Alignment Patterns, Crisis Bargaining, and Extended Deterrence: A Game-Theoretic Analysis

FRANK C. ZAGARE
State University of New York at Buffalo

D. MARC KILGOUR
Wilfrid Laurier University

To explore the impact of alignment patterns in a rudimentary state system, we develop and analyze the Tripartite Crisis Game, a three-person game among Challenger, Defender, and Protégé. This model captures some of the tensions implicit in the “Alliance” and “Adversary” games, two related but theoretically isolated models due to Snyder. Our analysis enables us to delineate and explore the circumstances that give rise to the “deterrence versus restraint” dilemma. It also provides an answer to Fearon’s empirical puzzle: when convincing commitments are possible, why are halfhearted signals sometimes sent?

Our most surprising result concerns the impact of Protégé’s threat on Challenger’s optimal behavior. When Challenger is willing to fight to back up its demand, but is nonetheless only weakly or moderately motivated, Protégé’s threat can dissuade Challenger from initiating a crisis. But when Challenger is willing to fight and stands to gain a great deal, Protégé’s threat may actually prompt Challenger to make a demand. Our analysis uncovers this unexpected pattern of behavior and suggests when it occurs. That Protégé’s threat to realign sometimes bolsters deterrence, and sometimes undermines it, has implications for the selection bias issue in studies of alliance reliability and helps to explain why some alliances are stabilizing while others are associated with crises and war. The nonlinear consequences of Protégé’s commitment seem to us to constitute another “paradox of war.”

Alignment patterns are fundamental to international politics, indeed to all political systems and most political interactions. Our objective is to understand the connection between alignment patterns and war.

In neorealist theory, alignment is considered one of the two principal constraints on state behavior; system structure is the other. Alignments are important because
they “supplement structure” by specifying the relationship of a system’s units (Snyder, 1997:22; see also Waltz, 1979:80).

System structure and alignment patterns interact. Conventional neorealist wisdom holds that bipolar systems (i.e., systems with only two great powers) are characterized by rigid alignment patterns and flexible policy options. Alignment patterns are rigid because realignment options are limited; policy is flexible because disgruntled parties have nowhere to turn. By contrast, multipolar systems, with three or more great powers, are characterized by flexible alignment patterns but constrained policy options. Alignment patterns are fluid because state options are plentiful; still, policy is restricted because discontented partners can more easily realign (Waltz, 1979:169–70).

Alliances register or reflect alignments, but alignments and alliances are distinct. Alliances are generally defined as an agreement among two or more states about contingent military behavior. For instance, the critical clause of the Franco-Russian alliance of 1894 committed France and Russia to each other’s defense should either state be attacked by Germany. By contrast, alignments refer to the “expectations of states about whether they will be supported or opposed by other states in future interactions.” Alignment, then, is a belief or “a state of mind that influences, or may be influenced, by interaction” (Snyder, 1997:6, 22).

An alignment pattern may exist in the absence of a formal system of alliances. Alignments include alliances, but not the other way around. “Alliances are simply one of the behavioral means to create or strengthen alignments” (Snyder, 1997:8). Alignment patterns, therefore, are more general and more basic than alliances.

What is the connection between alignment patterns and war? How do alignment patterns condition the settlement of, and constrain state behavior during, an acute inter-state crisis? Do fluid alignment patterns bolster or undermine extended deterrence relationships?

To answer these and related questions we develop a simple (i.e., a discrete) game model to understand, more precisely, how behavior (i.e., policy) may be constrained by expectations in a system with multiple alignment options. To our knowledge, this aspect of crisis bargaining has not previously been analyzed formally. As we show, alignment patterns can determine the resolution of acute crises and the onset of war.

We emphasize that, in addressing these questions, we are not attempting to explain either why or when alliances are likely to form. Rather, we seek to gain a deeper understanding of the impact of alignment patterns on crisis bargaining in a rudimentary state system. This is why alignment patterns are the givens of our analysis, why they are exogenous to our model.

Unlike most previous studies that assume that the defender alone may fail to honor its commitment, our extended deterrence model allows for perfidy by either the defender or its protégé. It is our belief that the additional complexity implied by the inclusion of a third active player is amply repaid by the new insights provided by the model.

To help place the model in strategic context, consider for now the situation facing German Chancellor Otto von Bismarck when, early in 1879, he attempted to establish the first peacetime alliance in Europe since the days of Napoleon. Looking to consolidate gains achieved in short and decisive wars with Austria-Hungary (1866) and with France (1870), Bismarck sought to isolate France and to convince other European states that Germany was no longer a revisionist power (Langer, 1950:195). The key to this policy was to draw Austria-Hungary (and eventually

---

1 Snyder (1976) was the first to lay out the theoretical implications of system structure for crisis bargaining. See also Snyder and Diesing (1977: ch. 6).

2 Many recent game-theoretic models have focused not on alignment patterns, but on the alliance formation process or other more specialized aspects of alliance behavior, such as the trade-off between autonomy and security. See, inter alia, Morrow (1991), Smith (1995), Sorokin (1994), or Powell (1999).
Russia) into a defensive alliance. Ultimately, Bismarck was successful: the Three Emperors’ League linking Germany, Austria-Hungary, and Russia was revived in 1881 (Kagan, 1995:100–14). In early 1879, however, this dénouement was far from certain. Austria and Russia were serious rivals in the Balkans. Fearing the worst, Bismarck sought to deflect Russian pressures against Austria that, if resisted, might draw Germany into a war with Russia or, if not resisted, might push Austria into an anti-German alliance with Russia or, even worse, France (Langer, 1950:180). Neither outcome would serve Germany’s interests. To help stabilize the status quo, then, Bismarck offered Austria an alliance—on terms that favored Austria. After a short period of negotiations in the fall of 1879, Austria accepted.

But why was an alliance necessary? And why did Bismarck, whose preference was for a general defensive arrangement, agree to an alliance directed against Russia only? We next address these and related questions in the context of a three-player crisis bargaining model we call the Tripartite Crisis Game.

**Game Form**

The Tripartite Crisis Game (Figure 1) is designed to capture the strategic dynamic that conditioned the Austro-German discussions in 1879 (and subsequent renegotiations of the Dual Alliance until 1914). It is not, however, limited to these situations. As we suggest below, the Tripartite Crisis Game applies as well to certain other extended deterrence relationships, such as Russia’s and Serbia’s just prior to World War I (Joll, 1992:125). The players in the model—Challenger, Protégé, and Defender—have distinct roles and divergent strategic interests. We begin by describing the players’ choices and the consequent outcomes. Then we discuss preferences.

In the Tripartite Crisis Game, Challenger begins play (at node 1), deciding whether to demand a concession from Protégé. If no demand is made, the game ends and the Status Quo (SQ) obtains. But if a demand is made, a crisis occurs. In this contingency, Protégé decides (at node 2) whether to concede or to hold firm. Concession leads to outcome Challenger Wins (CW), holding firm to a critical choice for Defender (at node 3), whether to support Protégé, Defender’s choice forces either Protégé or Challenger to make a subsequent decision. If Defender stands by Protégé, Challenger either presses on or backs down, leading to Conflict (C) or Challenger Concedes (CC), respectively. If Defender abandons Protégé, Protégé is forced to reach an accommodation with Challenger. In the wake of its diplomatic defeat, however, it may also reconsider its alignment policy, choosing whether to maintain its now tenuous relationship with Defender (as Austria did after it was forsaken by Germany in the 1913 Balkan crisis), or to find a more reliable partner by realigning, perhaps even with Challenger. The outcome is Protégé Loses (PL) in the first instance, and Protégé Realigns (PR) in the second. In

---

5 See Morrow (2000) for a general discussion of this question.  
6 We emphasize again that our model does not attempt to offer an explanation of the alliance formation process per se. Rather, our model suggests only why Bismarck wanted Austria to align itself more closely with Germany. In 1879, an explicit military alliance with Austria was the most obvious and, likely, the most efficacious, way of drawing Austria deeper into Germany’s orbit. After all, the initiative for the defense pact had come from Austria. But this does not mean that a formal alliance was the only mechanism available for strengthening the relationship of the two German-speaking states. For example, at the conclusion of the Austro-Prussian war in 1866, Bismarck wanted to leave open the possibility of a subsequent rapprochement. In consequence, he offered Austria lenient peace terms.  
7 In 1879, he could have offered Austria trade or other concessions rather than a lopsided defense agreement.  
8 The nature of the demand is exogenous to our model. In other words, we assume that the stakes are fixed and that Challenger is unable to adjust the size of its demand in order to manipulate play. For an insightful analysis of extended deterrence in which Challenger can select the level of its demand strategically, see Werner (2000).  
9 In effect, we assume that Challenger is stronger than Protégé, so that a confrontation would be unfavorable to an unprotected Protégé, and that both Challenger and Protégé know this.
either case, the concession made to Challenger is the same. But when Protégé Realigns, Defender loses an important strategic partner. Fear of this event drove much of German foreign policy in 1879 and afterward.7

To be sure, our model is a simplification of the crisis bargaining process in a multipolar system or subsystem. The most obvious simplification concerns the number of players: we assume just three. Larger and more complex games may be the rule, and not the exception, in international politics. In Bismarck’s time, for instance, five great powers dominated the European state system. As well, we have dichotomized the choices the players make at the various decision nodes. The Tripartite Crisis Game could be formulated to include additional players, more nuanced options, or more differentiated outcomes.

It is not immediately clear, however, that the political dynamic we are exploring would be made more transparent by extending the Tripartite Crisis Game to additional players or more refined choices. Rather than unduly complicating the model, then, we assume that the consequences of these and related refinements are captured by the players’ utility assessments. For instance, we assume that each player’s utility for Conflict takes into account the possibility that a war could expand.

7 During the 1908–1909 Bosnian crisis, for instance, both German Chancellor Bernhard Bülow and Friedrich von Holstein, a key foreign policy adviser at the Wilhelmstrasse, feared that Austria would tilt toward England and France if it were not supported in its dispute with Russia. In consequence, “Bülow and Holstein felt they had no choice but to support Austria in its bid to annex Bosnia and Herzegovina” (Mercer, 1996:125; see also Kagan, 1995: 162–63).
We also assume, *inter alia*, that the players’ utilities for *Prote´ge´ Realigns* reflect their best assessment of this outcome’s wider strategic implications.

**Outcomes and Preferences**

With three players and six possible outcomes, the Tripartite Crisis Game model could have many variants, reflecting different assumptions about players’ preferences. Not all of these variants, however, are consistent with the tacit bargaining process we are modeling. Others are inconsistent with our focus on crises in which alignment considerations are of critical importance. And still others are of little strategic or theoretical interest.

What follows is our justification for the specific assumptions we make about the players’ preferences (see Table 1). We do not argue that they are uniquely appropriate but, rather, defend them on both theoretical and pragmatic grounds. Theoretically, our assumptions (especially about Defenders’ preferences) reflect the standard neorealist position about the critical impact of alignment patterns on crisis bargaining. Pragmatically, our focus on an important special case makes the analysis tractable. In subsequent research we intend to study more general versions of the Tripartite Crisis Game.

The columns of Table 1 show the players’ preferences over the six outcomes, from best to worst. For example, we assume that Challenger most prefers *Challenger Wins*, next-most prefers *Prote´ge´ Realigns*, and so on. No fixed preference is assumed for outcomes contained in the same cell of Table 1. Thus, in our model, Challenger could prefer *Conflict* to *Challenger Concedes* or the reverse. The players’ relative preferences for these paired outcomes are crucial explanatory variables in the model.

**a. Challenger’s Preferences**

Consider first Challenger’s preferences. In the Tripartite Crisis Game, Challenger’s primary objective, as Bismarck saw Russia’s in 1879, is to obtain concessions from *Prote´ge´*. Three outcomes of the Tripartite Crisis Game offer Challenger a clear improvement of the *Status Quo*. We take Challenger’s preference for these outcomes over the others as a given. Of these, we assume: *Challenger Wins* $\succ_{Ch} *Prote´ge´ Realigns* $\succ_{Ch} *Prote´ge´ Loses*.

