Taiwan? Will the US Fight a Chinese Attack? *New York Times*, March 12.
- FEARON, JAMES D. (1997) Signaling Foreign Policy Interests. *Journal of Conflict Resolution* 41:68–90.
- FISCHER, FRITZ. (1974) *World Power or Decline: The Controversy Over Germany's Aims in the First World War*. New York: Norton.
- FREEDMAN, LAWRENCE. (1989) *The Evolution of Nuclear Strategy*, 2<sup>nd</sup> edition. New York: St. Martin's.
- HARRIS, J. PAUL. (2003) Great Britain. In *The Origins of World War I*, edited by Richard F. Hamilton and Holger H. Herwig. Cambridge: Cambridge University Press.
- HERWIG, HOLGER H. (2003) Germany. In *The Origins of World War I*, edited by Richard F. Hamilton and Holger H. Herwig. Cambridge: Cambridge University Press.
- HUTH, PAUL K. (1988) *Extended Deterrence and the Prevention of War*. New Haven: Yale University Press.
- KAGAN, DONALD. (1995) *On the Origins of War and the Preservation of Peace*. New York: Doubleday.
- KIESLING, EUGENIA C. (2003) France. In *The Origins of World War I*, edited by Richard F. Hamilton and Holger H. Herwig. Cambridge: Cambridge University Press.
- KUGLER, JACEK, AND WILLIAM DOMKE. (1986) Comparing the Strength of Nations. *Comparative Political Studies* 19:39–69.
- LANGER, W. L. (1950) *European Alliances and Alignments: 1871–1890*, 2<sup>nd</sup> edition. New York: Knopf.
- LEBOW, RICHARD NED. (1981) *Between Peace and War: The Nature of International Crisis*. Baltimore: Johns Hopkins University Press.
- MERCER, JONATHAN. (1996) *Reputation and International Politics*. Ithaca: Cornell University Press.

- QUESTER, GEORGE H. (1989) Some Thoughts on "Deterrence Failures." In *Perspectives on Deterrence*, edited by Paul C. Stern, Robert Axelrod, Robert Jervis, and Roy Radner. New York: Oxford University Press.
- SAZONOV, SERGE. (1928) *The Fateful Years, 1909–1916: The Reminiscences of Serge Sazonov*. New York: Frederick A. Stokes.
- SCHELLING, THOMAS C. (1966) *Arms and Influence*. New Haven: Yale University Press.
- SNYDER, GLENN H. (1984) The Security Dilemma in Alliance Politics. *World Politics* 36:461–495.
- SNYDER, GLENN H. (1997) *Alliance Politics*. Ithaca: Cornell University Press.
- SNYDER, GLENN H., AND PAUL DIESING. (1977) *Conflict Among Nations: Bargaining, Decision Making and System Structure in International Crises*. Princeton: Princeton University Press.
- WALTZ, KENNETH N. (1964) The Stability of the Bipolar World. *Daedalus* 93:881–909.
- WALTZ, KENNETH N. (1990) Nuclear Myths and Political Realities. *American Political Science Review* 84:731–745.
- ZAGARE, FRANK C., AND D.MARC KILGOUR. (2000) *Perfect Deterrence*. Cambridge: Cambridge University Press.
- ZAGARE, FRANK C., AND D.MARC KILGOUR. (2003) Alignment Patterns, Crisis Bargaining, and Extended Deterrence: A Game-Theoretic Analysis. *International Studies Quarterly* 47:587–615.