*Challenger Wins* brings immediate gratification and is relatively costless. *Prote´ge´ Realigns* humiliates Defender. *Prote´ge´ Loses* does not. Challenger’s postulated preference for *Prote´ge´ Realigns* over *Prote´ge´ Loses* means that Prote´ge´’s alignment policy matters, not only to Defender (see below), but to Challenger as well.8 As demonstrated shortly, this seemingly innocuous assumption has important strategic implications.

Of the remaining three outcomes, we assume that Challenger prefers the *Status Quo* to both *Conflict* and *Challenger Concedes*. According to this assumption, Defender’s threat to support Prote´ge´ is *capable*.9 As we show elsewhere (Zagare, 1987; Zagare and Kilgour, 2000), threat capability is a necessary condition for

---

8 The assumption that Challenger prefers *Prote´ge´ Realigns* to *Prote´ge´ Loses* is a simplification that does not hold all the time. For example, neither Britain nor France were interested in pursuing what Snyder (1997: 337–38) calls a “divide and rule” policy in the years leading up to World War I. Both Sir Edward Grey, Britain’s Foreign Secretary, and President Raymond Poincaré of France believed that separating Austria-Hungary from Germany would likely provoke Germany. By contrast, in 1905 during the first Moroccan crisis, Germany hoped to break up the Entente Cordiale by pressuring France for concessions in Morocco (Massie, 1991:363). Similarly, in July 1914, German Chancellor Theobald von Bethmann-Hollweg believed that a localized war in the Balkans would shatter Russia’s alliance with France. In consequence, he pushed Austria to move against Serbia (Levy, 1990/91:160). The conclusions of our model apply, therefore, to Germany in 1905 and 1914, but not to Britain and France immediately prior to the July crisis.

9 Following Schelling (1966), we define capability as the ability to hurt. Defender’s threat to support Prote´ge´, then, is capable if Challenger prefers that the threat not be carried out. For a more extensive discussion of this concept, see Zagare and Kilgour (2000: ch. 3).
deterrence success; without it, the Status Quo never survives in rational play and (extended) deterrence always fails.\textsuperscript{10}

Challenger’s preference between its two lowest ranked outcomes, Challenger Concedes and Conflict, determines its type. Challengers who prefer Conflict to Challenger Concedes are called determined. Challengers with the opposite preference, unwilling to fight, are termed hesitant.

b. \textit{Prote`ge´}’s Preferences

Consider now \textit{Prote`ge´}’s preferences. We assume that \textit{Prote`ge´}’s goal is to avoid acceding to Challenger. Among the three outcomes that do not involve concession, we assume Status Quo $\succ_{pro}$ Challenger Concedes $\succ_{pro}$ Conflict. Status Quo involves no loss or evident cost, whereas when Challenger Concedes\textit{Prote`ge´} incurs an obligation to Defender.\textsuperscript{11} With Conflict, on the other hand, there are additional costs as well as the risk of an unfavorable outcome.

Of the three outcomes that involve an outright loss, \textit{Prote`ge´} prefers Challenger Wins on the grounds that an immediate concession is less costly than a later concession made under duress. \textit{Prote`ge´}’s relative preference between the remaining two outcomes determines its type. A \textit{Prote`ge´} that least prefers to realign (i.e., prefers \textit{Prote`ge´} Loses to \textit{Prote`ge´} Realigns) is loyal;\textsuperscript{12} a \textit{Prote`ge´} with the opposite preference is disloyal.

c. Defender’s Preferences

Finally, consider Defender’s situation. Defender is stuck between a rock and a hard place. \textit{Ceteris paribus}, it would like to avoid Conflict and preserve its bond with \textit{Prote`ge´}. The rub, of course, is that Defender might have to threaten to fight to save the relationship. This conundrum is especially acute in a multipolar system where the costs of realignment are generally higher than in a bipolar system. Hence the additional constraints on policy under multipolarity.\textsuperscript{13}

\begin{table}
\centering
\begin{tabular}{lll}
\hline
Challenger: & \textit{Prote`ge´}: & Defender: \\
\hline
Challenger Wins & Status Quo & Status Quo \\
Prote`ge´ Realigns & Challenger Concedes & Challenger Concedes \\
Prote`ge´ Loses & Conflict & Challenger Wins \\
Status Quo & Challenger Wins & Prote`ge´ Loses \\
Conflict or Challenger Concedes & Prote`ge´ Loses or Prote`ge´ Realigns & Conflict or Prote`ge´ Realigns \\
\hline
\end{tabular}
\caption{Basic Preference Rankings for Tripartite Crisis Game}
\end{table}

\textsuperscript{10} Our model, therefore, applies only when it is possible to deter Challenger, either because Challenger’s dissatisfaction with the status quo is relatively low or because the costs of conflict are prohibitively high.

\textsuperscript{11} There may be circumstances when a \textit{prote`ge´} or a defender (see below) would derive considerable utility from facing down a challenger. Our model does not apply to these situations.

\textsuperscript{12} Loyalty, however, may have very little to do with the continuation of a strategic partnership, particularly after abandonment.

\textsuperscript{13} In a multipolar system such as Europe in the late 19th century, the defection of a single partner was potentially fatal to the viability of a blocking coalition, which is why Britain, in its “splendid isolation,” was able to play the role of the “balancer.” By contrast, in a strict bipolar system in which each pole is dominated by a single large state, the defection of any single partner would not necessarily be critical to the coalition’s survival.
Of course, neither cost is incurred when the *Status Quo* prevails. The same is true when *Challenger Concedes*. But since we take deterrence to be Defender’s primary motivation, we assume: *Status Quo* ><sub>Def</sub> *Challenger Concedes*.

We also assume: *Challenger Concedes* ><sub>Def</sub> *Challenger Wins* ><sub>Def</sub> *Protégé Loses* ><sub>Def</sub> *Protégé Realigns*. A common negative shared by the latter three outcomes is indirect: the concession that Protégé is pressured to make to Challenger. At *Challenger Wins*, however, Defender’s perfidiousness is not transparent. At *Protégé Loses* Defender’s type is apparent, but at least its relationship with Protégé, albeit weakened, is maintained. At *Protégé Realigns*, Protégé defects and Defender incurs a significant strategic cost.

Defender’s evaluation of *Conflict* relative to the other outcomes remains to be discussed. There are a number of reasonable possibilities. Not all are interesting or theoretically important. We focus on those that are.

When Defender is *always* willing to support Protégé, i.e., when *Conflict* ><sub>Def</sub> *Protégé Loses*, there is no tension between Defender’s goal of avoiding conflict and retaining Protégé’s allegiance. To accentuate this tension, we assume that *Protégé Loses* ><sub>Def</sub> *Conflict*. We make no fixed assumption, however, about Defender’s relative ranking of *Conflict* and *Protégé Realigns*. Defenders preferring *Conflict* to *Protégé Realigns* are called *staunch*; those with the opposite preference are called *perfidious*.

To summarize: the Tripartite Crisis Game is defined by the game form of Figure 1, the preference restrictions listed in Table 1, and the type designations given in Table 2. Notice that the Tripartite Crisis Game is a true three-person game in that all three players face dynamically interactive choices. No player can rationally ignore another’s actions. To be sure, Protégé’s decision at node 4a, and Challenger’s decision at node 4b, can be made without reference to the preferences and the choices of other players. But Protégé’s best choice at node 2 depends on what it expects of Defender at node 3, which in turn should depend on Defender’s estimates of *both* Protégé’s choice at node 4a and Challenger’s at node 4b. Similarly, Challenger’s best decision at node 1, and Defender’s at node 3, depend on each player’s expectations of both opponents.

By so structuring the Tripartite Crisis Game, we combine, in a single formal structure, elements of Snyder’s (1997: 37) “Alliance” and “Adversary” games. This theoretical integration is important because it permits a more thorough understanding of the “deterrence-vs.-restraint dilemma” (Snyder and Diesing, 1977:438). To deter Challenger, Defender is motivated to pledge strong support for Protégé. But stronger support also increases the probability that Protégé will hold firm if a demand is received, risking a conflict that Defender would prefer to avoid.

Protégé also faces a strategic dilemma in the Tripartite Crisis Game: the more credible Protégé’s threat to realign, the more likely the status quo to survive, because of the greater probability that Defender will support Protégé at node 3 and that Protégé, expecting this support, will hold firm against a demand. Of course,  

---

14 When Defender prefers *Conflict* to *Protégé Loses*, it always chooses to support Protégé at node 3. In consequence, Protégé never faces a choice at node 4a.
15 Defender’s preference for *Challenger Wins* and *Protégé Loses* over *Conflict* implies that Defender is not heavily invested in the issues at stake. One is reminded of Bismarck’s famous remark that all of the Balkans was “not worth the bones of a single Pomeranian grenadier.” Extended deterrence relationships with intrinsically valuable stakes have a fundamentally different strategic dynamic than the one modeled herein. For an innovative investigation of the role of high stakes in extended deterrence relationships, see Danilovic (2002).
16 We leave open precisely how Defender may signal its support. Depending on circumstances, Defender could, *inter alia*, back Protégé at an international conference, as Britain did with France during the Algeciras Conference in 1905 that marked the end of the first Moroccan crisis; it could mobilize its army or put its navy on alert in a way that Challenger would be forced to choose between war and peace; or it could make a visible and very public commitment to defend Protégé. For an innovative listing of “commitment tactics,” see Snyder (1972).
disloyalty is a double-edged sword and the prospect of dislodging Protégé from Defender’s orbit may actually prompt Challenger to instigate a crisis.

Analysis

In this section we offer a general discussion and summary of the strategic properties of the Tripartite Crisis Game. A formal treatment is given in the Appendix.

In what follows, we assume that, to the players of the Tripartite Crisis Game, all relevant information is common knowledge, except that players may be uncertain about the others’ types. (Players, of course, always know their own type.) Specifically, we assume that the players are fully informed about the game defined by (1) the rules of play, as reflected in the game tree of Figure 1, and (2) the preference orderings, as given in Table 1. As discussed below, we further simplify the problem by taking Challenger’s type to be common knowledge.

When information is complete and all player types are common knowledge, only two different outcomes are possible at a subgame-perfect equilibrium. The Status Quo obtains whenever Challenger is hesitant, or when Challenger is determined, Defender is staunch, and Protégé is disloyal. Otherwise, the outcome is Challenger Wins. (The outcomes associated with the subgame perfect equilibria of the Tripartite Crisis Game under complete information are summarized in Table 3.)

It is important to note that for the Status Quo to survive rational play against a determined Challenger, Protégé must be disloyal. In other words, given complete information, Protégé’s preference for Protégé Realigns over Protégé Loses is a necessary condition for extended deterrence success. As we shall show, however, when information is incomplete, a high likelihood that Protégé is disloyal is neither necessary nor sufficient for the survival of the Status Quo. In fact, there are conditions under which a higher probability that Protégé is disloyal actually undermines the stability of the Status Quo.

For the moment, however, consider the Tripartite Crisis Game when Challenger is known to be hesitant, but information about the other players’ types is incomplete. In this case, the Status Quo remains the only outcome that can be supported at any form of strategic equilibrium, as shown in the Appendix. To explain briefly, note that a hesitant Challenger strictly prefers Challenger Concedes to Conflict, so it always backs down at node 4b. Anticipating this choice, both types of Defender always support Protégé at node 3. (Defender strictly prefers Challenger Concedes to either of the outcomes that could result if it chooses not to support Protégé.) For similar reasons, Protégé always holds firm at node 2. In consequence, Challenger chooses not to issue a demand at node 1 knowing that, if it did, Protégé would hold firm, Defender would support Protégé, and that it (Challenger) would be forced to back down. Thus, whenever Challenger is known to be hesitant, the only outcome consistent with equilibrium is Status Quo.

Now consider the situation when Challenger is known to be determined. This is clearly the more interesting case, as only a determined Challenger can rationally
choose to initiate *Conflict* by pressing on at node 4b. And without the very real possibility of *Conflict*, the Tripartite Crisis Game loses much of its strategic and theoretical import.

The study of the complete information version of the Tripartite Crisis Game has shown that when Challenger is determined, at least two different outcomes are possible (see Table 3). The game is obviously more complicated when information about types is incomplete. To avoid some of these complications, however, we restrict our analysis of the incomplete information version of the Tripartite Crisis Game to the special case in which only Defender’s and Protegé’s types are unknown. In other words, we shall continue to assume that Challenger’s type is common knowledge.

We make this assumption fully cognizant that the absolute number of equilibria we uncover may be reduced. But, based on previous research, we conjecture that our special case analysis will result in no serious information loss (Zagare and Kilgour, 1998, 2000). In other words, we anticipate that the equilibrium structure of the Tripartite Crisis Game will be as rich but simpler—less subject to minor but complex exceptions—than when Challenger’s type is taken as private information. In consequence, our conclusions can be sharper, more focused.

To determine the range of behavioral possibilities when information about only Challenger’s type is common knowledge, we next outline a procedure for identifying the perfect Bayesian equilibria of the Tripartite Crisis Game. Our summary discussion is quite general. Leaving technicalities aside, we provide only their broad outline. All formal derivations and many technical details are confined to the Appendix.

A few specific assumptions are required to express the results. Assume now that Defender is staunch with probability $p_{Def}$ (where $0 < p_{Def} < 1$), that Protegé is disloyal with probability $p_{Pro}$ (where $0 < p_{Pro} < 1$), and that the values of $p_{Def}$ and $p_{Pro}$ are known to all the players.17 These assumptions allow us to model the consequences of the players’ uncertainty about Defender’s and Protegé’s types. We continue to assume that the realized value of $p_{Def}$ ($p_{Pro}$) is known to Defender (Protegé), that is, that each player knows its own type.

Observe that when Challenger is known to be determined, the outcome of the Tripartite Crisis Game will be *Conflict* should the play of the game reach node 4b. Knowing this enables us to simplify the game form. Figure 2 illustrates the reduced game, highlighting a proper subgame we call the Protegé–Defender subgame.18 To solve the Tripartite Crisis Game, we first solve this subgame.19 Then we consider Challenger’s node 1 choice in the context of the subgame solutions.

---

17 The combination of these belief variables defines a specific alignment pattern.
18 “Crisis subgame” would be equally apropos.
19 Although we do not focus on its strategic dynamic, the Protegé–Defender subgame is interesting in its own right. It models a situation of *extended immediate deterrence* in which the status quo has already broken down. By contrast, the Tripartite Crisis Game models a situation of *extended general deterrence*. For the distinction see Huth (1988: ch. 1). For related models of extended deterrence, see Smith (1998) and Werner (2000).
The standard solution concept of a game with incomplete information is perfect Bayesian equilibrium. A perfect Bayesian equilibrium consists of a plan of action (i.e., a strategy) for each player, plus the player’s beliefs about (i.e., subjective probabilities over) 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 (i.e., according to Bayes’s Rule) given the actions it observes during the play of the game. A perfect Bayesian equilibrium, therefore, specifies an action choice for every type of every player at every decision node or information set belonging to the player; it must also indicate how each player updates its beliefs about other players’ types.

Note that in the Protégé–Defender subgame, Protégé’s choice at node 4a is strictly determined by its type. A disloyal Protégé always realigns; a loyal Protégé never does. Note as well that before Defender can make a choice at node 3, it will observe Protégé’s prior action choice at node 2. Defender can use this new information to reassess Protégé’s type (and likely choice) at node 4a. Since Defender’s choice at node 3 depends, in part, on its assessment of Protégé’s type, the additional information it obtains is useful. By contrast, any observations that Protégé makes before choosing at node 4a will be beside the point. As noted, Protégé’s choice at this decision node is strictly determined by its type.
Given these considerations, it follows that a perfect Bayesian equilibrium of the Protege–Defender subgame will consist of a five-tuple of probabilities $[y_D, y_L, z_S, z_P, q]$ where:

- $y_D =$ the probability that a disloyal Protege will choose to hold firm at node 2
- $y_L =$ the probability that a loyal Protege will choose to hold firm at node 2
- $z_S =$ the probability that a staunch Defender will choose to support Protege at node 3
- $z_P =$ the probability that a perfidious Defender will choose to support Protege at node 3
- $q =$ Defender’s updated probability that Protege is disloyal, given that Protege holds firm at node 2.

The first four probabilities are strategic variables describing Protege’s and Defender’s choices, contingent on type. The fifth probability is the a posteriori probability, updated by Defender once Protege’s choice to hold firm at node 2 has been observed. Equilibrium values for the belief variable, $q$, are reported in the Appendix, and are discussed below only when relevant.

There are four nontransitional perfect Bayesian equilibria in the Protege–Defender subgame: Settlement, Separating, Hold Firm, and Bluff. Table 4 summarizes their technical properties. As we discuss below, the behavioral patterns associated with these perfect Bayesian equilibria are crucial determinants of Challenger’s node 1 choice and, consequently, the larger dynamic of the Tripartite Crisis Game.

<table>
<thead>
<tr>
<th>Equilibrium</th>
<th>Strategic and Belief Variables</th>
<th>Existence Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settlement I</td>
<td>$y_D = 0$, $y_L = 0$, $z_S = 1$, $z_P = 0$</td>
<td>$p_{Df} &lt; e_2$</td>
</tr>
<tr>
<td>Settlement II</td>
<td>$y_D = 0$, $y_L = 0$, $z_S = 0$, $z_P = 1$</td>
<td>$d_1 &lt; p_{Df}$</td>
</tr>
<tr>
<td>Separating</td>
<td>$y_D = 1$, $y_L = 0$, $z_S = 0$, $z_P = 1$</td>
<td>$e_2 &lt; p_{Df} &lt; e_1$</td>
</tr>
<tr>
<td>Hold Firm</td>
<td>$y_D = 1$, $y_L = 1$, $z_S = 0$, $z_P = 0$</td>
<td>$p_{Df} &gt; e_1$, $p_{P_B} &gt; d_1$</td>
</tr>
<tr>
<td>Bluff</td>
<td>$y_D = 1$, $y_L = 0$, $z_S = 0$, $z_P = 1$</td>
<td>$p_{Df} &gt; e_1$, $p_{P_B} &lt; d_1$</td>
</tr>
</tbody>
</table>

Key: “$\bullet$” = fixed value between 0 and 1; “--” = value not fixed although some restrictions apply.

20 We ignore one transitional equilibrium, which is an equilibrium that exists only when the parameters of a model satisfy a specific functional relationship (i.e., an equation). The justification for ignoring transitional equilibria is that, however the parameter values are obtained, they are very unlikely to satisfy any specific equation.

21 See Appendix for the details, a brief discussion, and references.
subgame under a Separating equilibrium, then, is **Challenger Wins** when Protégé is loyal; **Conflict** when Protégé is disloyal and Defender is staunch; and **Protégé Realigns** when Protégé is disloyal and Defender is perfidious.

Under the **Hold Firm** perfect Bayesian equilibrium, Protégé always holds firm (i.e., \( y_D = y_L = 1 \)). Hence its name. Staunch Defenders always support Protégé, perfidious Defenders always withhold support (i.e., \( z_S = 1, z_P = 0 \)). Thus, the outcome under a Hold Firm equilibrium is **Conflict** if Defender is staunch. When Defender is perfidious the outcome depends on Protégé’s type: **Protégé Loses** when Protégé is loyal; **Protégé Realigns** when Protégé is disloyal.

The final, and most interesting, perfect Bayesian equilibrium of the Protégé–Defender subgame is the **Bluff** equilibrium. At a Bluff equilibrium, disloyal Protégés always hold firm; loyal Protégés sometimes do the same (i.e., \( y_D = 1; 0 < y_L < 1 \)). This means that a loyal Protégé’s node 2 choice is a bluff. Hoping to elicit Defender’s support by sending a (false) signal, it holds firm, even though it has no intention of realigning. When Defender is staunch, the bluff works, sometimes, though a perfidious Defender never supports Protégé. This is not surprising. Perfidious Defenders never support Protégé under any perfect Bayesian equilibrium of the Protégé–Defender subgame (i.e., \( z_P \) always equals 0).

The action choices of a staunch Defender under a Bluff equilibrium are difficult to categorize. A staunch Defender sometimes, but not always, supports Protégé (i.e., \( 0 < z_S < 1 \)). This behavioral tendency is in fact surprising. It runs counter to Defender’s equilibrium behavior under all other perfect Bayesian equilibria of the Protégé–Defender subgame (except for the Form II Settlement equilibrium that we ignore). Protégé would probably view a staunch Defender’s support by sending a (false) signal, it holds firm, even though it has no intention of realigning. When Defender is staunch, the bluff works, sometimes, though a perfidious Defender never supports Protégé. This is not surprising. Perfidious Defenders never support Protégé under any perfect Bayesian equilibrium of the Protégé–Defender subgame (i.e., \( z_P \) always equals 0).

The action choices of a staunch Defender under a Bluff equilibrium are difficult to categorize. A staunch Defender sometimes, but not always, supports Protégé (i.e., \( 0 < z_S < 1 \)). This behavioral tendency is in fact surprising. It runs counter to Defender’s equilibrium behavior under all other perfect Bayesian equilibria of the Protégé–Defender subgame (except for the Form II Settlement equilibrium that we ignore). Protégé would probably view a staunch Defender’s willingness to withhold support as a veiled threat. Challenger might interpret it as a promise—of sorts. For Defender it is a way to resolve the “deterrence-vs.-restraint” dilemma. By hedging its bet, it achieves an optimal mix of both of these important security commodities.

Defender’s equilibrium behavior under a Bluff equilibrium provides insight into an empirical puzzle raised by Fearon (1997). In Fearon’s game model of costly foreign policy signaling, a player has two strategies to communicate its interests: it can signal that its “hands are tied” or that its “costs are sunk.” Players never rationally bluff with either signal, leading Fearon (1997: 71) to wonder “why we sometimes observe halfhearted signals when convincing ones are possible?”

But this behavior arises naturally in the Tripartite Crisis Game. Under the Bluff equilibrium in the Protégé–Defender subgame, a staunch Defender’s strategy corresponds to a signal that is strong enough to deter all but the most determined Challengers, yet not so strong that loyal Protégés become intransigent and provoke crises. Defender’s rational objective is balance: too strong a commitment enflames Protégé whereas too weak a commitment incites Challenger.

The signal is fuzzy, then, because it has two different audiences. By deterring Challenger, the signal minimizes the risk of conflict and helps to stabilize the status quo; by restraining Protégé, it reduces the risk of chain ganging (Christensen and Snyder, 1990) and protects Defender’s alignment relationship with Protégé. What is surprising about this mixed message is that it is delivered by a staunch Defender, one that would prefer to fight to save its relationship with Protégé.

The “intentionally vague commitment” made by the United States in the Taiwan Relations Act of 1979 is a good example (Erlanger, 1996). To restrain China, the U.S. signaled its intention to back Taiwan. But to restrain Taiwan, the U.S. signaled that its support was not unconditional. In sum, the signal was halfhearted.22

With the exception of the implausible Form II Settlement equilibrium, the perfect Bayesian equilibria of the Protégé–Defender subgame do not co-exist. To see this, consider now Figure 3. Along the horizontal and vertical axes of this figure

---

22 Benson and Niou (2001) make the case (game-theoretically) for the United States' policy of “strategic ambiguity” in the Taiwan Strait security situation.
are graphed, respectively, the two belief variables, \( p_{\text{Def}} \) and \( p_{\text{Pro}} \). Each possible combination of these variables defines a particular alignment pattern. Several constants, such as \( d_1 \) and \( e_2 \), are also indicated along the axes of Figure 3. These constants, which are defined and discussed in detail in the Appendix, are convenient thresholds for categorizing and interpreting the perfect Bayesian equilibria of the Protégé–Defender subgame.

The two belief variables can be interpreted as measures of Protégé’s and Defender’s threat credibility (Zagare and Kilgour, 2000): the higher \( p_{\text{Pro}} \), the more likely/credible Protégé’s threat to realign; and the higher \( p_{\text{Def}} \), the more likely/credible Defender’s threat/promise to support Protégé. As well, since the belief variables indicate Defender’s and Protégé’s expectations of support or nonsupport as the crisis unfolds, the unrealized values of these variables capture the particular alignment pattern that constrains behavior in the Tripartite Crisis Game.

As Figure 3 shows, and as one might very well expect, Settlement equilibria exist precisely when Defender’s credibility is lowest (i.e., when \( p_{\text{Def}} < e_2 \)). Intuitively, when Defender is unlikely to be staunch, Protégé will (rationally) choose to settle with Challenger if and when it is faced with a decision at node 2. Notice that when Defender is very likely perfidious, Protégé’s threat to realign is of no strategic consequence. In other words, the conditions associated with the existence of a
Settlement equilibrium do not depend on Protégé’s type. The reason is straightforward: given Protégé’s preferences, node 4a is never reached in the play of the game. Thus, Protégé’s threat to realign never factors into either Challenger’s or Defender’s strategy choice.

Protégé’s credibility determines which of two perfect Bayesian equilibria will exist when Defender’s credibility is high (i.e., when \( p_{Def} > \epsilon_1 \)). When Protégé’s threat to realign is highly credible (i.e., when \( p_{Pro} > d_1 \)), a Hold Firm equilibrium exists. But when it is low (i.e., when \( p_{Pro} < d_1 \)), a Bluff equilibrium comes into play. Again, this result is consistent with intuition. Staunch Defenders should be more willing to support Protégé as the credibility of Protégé’s realignment threat increases. When Protégé’s threat to realign is relatively credible (as it is under a Hold Firm equilibrium), a staunch Defender will always choose to support Protégé. After all, a staunch Defender, by definition, prefers Conflict to Protégé Realigns. But as the credibility of Protégé’s threat erodes (as it does under a Bluff equilibrium), Defender becomes progressively more likely to withhold support.

A Separating equilibrium occurs between the two extremes of Defender’s credibility (i.e., when \( e_2 < p_{Def} < \epsilon_1 \)). As indicated, under a Separating equilibrium, play in the Protégé–Defender subgame is determined strictly by the players’ types. Only disloyal Protégés hold firm; and only staunch Defenders offer support. As with play under a Settlement equilibrium, Protégé’s initial credibility (\( p_{Pro} \)) is of little moment. The reason is that under a Separating equilibrium Protégé signals its type by its node 2 choice. This means that when Defender makes its choice at node 3, it will know for sure what choice Protégé will make at node 4a should it choose not to support Protégé. In consequence, Defender’s node 2 choice is dictated by its type (i.e., by its preferences).

b. The Tripartite Crisis Game

Having laid out the contours of the Protégé–Defender subgame, we now return to an analysis of the Tripartite Crisis Game. In our analysis, we focus on Challenger’s node 1 choice. Recall that at node 1 Challenger either makes no demand, in which case the outcome is Status Quo, or makes a demand, in which case the outcome is that of the Protégé–Defender subgame. The outcome of the subgame, in turn, depends on which perfect Bayesian equilibrium will come into play, that is, on the two belief variables \( p_{Pro} \) and \( p_{Def} \). Clearly, Challenger will choose to make a demand if and only if its utility for the outcome of the Protégé–Defender subgame exceeds its utility for the Status Quo.

As explained in the Appendix, the key to Challenger’s decision is the probability

\[
\epsilon_1 = \frac{c_{PR} - c_{SQ}}{c_{PR} - c_{C+}}
\]

relative to the threshold probabilities that separate the various equilibrium regions of the Protégé–Defender subgame. Three parameters establish the value of \( \epsilon_1 \): Challenger’s utility for Protégé Realigns (\( c_{PR} \)); Challenger’s utility for the Status Quo (\( c_{SQ} \)); and (a determined) Challenger’s utility for Conflict (\( c_{C+} \)).

The probability \( \epsilon_1 \) is the threshold at which Challenger’s utility for a lottery between Protégé Realigns and Conflict equals its utility for the Status Quo. If the probability of Conflict exceeds \( \epsilon_1 \), Challenger prefers the Status Quo. If it is less, Challenger prefers the lottery.

The threshold probability \( \epsilon_1 \) marks the point beyond which Challenger is unwilling to risk Conflict, given that Protégé is known to be disloyal (i.e., \( p_{Pro} = 1 \)).

---

23 Disloyal protégés always hold firm; loyal protégés always concede. More technically, under a Separating equilibrium, \( q \) (Defender’s updated probability that Protégé is disloyal given that Protégé holds firm at node 2) equals 1.
Following Ellsberg (1959), we refer to $c_1$ as Challenger’s critical risk.\textsuperscript{24} As will be seen, the likelihood of a crisis increases as $c_1$ increases, so we interpret this quantity as a measure of Challenger’s propensity to foment a crisis.

Notice that Challenger’s critical risk increases as its utility for Protégé Realigns increases, or as its utility for Conflict increases. This suggests that a Challenger who places a relatively high value on breaking up the relationship between Protégé and Defender may be more likely to precipitate a crisis than one who does not. And, as one might expect, a Challenger averse to Conflict may be less likely to issue a demand than one who is not.

For expository purposes it will be useful to think of Challenger’s critical risk as being at one of three levels: low, medium, and high. Low values are those below the threshold probability $e_2$ that serves as the boundary between the regions of Settlement and Separating equilibria (see Figure 3). High values are those above the threshold probability $e_1$ that sets the region of Separating equilibria apart from the regions of Hold Firm and Bluff equilibria. Medium values are in-between (i.e., $e_2 < c_1 < e_1$), falling squarely in the region of Separating equilibria.

Figure 4 provides a graphical summary of Challenger’s node 1 decision for each theoretically meaningful level of Challenger’s critical risk: low (Case A), medium (Case B), and high (Cases C\textsubscript{1} and C\textsubscript{2}).\textsuperscript{25} In each case an area of gray shading is superimposed on the equilibrium structure of the Protégé–Defender subgame as given in Figure 3. The shading delimits the conditions under which Challenger demands an adjustment of the Status Quo at node 1. (No shading means that Challenger prefers to accept the Status Quo.) Table 5 provides a verbal summary.

As is apparent from an examination of Figure 4, Challenger becomes progressively more likely to issue a demand as $c_1$ increases. The shaded area, where a demand is issued, grows larger and larger as one moves rightward and downward from Case A through Case B to Cases C\textsubscript{1} and C\textsubscript{2}.

Consider now Challenger’s node 1 choice given that Defender’s credibility is low (i.e., when $p_{Def} > e_2$). This contingency is just about Protégé’s worst nightmare. Recall that when Defender is very likely perfidious, a Settlement equilibrium uniquely exists. Under any Settlement equilibrium, Protégé always concedes at node 2, and the outcome is always Challenger Wins. Since our assumption is that a determined Challenger prefers Challenger Wins to Status Quo, Challenger will always issue a demand at node 1 in order to secure a more preferred outcome, even when its critical risk is low. Thus, in each case depicted in Figure 4, the region indicating the existence of a Settlement equilibrium is shaded.

What happens at intermediate levels of Defender’s resolve (i.e., where $e_2 < p_{Def} < e_1$) and a Separating equilibrium governs play of the Protégé–Defender subgame? As Figure 4 suggests, Challenger’s critical risk interacts with the credibility of both Protégé’s and Defender’s threats to determine Challenger’s node 1 choice. When Challenger’s critical risk is low (Case A), Challenger rationally accepts the Status Quo at node 1, but only when the credibilities of Protégé’s and Defender’s threats (i.e., the values of $p_{Prot}$ and $p_{Def}$) are relatively high. Much the same can be said when Challenger is moderately motivated (Case B), except that proportionately higher levels of Protégé and Defender credibility are required to deter Challenger. Finally, when the magnitude of Challenger’s critical risk is high (i.e., exceeds the threshold probability $e_1$), Challenger always issues a demand and is never deterred (Cases C\textsubscript{1} and C\textsubscript{2}). Thus, under a Separating equilibrium,

\textsuperscript{24} In Ellsberg’s model, a player’s critical risk represents the maximum risk of conflict the player is willing to tolerate. At any higher risk level, a rational player simply cooperates. Hence, the lower a player’s critical risk, the more likely it is to cooperate; the higher a player’s critical risk, the less likely it is to cooperate.

\textsuperscript{25} The distinction between subcases C\textsubscript{1} and C\textsubscript{2} has no bearing on the present discussion. See Appendix for details.
Challenger’s behavior is fully consistent with our initial conclusion: ceteris paribus, as its critical risk increases, so does Challenger’s propensity to foment a crisis.

The same general pattern holds when Defender is likely staunch (i.e., when $p_{Def} > e_1$) and either a Bluff or Hold Firm equilibrium is in play—as long as Challenger’s critical risk is low (Case A) or moderate (Case B). But a very interesting anomaly arises when Challenger is highly motivated (Cases C1 and C2).

Before we discuss this irregularity, however, notice that even when Challenger’s motives are intense (i.e., $c_1 > e_1$), there is always a region under which no demand is made and Challenger is deterred. One might expect as much in these cases, given that Defender is most likely staunch. But the high likelihood that Defender is staunch and will stand by Protégé is a necessary, but not a sufficient, condition for deterrence to hold against a determined Challenger. In general, for the Status Quo to survive, Protégé must also likely be disloyal. Since a disloyal Protégé always holds firm (under either equilibrium form), and a staunch Defender always supports a disloyal Protégé (under a Hold Firm equilibrium), Challenger can always be dissuaded from issuing a demand at node 1. A reputation for disloyalty, then, can serve Protégé well. In the context of this conclusion it is easy to appreciate the strategic rationale of Lord Palmerston’s well-known maxim about the impermanence of Great Britain’s commitments: “We have no eternal allies and we have no perpetual enemies. Our interests are eternal and perpetual, and these interests it is our duty to follow.”

Observe, though, that in Cases C1 and C2, the left-hand boundary of the (unshaded) region where no challenge takes place slopes upward and to the right. This means that, against a determined Challenger, there are conditions under which Protégé’s threat to realign (directed at Defender) is in fact counterproductive. What is particularly striking about this observation is that exactly the opposite pattern occurs...
under either a Settlement or a Separating equilibrium where the slope of the left-hand boundary of the no challenge region is negative. Under most conditions, then, Protegé’s threat to realign increasingly induces Defender’s support (at node 3) which, in turn, tends to discourage Challenger from initiating a crisis. But when Challenger is highly motivated and its critical risk high, it is possible that a heightened likelihood that Protegé is disloyal will actually make matters worse and provoke a challenge! Not only does our model reveal this unexpected behavioral consequence, it also identifies, precisely, the conditions under which it comes into play.

One could attempt to explain this anomaly by referring to Challenger’s preferences. Recall that we assume that Challenger prefers Protegé Realigns to Protegé Loses, so that one might conjecture that, ceteris paribus, Challenger would be more motivated to issue a demand as the likelihood of realignment increases. But this explanation does not generally hold. Indeed, when Challenger’s motivation is low or moderate, it runs the other way. The opportunity to break up Defender’s and Protegé’s relationship only changes Challenger’s behavior when Challenger’s critical risk is high, beyond the threshold $e_1$.

To put this in a slightly different way, under certain conditions a determined Challenger might chose to foment a crisis in the hope of disrupting a fragile strategic relationship. The prospect of separating Defender and Protegé, however, is a long shot; it is also quite risky since this motivation comes into play only under a Hold Firm equilibrium. Although Protégé is likely disloyal under a Hold Firm equilibrium ($p_{p_{df}}>d_1$), it is also the case that Defender is likely staunch ($p_{d_{df}}>e_1$). Since a staunch Defender always supports Protégé under a Hold Firm equilibrium, crises played out in this strategic environment tend to escalate to Conflict.

One can well imagine probes and challenges to the status quo, even in a strict bipolar system, designed not only to demonstrate a defender’s perfidiousness, but also to weaken the bonds of its supporting coalition. But if alignment considerations are generally of greater salience in multipolar systems, as Snyder (1976) and Waltz (1979) contend, then our model suggests yet another reason, beyond those traditionally cited, why multipolar systems are likely less stable than bipolar systems (Waltz, 1964; Gaddis, 1986; Mearsheimer, 1990). As corollaries, we hypothesize, ceteris

---

**Table 5. Challenger’s Node 1 Choice at the Perfect Bayesian Equilibrium of the Tripartite Crisis Game**

<table>
<thead>
<tr>
<th>Challenger’s Critical Risk $e_1$</th>
<th>Settlement Equilibrium $p_{d_{df}}&lt;e_2$</th>
<th>Separating Equilibrium $e_2&lt;p_{d_{df}}&lt;e_1$</th>
<th>Hold Firm Equilibrium $p_{d_{df}}&gt;e_1$, $p_{p_{df}}&gt;d_1$</th>
<th>Bluff Equilibrium $p_{d_{df}}&gt;e_1$, $p_{p_{df}}&lt;d_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low: $e_1&lt;e_2$</td>
<td>Case A</td>
<td>Challenger sometimes demands</td>
<td>Challenger never demands</td>
<td>Challenger sometimes demands</td>
</tr>
<tr>
<td>Medium: $e_2&lt;e_1&lt;e_1$</td>
<td>Case B</td>
<td>Challenger always demands</td>
<td>Challenger sometimes demands</td>
<td>Challenger never demands</td>
</tr>
<tr>
<td>High: $e_1&lt;e_1$</td>
<td>Case C_1</td>
<td>Outcome is always Challenger Wins</td>
<td>Challenger always demands</td>
<td>Challenger sometimes demands</td>
</tr>
<tr>
<td></td>
<td>Case C_2</td>
<td></td>
<td></td>
<td>Challenger always demands</td>
</tr>
</tbody>
</table>

under either a Settlement or a Separating equilibrium where the slope of the left-hand boundary of the no challenge region is negative. Under most conditions, then, Protegé’s threat to realign increasingly induces Defender’s support (at node 3) which, in turn, tends to discourage Challenger from initiating a crisis. But when Challenger is highly motivated and its critical risk high, it is possible that a heightened likelihood that Protegé is disloyal will actually make matters worse and provoke a challenge! Not only does our model reveal this unexpected behavioral consequence, it also identifies, precisely, the conditions under which it comes into play.

One could attempt to explain this anomaly by referring to Challenger’s preferences. Recall that we assume that Challenger prefers Protegé Realigns to Protegé Loses, so that one might conjecture that, ceteris paribus, Challenger would be more motivated to issue a demand as the likelihood of realignment increases. But this explanation does not generally hold. Indeed, when Challenger’s motivation is low or moderate, it runs the other way. The opportunity to break up Defender’s and Protegé’s relationship only changes Challenger’s behavior when Challenger’s critical risk is high, beyond the threshold $e_1$.

To put this in a slightly different way, under certain conditions a determined Challenger might chose to foment a crisis in the hope of disrupting a fragile strategic relationship. The prospect of separating Defender and Protegé, however, is a long shot; it is also quite risky since this motivation comes into play only under a Hold Firm equilibrium. Although Protégé is likely disloyal under a Hold Firm equilibrium ($p_{p_{df}}>d_1$), it is also the case that Defender is likely staunch ($p_{d_{df}}>e_1$). Since a staunch Defender always supports Protégé under a Hold Firm equilibrium, crises played out in this strategic environment tend to escalate to Conflict.

One can well imagine probes and challenges to the status quo, even in a strict bipolar system, designed not only to demonstrate a defender’s perfidiousness, but also to weaken the bonds of its supporting coalition. But if alignment considerations are generally of greater salience in multipolar systems, as Snyder (1976) and Waltz (1979) contend, then our model suggests yet another reason, beyond those traditionally cited, why multipolar systems are likely less stable than bipolar systems (Waltz, 1964; Gaddis, 1986; Mearsheimer, 1990). As corollaries, we hypothesize, ceteris...
paribus, (1) that extended deterrence relationships are more prone to succeed under bipolarity and (2) that crises that erupt under multipolarity are more likely to escalate.

What should Protégé do? The answer is not clear-cut. Against a weakly or even a moderately dissatisfied Challenger, an ambiguous alignment policy can help by strengthening Defender’s incentive to support Protégé, thereby deterring Challenger. But against a highly motivated Challenger, a shaky relationship could prove fatal by encouraging Challenger to initiate a crisis. To bolster deterrence in this case, Protégé must tilt, ever so gingerly, toward Defender.26 Of course, this strategem may cost Protégé some autonomy, as well as some leverage in its relationship with Defender (Morrow, 1991). Moreover, other potential challengers may see Defender as less likely to stand by Protégé in a crisis (because Defender will have less to lose), so Protégé may face additional challenges.

Protégé, then, is conflicted. Sometimes, a reputation for disloyalty makes a challenge less likely. But not always. Paradoxically, Protégé may attain its greatest bargaining leverage prior to negotiating its relationship with Defender—aligning too closely with Defender may increase Protégé’s risk, so it may demand greater compensation for doing so.

Provided that Defender’s (and Protégé’s) credibilities are past certain thresholds (see Appendix for details), Defender’s best bet would be to demonstrate a stronger relationship with Protégé. A general alliance or a mutual defense pact would serve this purpose well. Increasing the likelihood of Defender’s intervention or reducing the likelihood of Protégé’s realignment strengthens deterrence by dampening a highly motivated Challenger’s incentive to contest the status quo. And, of course, the absence of a crisis makes realignment even less likely.

It is clear that conflict avoidance was no small part of Bismarck’s motivation for negotiating the Dual Alliance with Austria-Hungary in 1879 (Massie, 1991:79; Kissinger, 1994:158–159). But this explanation for the union of the two German-speaking states is incomplete, for this was no ordinary alliance.

To begin, it was “the first permanent arrangement in peace-time between two Great Powers since the end of the ancien régime” (Taylor, 1954:264). In addition, the particulars of the agreement fully reflected the standpoint of Austria, the weaker and more dependent partner. Austria, afraid of offending France and England, sought a defensive pact aimed directly against Russia. Although Bismarck preferred a more general arrangement, the negotiated treaty embraced the Austrian position.

As well, Germany received few defensive benefits from the alliance: Germany did not need Austria if Russia attacked unless, of course, France joined in. More over, since German interests almost dictated that Germany would have to defend Austria in the event of a Russian challenge, a formal alliance could have only a minimal impact on the credibility of Germany’s (extended) deterrent threat. In other words, $p_{Def}$ was already quite high. (Note that this is precisely the condition under which the high likelihood that Protégé is disloyal can increase the probability of a crisis.)

Still, Bismarck pushed hard to finalize this treaty. When the Kaiser resisted Bismarck’s policy recommendation, he even threatened to resign. Why then would Bismarck place so much importance on a one-sided arrangement that did not fully reflect his own preferences and, seemingly, offered Germany limited strategic advantages? The answer lies, in large part, in the impact the alliance would have on Austria’s likely behavior and, by extension, the alignment status of the European system. As Snyder (1997: 90) explains:

Bismarck’s primary motive in negotiating an alliance with Austria was preclusive: to prevent Austria from allying elsewhere. ... Bismarck reasoned ... that if Austria were unsupported by Germany in a severe crisis with Russia, the internal

26 Subtlety is required here, as too much loyalty could also precipitate a crisis.
political balance might shift and Austria might be driven by its weakness to make a deal with Russia about the Balkans, leading to an anti-German alliance with Russia and France.

If Snyder is correct, Bismarck must have thought that, in the event of a crisis with Russia, there was some good chance that German support for Austria could fail to materialize.\textsuperscript{27} Still, it is unlikely that Bismarck sought an alliance with Austria solely to tie Germany’s (i.e., the Kaiser’s) hands, since actual German intentions (and likely contingent behavior) were somewhat beside the point, of little \textit{immediate} strategic consequence. What really mattered, in both theory and practice, was Russia’s perception of Germany’s intentions (i.e., $p_{D(i)}$). The German extended deterrent threat, however, was naturally high.\textsuperscript{28} In consequence, it did not need to be shored up by means of a formal alliance.

By contrast, the risk that a forsaken Austria might drift outside the German orbit weighed heavily on Bismarck’s mind (Langer, 1950:180). It is important to keep in mind that until mid-1879 Bismarck had resisted repeated Austrian overtures for a formal alliance. The explanation is simple: until then there was no good reason to worry about Austrian loyalty or, by extension, a Russian challenge. What changed suddenly was news that Austria’s pro-German Foreign Minister, Count Julius Andrásy, planned to resign. Anticipating an Austrian foreign office far less committed to Germany, Bismarck moved quickly to conclude an alliance before Andrásy left office.

Our model explains why. Clearly, Bismarck’s policy push was driven by a desire to alter Russia’s expectation of Austrian behavior in the event of a challenge (Taylor, 1954:263). Tellingly, Bismarck exaggerated the extent of the Russian threat in order to convince the Kaiser of the necessity of a formal agreement (Taylor, 1954:259–60). By reducing the probability that Austria might move to join an anti-German coalition, Bismarck hoped to forestall an immediate crisis with Russia. In turn, this would buy him the time necessary to revive the Three Emperors’ League, thereby isolating France and solidifying the gains made in the wars of 1866 and 1870. The intricate system that Bismarck devised helped keep Europe at peace until 1914. It is likely no coincidence that, when war finally came, it came in the context of a confrontation between Austria and a determined Russia in the Balkans that escalated to war only after Germany, fearful that the Dual Alliance would otherwise collapse, issued its partner a “blank check” (Kagan, 1995:191).

Summary and Conclusions

Our purpose has been to explore the dynamics of crisis bargaining in a state system in which policy options are conditioned by flexible alignment patterns. To this end we developed and analyzed the Tripartite Crisis Game, a three-person game among Challenger, Defender, and Protégé. Challenger initiates play by either demanding or not demanding a concession from Protégé. If a demand is made, Protégé either concedes or holds firm. If Protégé holds firm, Defender either supports or does not support Protégé. If Defender chooses to support Protégé, Challenger either backs down or presses on. If Defender withholds support and Protégé must capitulate, Protégé decides whether to realign.

\textsuperscript{27} In the absence of this belief, Snyder’s argument is dubious. Langer (1950: 175–76) asserts that as early as 1876 Bismarck was of the opinion that “Germany could not afford to see Austria completely defeated and deprived of her position as a great power” (emphasis added), suggesting that there were thresholds that would trigger German support (or nonsupport). Langer concludes that “it does not follow from [Bismarck’s pursuit of an alliance with Austria] that the chancellor had definitely decided to back Austria against Russia.”

\textsuperscript{28} This is not to say that there was no doubt about German intentions. British prime minister Benjamin Disraeli “was not at all certain that Germany would come to the aid of Austria in the event of a Russian attack” (Langer, 1950:187). And if there was some doubt in Britain, it is likely that there was also some uncertainty in Russia about the extent to which Germany would stand by Austria in a crisis.
The Tripartite Crisis Game is designed to capture in a single game form some of the tensions implicit in Glenn Snyder’s “Alliance” and “Adversary” games, two theoretically isolated models. For example, in the Tripartite Crisis Game, Defender’s threat to support Protégé acts to deter Challenger. At the same time this threat puts Defender at risk by encouraging Protégé to resist reaching a low-cost accommodation with Challenger. Similarly, Protégé’s threat to realign should help induce Defender’s support. But it may also encourage Challenger to foment a crisis. The specific preference assumptions we make in the Tripartite Crisis Game are motivated by a desire to heighten such tensions. Of course, they also limit the cases to which the conclusions of the model apply. As well, the assumptions define two types of each player: Challenger may be determined or hesitant; Defender may be staunch or perfidious; and Protégé may be loyal or disloyal.

Our results under complete information about preferences are straightforward and unexceptional. No demand is made, and the status quo obtains, whenever Challenger is hesitant, or if Defender is staunch and Protégé is disloyal. Otherwise, Protégé concedes and Challenger wins.

More subtle insights begin to emerge, though, when information is incomplete. To simplify our analysis, we assume that Challenger is always determined, that is, prefers conflict to backing down in a crisis it precipitates, and that this is known to all the players. Still, we vary the extent to which Challenger is motivated, distinguishing three levels of incentives: low, medium, and high. These incentives interact with Defender’s and Protégé’s preferences and determine behavior in the Tripartite Crisis Game.

For the most part, rational behavior in the Tripartite Crisis Game is consistent with expectations. In general, Challenger’s tendency to foment a crisis increases with its motivation (critical risk). But the rupture of deterrence is not certain even when Challenger’s motivation is extremely strong, provided Defender is very likely to support Protégé in the face of Challenger’s demand, and that Protégé is very likely to realign should it be abandoned by Defender. Our findings confirm the conclusions of several recent theoretical and empirical studies: Challenger’s relative satisfaction with the status quo makes deterrence success more likely. But a highly motivated Challenger is not sufficient to break down deterrence. Only if Protégé is sufficiently loyal to Defender, or if Defender’s commitment to Protégé is sufficiently weak, is the status quo unlikely to survive.

Our most surprising result concerns the impact of Protégé’s intentions on Challenger’s optimal behavior. When Challenger’s motivation is low or moderate, Protégé’s promise/threat serves only to dissuade Challenger from initiating a crisis. But when Challenger is highly motivated, Protégé’s intention to realign may actually provoke Challenger. Our analysis not only reveals this unusual behavioral pattern, but suggests when it is operative. This pattern is not manifest in any complete information variant of the Tripartite Crisis Game.

This result puts a somewhat finer point on Smith’s (1995:418) conclusion that “the empirical observation that alliances are often unreliable is a result of sampling.” In Smith’s model an aggressor is more likely to attack an unreliable alliance than a reliable one. But our analysis of the Tripartite Crisis Game indicates that this principle may apply for one partner (Defender), but not always for the other (Protégé). In other words, dependability usually—but not always—enhances a coalition’s ability to deter. Another feature of our model is that deterrence can fail even when Defender is very likely staunch, suggesting that Challenger’s decision to initiate a crisis may depend more critically on Protégé’s attitude than on Defender’s. In sum, the selection bias issue with respect to alliance reliability is not nearly as straightforward as Smith’s model implies.29

29 Werner (2000) comes to a similar conclusion.
That Protégé’s disloyalty can cut both ways, sometimes deterring a crisis and sometimes encouraging one, seems to us another “paradox of war.” It has been noted before that great power interactions often produce unintended consequences, that the relationship between inputs and outputs is not always linear. Highly capable states may risk defeat just like weak ones (Maoz, 1990) and, as we have shown, peace may depend on an alignment that is neither too tight nor too loose.

In addition to integrating Snyder’s Alliance and Adversary games in a single theoretical structure, and setting out optimal behavior in the Tripartite Crisis Game, our analysis of alignment patterns also allows us to isolate the circumstances that give rise to the “deterrence-vs.-restraint” dilemma. Specifically, only under the conditions that define a Bluff equilibrium does a staunch Defender have an incentive to offer less-than-certain support to Protégé. Bluff equilibria occur when Defender is likely staunch and Protégé is likely loyal. By withholding a modicum of support, Defender encourages Protégé to reach an accommodation with Challenger should deterrence fail and, at the same time, tries to discourage Challenger from fomenting a crisis. There are times when the stratagem works.30

This is an important theoretical insight for two reasons. First, Defender’s equilibrium behavior runs counter to the conventional wisdom that the probability of deterrence success is maximized when threats are communicated clearly.31 Second, it provides a compelling answer to Fearon’s puzzle: when strong commitments are possible, why are signals sometimes halfhearted? The deliberately vague support of the United States codified in the Taiwan Relations Act of 1979 is explained by the Bluff equilibrium in the Protégé–Defender subgame.32

Finally, our analysis of the Tripartite Crisis Game confirms Snyder’s contention that, in an extended deterrence relationship, the stronger motivation to align formally may lie with Defender, not Protégé, which may help to explain why Bismarck felt compelled to strike such a one-sided deal with Austria in 1879. Moreover, Defender’s motivation to draw closer to Protégé need not be rooted in an effort to bolster its own credibility. Rather, it may be motivated by the necessity to dampen Protégé’s proclivity to abrogate its relationship with Defender. To be sure, there are conditions under which Protégé’s willingness to realign actually enhances deterrence—because it makes Defender more likely to stand by Protégé. But there are also conditions under which flexible alignment policies serve both to undermine the status quo and to increase the chance of conflict. Such considerations help to explain why some alliances are associated with crises and war, and why others can be stabilizing forces in international relations.

Appendix

This Appendix contains the analysis of the Tripartite Crisis Game (TCG) discussed in the text. The TCG is a three-person non-cooperative game of incomplete information with three players, Challenger (Ch), Defender (Def), and Protégé (Pro). Figure 1 shows the TCG in extensive form. The six possible outcomes are Status Quo (SQ), Challenger Wins (CW), Protégé Realigns (PR), Protégé Loses (PL), Challenger Concedes (CC), and Conflict (C). The players’ preferences over these outcomes are

\begin{align*}
\text{Challenger: } & \text{CW} > \text{Ch PR} > \text{Ch PL} > \text{Ch SQ} > \text{Ch [C or CC]} \\
\text{Defender: } & \text{SQ} > \text{Def CC} > \text{Def CW} > \text{Def PL} > \text{Def [C or PR]} \\
\text{Protégé: } & \text{SQ} > \text{Pro CC} > \text{Pro C} > \text{Pro CW} > \text{Pro [PR or PL]} \\
\end{align*}

30 In July 1914, the British attempt to restrain Russia and deter Germany failed. By contrast, in the 1870s and 1880s, Bismarck had far greater success implementing what Snyder (1984:481–83) calls a “saddle strategy.”
31 Huth (1988:3) asserts that “in the context of extended deterrence, uncertainty is likely to undermine the credibility of a defender’s threat.” Similarly, Lebow (1981:85) lists a clearly communicated threat as one of four conditions crucial for deterrence success.
32 Christensen (2000) makes the case for an unambiguous U.S. commitment to Taiwan.
where a player’s relative preference for outcomes in brackets depends on its type. The TCG allows two possible types for each player. A determined Challenger prefers Conflict (C) to Challenger Concedes (CC) whereas a hesitant Challenger prefers Challenger Concedes (CC) to Conflict (C). A staunch Defender prefers Conflict (C) to Protégé Realigns (PR) whereas a perfidious Defender prefers Protégé Realigns (PR) to Conflict (C). A loyal Protégé prefers Protégé Loses (PL) to Protégé Realigns (PR) whereas a disloyal Protégé prefers Protégé Realigns (PR) to Protégé Loses (PL).

In general, we denote the utility of outcome $O$ to Challenger by $c_{O}$, to Defender by $d_{O}$, and to Protégé by $e_{O}$. However, in order to account for the different types of each player, we allow two different values for the utility of one crucial outcome for that player, as follows:

- $c_{C+}$ is a determined Challenger’s utility for Conflict (C);
- $c_{C-}$ is a hesitant Challenger’s utility for Conflict (C);
- $d_{C+}$ is a staunch Defender’s utility for Conflict (C);
- $d_{C-}$ is a perfidious Defender’s utility for Conflict (C);
- $e_{PR+}$ is a disloyal Protégé’s utility for Protégé Realigns (PR); and
- $e_{PR-}$ is a loyal Protégé’s utility for Protégé Realigns (PR).

Consistent with the preference restrictions above, we assume that

\[
\begin{align*}
    c_{CW} > e_{PR} > e_{PL} > e_{SQ} > c_{C+} > c_{CC} > c_{C-} \\
    d_{SQ} > d_{CC} > d_{CW} > d_{PL} > d_{PR} > d_{C-} \\
    e_{SQ} > e_{CC} > e_{C} > e_{CW} > e_{PR+} > e_{PL} > e_{PR-}
\end{align*}
\]  

Together these 21 fixed utility values, called the preference parameters of the TCG, constitute a discrete (binary) model of types.

When each player’s type is common knowledge, the TCG is a game of complete information. Each of the eight complete information versions of the TCG has a unique subgame-perfect equilibrium, given in Table 3 of the text.

To express the TCG as a game of incomplete information, we represent Challenger’s utility for Conflict as a binary random variable, $C_{C}$, satisfying

\[
C_{C} = \begin{cases} 
    c_{C+} & \text{with probability } p_{Ch} \\
    c_{C-} & \text{with probability } 1 - p_{Ch}
\end{cases}
\]

so that Ch is determined with probability $p_{Ch}$ and hesitant otherwise. Similarly, Defender’s utility for Conflict is a binary random variable, $D_{C}$, satisfying

\[
D_{C} = \begin{cases} 
    d_{C+} & \text{with probability } p_{Def} \\
    d_{C-} & \text{with probability } 1 - p_{Def}
\end{cases}
\]

so that Def is staunch with probability $p_{Def}$ and perfidious otherwise. Finally, Protégé’s utility for Protégé Realigns is a binary random variable, $E_{PR}$, satisfying

\[
E_{PR} = \begin{cases} 
    e_{PR+} & \text{with probability } p_{Pro} \\
    e_{PR-} & \text{with probability } 1 - p_{Pro}
\end{cases}
\]

so that Pro is disloyal with probability $p_{Pro}$ and loyal otherwise. In games of incomplete information, the values of the probabilities $p_{Ch}$, $p_{Def}$, and $p_{Pro}$ are assumed to be common knowledge, but each player’s realized (actual) type is known only to that player. There are eight possible combinations of extreme values, either 0 or 1, for $p_{Ch}$, $p_{Def}$, and $p_{Pro}$; these are the eight complete information versions of the TCG.

When information is genuinely incomplete (at least one $p$ lies strictly between 0 and 1), the appropriate solution concept is the Perfect Bayesian Equilibrium (PBE) (Fudenberg and Tirole, 1991). A PBE specifies each player’s action (which may depend on its type) at each of its decision nodes; in addition, a PBE specifies how each player’s beliefs about other players’ types evolve as information about their
actions becomes available. In this Appendix, we discuss the PBE of the TCG in two cases: when \( p_{Ch} = 0, 0 < p_{Def} < 1, \) and \( 0 < p_{Pro} < 1 \) (Ch is known to be hesitant, but the other players’ types are uncertain), and when \( p_{Ch} = 1, 0 < p_{Def} < 1, \) and \( 0 < p_{Pro} < 1 \) (Ch is known to be determined, but the other players’ types are again uncertain).

First we consider the case \( p_{Ch} = 0, 0 < p_{Def} < 1, \) and \( 0 < p_{Pro} < 1 \) (Ch is known to be hesitant). The game of Figure 1 is then easy to analyze. First, Ch always picks Back Down (resulting in outcome CC) at node 4b. Consequently, Def always picks Support at node 3, because then the outcome (for certain) will be CC, which it prefers to either of the two outcomes PR and PL, which could result if it chose Not Support. At node 2, Pro picks Hold Firm because it leads inevitably to CC, which it prefers to the only available alternative, CW. At node 1, therefore, Ch is faced with a choice of Demand, leading eventually to CC, or Not Demand, leading to SQ. Ch prefers SQ to CC, so it always chooses Not Demand at node 1.

Because all of these choices are strictly dominant when Ch is known to be hesitant, it follows that the only possible PBE in this circumstance is as described: Ch selects Not Demand at node 1; if the game reaches node 2, Pro plans to select Hold Firm; if the game reaches node 3, Def plans to select Support; if the game reaches node 4b, Ch plans to select Back Down. Node 4a is off the equilibrium path, but Pro’s choice there is fixed also; it is Realign if Pro is disloyal, and Not Realign if Pro is loyal. This analysis provides a very strong argument that the outcome of the TCG if Challenger is known to be hesitant is Status Quo.

The remainder of this Appendix concerns the case \( p_{Ch} = 1, 0 < p_{Def} < 1, \) and \( 0 < p_{Pro} < 1 \), when Ch is known to be determined. In the complete information case (see the text), the unique SPE outcome of the TCG is SQ if Def is staunch and Pro is disloyal, and CW otherwise. Thus we expect that a PBE will almost surely yield outcome SQ near \((p_{Def}, p_{Pro}) = (1, 1)\), and that a PBE will almost surely yield outcome CW near \((p_{Def}, p_{Pro}) = (0, 0), (0, 1), \) or \((1, 0)\).

If Ch is known to be Hard, then the outcome of the game will be C should the game ever reach node 4b. Figure 2 shows how the TCG is reduced when \( p_{Ch} = 1, \) and highlights the subgame played by Pro ete and Defender after an initial demand by Ch. This subgame will be called the Protégé–Defender Subgame (PDS).

Our first objective is to solve the PDS. Define the strategic variables \( y \) and \( z \) as shown in Figure 2. Note that a PBE in the PDS must specify

\[
\begin{align*}
\gamma_D &= \Pr\{\text{Pro chooses Hold Firm | Pro is disloyal}\} \\
\gamma_L &= \Pr\{\text{Pro chooses Hold Firm | Pro is loyal}\} \\
z_S &= \Pr\{\text{Def chooses Support | Def is staunch}\} \\
z_P &= \Pr\{\text{Def chooses Support | Def is perfidious}\}
\end{align*}
\]

The probabilities of the complementary choices are given by the complements of these strategic variables; for example, a disloyal Pro chooses Concede with probability \( 1 - \gamma_D \), and a staunch Def chooses Not Support with probability \( 1 - z_S \).

At the beginning of the PDS, Pro believes that Def is staunch with probability \( p_{Def} \), and Def believes that Pro is disloyal with probability \( p_{Pro} \). A PBE cannot depend on the evolution of Pro’ s belief through the game, as its decision at node 2 is based on its initial beliefs about Def, and its decision at node 4a depends only on its realized (actual) type. But the evolution of Def’s belief about Pro does figure into a PBE: prior to acting at node 3, Def can revise its probability that Pro is disloyal on the basis of Pro’s choice at node 2, which is of course known to Def. Denote by \( q \) Def’s revised probability that Pro is disloyal, given that the game reaches node 3.

Before determining the PBE of the PDS, define the following parameters:

\[
\begin{align*}
d_1 &= \frac{d_{PL} - d_{CP}}{d_{PL} - d_{PR}}; e_1 = \frac{e_{CW} - e_{PL}}{e_{C} - e_{PL}}; e_2 = \frac{e_{CW} - e_{PR+}}{e_{C} - e_{PR+}}.
\end{align*}
\]

It is easy to show directly, using (1), that \( 0 < d_1 < 1 \) and \( 0 < e_2 < e_1 < 1 \). The constants \( d_1, e_1, \) and \( e_2 \) define the conditions for a PBE.
Theorem 1:

A Perfect Bayesian Equilibrium (PBE) of the Protégé–Defender Subgame (PDS) is a 5-tuple \((y_D, y_L; z_S, z_P, q)\) satisfying the following conditions:

\[(C1) \quad y_D = \arg\max_{0 \leq y \leq 1} y(z_a - e_2)\]

\[(C2) \quad y_L = \arg\max_{0 \leq y \leq 1} y(z_a - e_1)\]

\[(C3) \quad \text{If } y_D + y_L > 0, \quad q = \frac{p_{EYD}}{p_{EYD} + (1 - p_E)y_L}\]

\[(C4) \quad z_S = \arg\max_{0 \leq z \leq 1} z(q - d_1)\]

\[(C5) \quad z_P = 0\]

where \(z_a = p_{Def} z_S + (1 - p_{Def}) z_P\).

Proof:

Should the PDS reach node 4a, Pro will choose exactly as dictated by its type, so the outcome will be PR if Pro is disloyal and PL if Pro is loyal. Therefore, should the PDS reach node 3, Def must choose either Support (probability \(z\)) or Not Support (probability \(1 - z\)), knowing that Pro is disloyal with probability \(q\) and loyal with probability \(1 - q\). At the time Def makes its choice, its expected utility is

\[z D_C + (1 - z)[q d_{PR} + (1 - q)d_{PL}]\]

where \(D_C = d_{C+}\) if Def is staunch and \(D_C = d_{C-}\) if Def is perfidious. In the latter case, \(d_{C-} < \min\{d_{PR}, d_{PL}\}\), so that \(z_P\) should be minimized to maximize Def’s expected utility, producing (C5). If Def is staunch, Def’s expected utility at node 3 can be written

\[z_S d_{C+} + (1 - z_S)[q d_{PR} + (1 - q)d_{PL}] = q d_{PR} + (1 - q)d_{PL} + z_S(d_{PL} - d_{PR})(q - d_1)\]

Because \(d_{PL} > d_{PR}\), (C4) follows immediately.

Working backward, (C3) is the standard Bayesian update condition on Def’s probability that Pro is disloyal, given that a disloyal Pro Holds Firm at node 2 with probability \(y_D\) and a loyal Pro Holds Firm at node 2 with probability \(y_L\). Of course, if \(y_D = y_L = 0\), node 3 is not on the equilibrium path, and Bayesian updating cannot be applied.

Now consider the choice facing Pro at node 2. Because Pro knows Def to be staunch with probability \(p_{Def}\), Pro calculates the probability that Defender Supports at node 3 to be \(z_a = p_{Def} z_S + (1 - p_{Def}) z_P\). It follows that if Pro is disloyal, its expected utility at node 2 is

\[(1 - y_D)e_{CW} + y_D[z_a e_C + (1 - z_a)e_{PR+}] = e_{CW} + y_D(e_C - e_{PR+})(z_a - e_2)\]

(C1) now follows because \(e_C > e_{PR+}\). Analogously, the expected utility at node 2 of a loyal Pro is

\[(1 - y_L)e_{CW} + y_L[z_a e_C + (1 - z_a)e_{PL}] = e_{CW} + y_L(e_C - e_{PL})(z_a - e_1)\]

so that (C2) follows from \(e_C > e_{PL}\), completing the proof. ■

Theorem 2:

The non-transitional Perfect Bayesian Equilibria of the Protégé–Defender subgame are as given in Table A1.
This shows that by (C3). But then (C4) implies that therefore necessary. Clearly this class are called Settlement PBE. By (C1) and (C2), a necessary condition is that is another PBE with is possible if provided that these conditions are sufficient for a PBE, called a Bluff PBE. By (C2) are satisfied iff (C3) applies, and shows that is the only necessary condition for a Separating PBE; it is straightforward to verify that it is also sufficient. For a PBE with where, as usual, is required where, as usual, is required for a PBE with is required for a PBE with is required for a PBE with which, by (C4), is possible if . This shows that the conditions in classes 1a and 1b of Table A1 are necessary. It is easy to verify that they are also sufficient. The PBE of this class are called Settlement PBE.

Next we identify all PBE with which, by (C4), is possible if is transitional, as it exists only when as (C4) makes clear.

A PBE with and would have , by (C1) and (C2), and , by (C3). But then (C4) implies that so as required. It follows that is necessary. Clearly is impossible, as it would imply , and therefore , Likewise, (C4) shows that implies , and therefore . It follows that any PBE with is transitional, as it must satisfy . Only one possibility remains, namely, , which would be consistent with if in which turn can be arranged iff . Moreover, (C3) requires that so that can be arranged iff . Again, it is straightforward to show that these conditions are sufficient for a PBE, called a Bluff PBE.

<table>
<thead>
<tr>
<th>PBE Class</th>
<th>Existence Conditions</th>
<th>( \gamma_D )</th>
<th>( \gamma_L )</th>
<th>( z_S )</th>
<th>( z_H )</th>
<th>( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Settlement</td>
<td>( p_{D</td>
<td>j} \leq e_2 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1b. Steadfast Settlement</td>
<td>( p_{D</td>
<td>j} &gt; e_2 )</td>
<td>0</td>
<td>0</td>
<td>( \leq e_2 p_{Def} )</td>
<td>0</td>
</tr>
<tr>
<td>2. Separating</td>
<td>( e_2 \leq p_{Def} \leq e_1 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3. Hold Firm</td>
<td>( p_{Def} \geq e_1 ) and ( p_{Pr_0} \geq d_1 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( p_E )</td>
</tr>
<tr>
<td>4. Bluff</td>
<td>( p_{Def} \geq e_1 ) and ( p_{Pr_0} &lt; d_1 )</td>
<td>1</td>
<td>( u_L )</td>
<td>( e_1 p_{Def} )</td>
<td>0</td>
<td>( d_1 )</td>
</tr>
</tbody>
</table>

Proof:

Recall that \( e_1 > e_2 \). (C1) shows that at any PBE with \( \gamma_D < 1 \), \( z_a < e_2 \), from which it follows that \( z_a < 1 \), so that \( \gamma_L = 0 \) by (C2). Similarly, (C2) shows that at any PBE with \( \gamma_L > 0 \), \( z_a \geq e_1 \), from which it follows that \( z_a > e_2 \), so that \( \gamma_D = 1 \) by (C1). These observations limit the possibilities for a PBE; it must be the case that \( \gamma_L \leq \gamma_D \), and furthermore at least one of these probabilities must equal zero or one.

First we identify all PBE with \( \gamma_D = 0 \) and \( \gamma_L = 0 \). Clearly (C3) is void, and (C1) and (C2) are satisfied iff \( z_a \leq e_2 \). If \( p_{Def} \leq e_2 \), then there is a PBE of this class with \( z_S = 1 \) provided \( q \geq d_1 \). If \( p_{Def} > e_2 \), the condition on \( z_a \) is satisfied only if \( z_S \leq e_2 p_{Def} \), which, by (C4), is possible if \( q = d_1 \). This shows that the conditions in classes 1a and 1b of Table A1 are necessary. It is easy to verify that they are also sufficient. The PBE of this class is called Settlement PBE.

Next we identify all PBE with \( \gamma_D = 1 \) and \( \gamma_L = 0 \), which are called Separating PBE. By (C1) and (C2), a necessary condition is that \( e_2 \leq z_a \leq e_1 \), where \( z_a = p_{Def} z_S \). (C3) applies, and shows that \( q = 1 \), which implies that \( z_S = 1 \) by (C4), because \( d_1 < 1 \). This shows that \( e_2 \leq z_a \leq e_1 \) is the only necessary condition for a Separating PBE; it is straightforward to verify that it is also sufficient.

For a PBE with \( \gamma_D = 1 \) and \( \gamma_L = 1 \), (C1) and (C2) show that \( z_a \geq e_1 \) is required where, as usual, \( z_a = p_{Def} z_S \). By (C3), \( q = p_{Pr_0} \). Therefore, (C4) shows that \( p_{Pr_0} \geq d_1 \) is required for a PBE with \( z_S = 1 \), as is \( p_{Def} \geq e_1 \). It is easily verified that these conditions are necessary and sufficient for a PBE, called the Hold Firm PBE. There is another PBE with \( \gamma_D = 1 \), \( \gamma_L = 1 \), and \( z_S < 1 \), but it is transitional as it exists only when \( p_{Pr_0} = d_1 \), as (C4) makes clear.

A PBE with \( 0 < \gamma_D < 1 \) and \( \gamma_L = 0 \) would have \( z_a = e_2 \), by (C1) and (C2), and \( q = 1 \), by (C3). But then (C4) implies that \( z_S = 1 \), so \( z_a = p_{Def} \). It follows that \( p_{Def} = e_2 \) is required for a PBE with \( 0 < \gamma_D < 1 \) and \( \gamma_L = 0 \), so that any such PBE is transitional.

Finally, for a PBE with \( \gamma_D = 1 \) and \( 0 < \gamma_L < 1 \), (C1) and (C2) show that \( z_a = e_1 \) is necessary. Clearly \( q < d_1 \) is impossible, as by (C4) it would imply \( z_S = 0 \), and therefore \( z_a = 0 < e_1 \). Likewise, (C4) shows that \( q > d_1 \) implies \( z_S = 1 \), and therefore \( z_a = p_{Def} \). It follows that any PBE with \( \gamma_D = 1 \), \( 0 < \gamma_L < 1 \), and \( q > d_1 \) is transitional, as it must satisfy \( p_{Def} = e_1 \). Only one possibility remains, namely, \( q = d_1 \), which would be consistent with \( z_a = p_{Def} z_S = e_1 \) iff \( z_S = e_1 p_{Def} \), which in turn can be arranged iff \( p_{Def} \geq e_1 \). Moreover, (C3) requires that

\[
q = \frac{p_{Pr_0}}{p_{Pr_0} + (1 - p_{Pr_0}) \gamma_L} = d_1,
\]

which is equivalent to

\[
\gamma_L = u_L = \frac{p_{Pr_0}(1 - d_1)}{d_1(1 - p_{Pr_0})},
\]

so that \( 0 < \gamma_L < 1 \) can be arranged iff \( p_{Pr_0} < d_1 \). Again, it is straightforward to show that these conditions are sufficient for a PBE, called a Bluff PBE. ■
because the denominator of the fractional argument of the minimum is strictly positive, which together determine the outcome of the PDS, as given in Tables A1 and A2.

PBE Class | Def staunch Pro disloyal | Def staunch Pro loyal | Def perfidious Pro disloyal | Def perfidious Pro loyal
---|---|---|---|---
1. Settlement | CW | CW | CW | CW
2. Separating | C | CW | PR | CW
3. Hold Firm | C | C | PR | PL
4. Bluff | PR \(1 - e_1[p_{D\text{def}}]\) | CW \(1 - u_L\), PL \(u_L \left(1 - e_1[p_{D\text{def}}]\right)\) | PR | CW \(1 - u_L\)
   | C \(e_1[p_{D\text{def}}]\) | C \(u_L, e_1[p_{D\text{def}}]\) | PL \(u_L\)

Table A2 also lists the outcomes associated with each equilibrium class, as a function of the types of the players. These outcomes are easy to determine using Figure 1 and Table A1. For the Bluff PBE, note that several outcomes may be possible even when the players’ types are fixed.

In the following, we will ignore Steadfast Settlement PBE, as they are “implausible” PBE since they require \(q\) to be bounded away from 1. The implausibility arises from the fact that at a Settlement PBE, Def expects both types of Pro to Concede for certain, so that node 3 is “off the equilibrium path.” But should Pro unexpectedly hold firm, it is reasonable for Def [based, for example, on a comparison of (C1) and (C2)] to conclude that Pro is quite likely disloyal. But the requirement that \(q \leq e_2/p_{D\text{def}} < 1\) may in fact require Def to reduce its probability that Pro is disloyal. For further discussion see Zagare and Kilgour (2000, p. 152 and p. 206, n.6).

Figure 3 shows the regions of existence of the remaining four equilibrium classes in the \((p_{D\text{def}}, p_{P\text{pro}})\)-unit square. By discarding the Steadfast Settlement PBE, we have identified a specific PBE (and therefore a specific outcome of probabilistic combination of outcomes) for the PDS in each of the four regions shown.

With this information about the outcome of the PDS, we return to the TCG when Ch is known to be determined, i.e., in the case \(p_{C\text{h}} = 1\), \(0 < p_{D\text{def}} < 1\), and \(0 < p_{P\text{pro}} < 1\). Figure 2 makes it clear that we must consider whether Ch would prefer to choose No Demand, leading to outcome SQ, or Demand, leading to the outcome of the PDS. Obviously, which choice is in Ch’s interest depends on the values of \(p_{D\text{def}}\) and \(p_{P\text{pro}}\), which together determine the outcome of the PDS, as given in Tables A1 and A2.

Define the parameters \(e_1\) and \(e^*\) and the functions \(r(p_{D\text{def}})\) and \(s(p_{D\text{def}})\) as follows:

\[
e_1 = \frac{c_{PR} - c_{SQ}}{c_{PR} - c_{C+}}; \quad e^* = d_1 \min \left\{1, \frac{c_{CW} - c_{SQ}}{c_{CW} - e_1c_{C+} - d_1(1 - e_1)c_{PR} - (1 - d_1)(1 - e_1)c_{PL}} \right\} \tag{3}
\]

It is easy to show directly, using (1), that \(0 < e_1 < 1\). The quantity \(e^*\) is well-defined, as the denominator of the fractional argument of the minimum is strictly positive, because

\[
e_1c_{C+} + (1 - e_1)d_1c_{PR} + (1 - e_1)(1 - d_1)c_{PL} < e_1c_{PR} + (1 - e_1)d_1c_{PR} + (1 - e_1)(1 - d_1)c_{PR} = e_{PR} < c_{CW}.
\]

This shows that \(0 < e^* < d_1\). For \(0 \leq p_{D\text{def}} < 1\), define the functions \(r(p_{D\text{def}})\) and \(s(p_{D\text{def}})\) according to

\[
r(p_{D\text{def}}) = \frac{p_{D\text{def}}(c_{PL} - c_{C+}) - (c_{PL} - c_{SQ})}{(1 - p_{D\text{def}})(c_{PR} - c_{PL})}, \quad s(p_{D\text{def}}) = \frac{e_{CW} - e_{SQ}}{p_{D\text{def}}(c_{PR} - c_{C+}) + (c_{CW} - c_{PR})}. \tag{4}
\]

The graphs of \(r(p_{D\text{def}})\) and \(s(p_{D\text{def}})\) appear in each of the four parts of Figure 4. The curve sloping upward to the right represents \(r(p_{D\text{def}})\), while \(s(p_{D\text{def}})\) appears as a curve.
sloping downward to the right. More information about the geometric properties of the graphs of \(r(p_{\text{Def}})\) and \(s(p_{\text{Def}})\) is given below. These functions, and the constants \(c_1\) and \(c^*\), determine the PBE of the TCG.

**Theorem 3:**

When \(p_{\text{Ch}} = 1, 0 < p_{\text{Def}} < 1,\) and \(0 < p_{\text{Pro}} < 1,\) the Tripartite Crisis Game has a unique Perfect Bayesian Equilibrium, which results in the outcome Status Quo (if Challenger chooses No Demand) or the outcome of the Perfect Bayesian Equilibrium of the Protégé–Defender subgame (if Challenger chooses Demand). Challenger’s choices (for certain) are as follows:

1. If \(p_{\text{Def}} < e_2,\) Challenger Demands.
2. If \(e_2 < p_{\text{Def}} < e_1,\) Challenger Demands if \(p_{\text{Pro}} < s(p_{\text{Def}})\) and not if \(p_{\text{Pro}} > s(p_{\text{Def}}).
3. If \(e_1 < p_{\text{Def}}\) and \(d_1 < p_{\text{Pro}},\) Challenger Demands if \(p_{\text{Pro}} > r(p_{\text{Def}})\) and not if \(p_{\text{Pro}} < r(p_{\text{Def}})\).
4. If \(e_1 < p_{\text{Def}}\) and \(p_{\text{Pro}} < d_1,\) Challenger Demands if \(p_{\text{Pro}} < c^*\) and not if \(p_{\text{Pro}} > c^*.

**Proof:**

Given \(p_{\text{Def}}\) and \(p_{\text{Pro}},\) denote Ch’s utility for the outcome of the PDS to be \(u_{\text{Ch}}(p_{\text{Def}}, p_{\text{Pro}})\). Then Ch optimally chooses No Demand, leading to outcome SQ, if and only if

\[
\epsilon_{SQ} > u_{\text{Ch}}(p_{\text{Def}}, p_{\text{Pro}})
\]

If \(p_{\text{Def}} < e_2,\) the PDS has the Class 1 (Settlement) PBE, with outcome CW for certain. Thus, \(u_{\text{Ch}}(p_{\text{Def}}, p_{\text{Pro}}) = \epsilon_{CW} > \epsilon_{SQ}\). Clearly, Ch optimally chooses Demand, and the outcome is CW for certain.

Now suppose that \(e_2 < p_{\text{Def}} < e_1.\) The PBE of the PDS is of Class 2 (Separating), and Table A2 shows that Ch’s rational choice is Demand if and only if

\[
\epsilon_{SQ} < p_{\text{Def}} p_{\text{Pro}} c_{C+} + (1 - p_{\text{Def}}) p_{\text{Pro}} c_{PR} + (1 - p_{\text{Pro}}) c_{CW}.
\]

This condition is equivalent to \(p_{\text{Pro}} \leq s(p_{\text{Def}})\). Moreover, Ch’s uniquely optimal choice is Demand if the inequality is strict, and No Demand if the inequality fails.

Next suppose that \(e_1 < p_{\text{Def}}\) and \(d_1 < p_{\text{Pro}}.\) The PBE of the PDS is of Class 3 (Hold Firm), and Table A2 shows that Ch’s rational choice is Demand if and only if

\[
\epsilon_{SQ} < p_{\text{Def}} e_{C+} + (1 - p_{\text{Def}}) p_{\text{Pro}} c_{PR} + (1 - p_{\text{Def}})(1 - p_{\text{Pro}}) c_{PL}.
\]

This condition is equivalent to \(p_{\text{Pro}} \geq r(p_{\text{Def}})\). Moreover, Ch’s uniquely optimal choice is Demand if the inequality is strict, and No Demand if the inequality fails.

Finally, suppose that \(e_1 < p_{\text{Def}}\) and \(p_{\text{Pro}} < d_1.\) The PBE of the PDS is the Class 4 (Bluff) PBE. Using Table A2, it can be shown that the possible outcomes are CW, PR, PL, and C, with probabilities

\[
\frac{d_1 - p_{\text{Pro}}}{d_1}, \quad \frac{p_{\text{Pro}}(1 - d_1)}{d_1}, \quad \frac{p_{\text{Pro}}(1 - d_1)(1 - e_1)}{d_1}, \quad \text{and} \quad \frac{p_{\text{Pro}}e_1}{d_1},
\]

respectively. (Note that these probabilities are well-defined since \(0 < p_{\text{Pro}} < d_1.\) It follows that Ch’s rational choice is Demand if and only if

\[
\epsilon_{SQ} < \frac{p_{\text{Pro}} e_1}{d_1} c_{C+} + p_{\text{Pro}}(1 - e_1) c_{PR} + p_{\text{Pro}}(1 - d_1)(1 - e_1) c_{PL} + \frac{d_1 - p_{\text{Pro}}}{d_1} c_{CW}.
\]

Manipulation now shows that this condition is equivalent to \(p_{\text{Pro}} \leq c^*,\) and the theorem follows.
To summarize the conclusions of Theorem 3, the outcome is SQ whenever Ch’s rational choice is No Demand. But when Ch chooses Demand, the outcome of the TCG is the outcome of the PDS.

The set of values of \((p_{Def}, p_{Pro})\) such that the PBE of the TCG is No Demand by Challenger, leading to the Status Quo outcome, is called the No-Demand region. Based on Theorem 3, the geometric properties of the No-Demand region can be determined.

First, the geometric properties of the graphs of \(r(p_{Def})\) and \(s(p_{Def})\) must be understood. It is easy to verify from (4) that the function \(r(p_{Def})\) is strictly increasing and satisfies \(r(0) < 0\) and \(r(p_{Def}) \to \infty\) as \(r(p_{Def}) \to 1^-\).

Moreover, \(r(p_{Def}) = 1\) if and only if \(p_{Def} = e_1\). Also from (4), the function \(s(p_{Def})\) is strictly decreasing, with \(s(0) > 0\), \(0 < s(1) < 1\), and \(s(p_{Def}) = 1\) if and only if \(p_{Def} = e_1\). Thus the graphs of \(r(p_{Def})\) and \(s(p_{Def})\) cross at the point \((p_{Def}, p_{Pro}) = (e_1, 1)\); to the left of this point, \(r(p_{Def}) < 1\), and to the right of this point, \(s(p_{Def}) < 1\).

The other important information needed to draw the No-Demand region is the connection between the value of \(c_1\) and the relative values of \(c^*\) and \(d_1\). As is clear from (3), \(c^* \leq d_1\). It can be shown directly that \(c^* = d_1\) if and only if \(r(e_1) \leq d_1\). Because \(r(p_{Def})\) is a strictly increasing function satisfying \(r(e_1) = 1\), it follows that a necessary condition for \(c^* = d_1\) is \(e_1 \geq e_1\). In particular, if \(e_1 < e_1\), then \(c^* < d_1\).

With this information, the four cases depicted in Figure 4 can be distinguished. Note that the No-Demand region never includes any \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 1 (in the PDS), where Pro always chooses Concede. However, the remaining three PBE classes can appear in the No-Demand region, as follows:

Case A. \(e_1 < e_2\). The No-Demand region includes some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 2 (Separating), all of the \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 3 (Hold Firm), and some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 4 (Bluff). Moreover, the No-Demand region includes points where \(p_{Def}\) is arbitrarily close to \(e_2\).

Case B. \(e_2 < c_1 < e_1\). The No-Demand region includes some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 2 (Separating), all of the \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 3 (Hold Firm), and some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 4 (Bluff). However, all points in the No-Demand region have \(p_{Def}\) bounded away from \(e_2\).

Case C. \(e_1 < c_1 \) and \(c^* < d_1\). The No-Demand region includes some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 3 (Hold Firm) and some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 4 (Bluff). However, all points corresponding to PBE of Class 2 (Separating) are excluded from the No-Demand Region, as are some points corresponding to PBE of Class 3.

Case D. \(e_1 < c_1 \) and \(c^* = d_1\). The No-Demand region includes some \((p_{Def}, p_{Pro})\) points corresponding to PBE of Class 3 (Hold Firm) only. All points corresponding to PBE of Class 2 (Separating) and Class 4 (Bluff) are excluded from the No-Demand Region, as are some points corresponding to PBE of Class 3.

References


