Contracts Between Friends

Alliances, Reputation, and International Politics*

Erik Gartzke[†]
University of California, San Diego

4 June 2012

Abstract

Reputation is the "dark matter" of international politics. Orthodox explanations for diplomacy, treaties, alliances, deterrence, and other processes rely on reputation to link actions and payoffs across time. However, influential empirical research argues that reputations do not exist, or should not matter (Mercer 1996, 1997; Press 2004, 2005). The search for reputation is complicated because effects will generally be obscured by strategic interaction. Here, I examine a context where the consequences of reputation should be readily observable. I offer two versions of a formal model of alliance formation, one in which a potential attacker has no information about the distribution of defender preferences, and a second model in which the potential attacker has "noisy" information about preferences, mimicking the effects of reputation. Consistent with the noisy model, logit and GAM estimates report a non-linear relationship between state affinities and alliance formation. "Friends" have less need to signal an alignment by forming an alliance.

*Paper prepared for presentation at the European Political Science Association conference, Berlin, 21-23 June, 2012. I thank Scott Bennett, Lawrence Broz, Bruce Bueno de Mesquita, Christopher Butler, John Conybeare, Kristian S. Gleditsch, Robert Jervis, Brett Ashley Leeds, Forrest Nelson, William L. Reed, Jeffrey Ritter, Alastair Smith, Erik Voeten and seminar participants at New York University, the University of California, San Diego, and Texas A & M University for their comments and suggestions. Data will be made available by the author upon publication.

[†]University of California, San Diego. E-mail: `egartzke@ucsd.edu`. Web: `http://dss.ucsd.edu/ egartzke`.

Nations dispense with an alliance...when their interests so obviously call for concerted policies and actions that an explicit formulation of these interests, policies, and actions in the form of a treaty of alliance appears to be redundant. Hans Morgenthau (1973 [1948], page 181-182)

Most communications convey two messages: what the actor is saying and the fact that he needs to say it.... Robert Jervis (2001, page 298)

1 Introduction

Perceptions guide decision making in an uncertain world, providing individuals and societies with incentives to manipulate the beliefs of partners and competitors. Reputations reflect the residual effect of these manipulations in terms of common conjectures about subsequent behavior. Reputation thus functions as cognitive and conceptual "glue," binding past behaviors and outcomes to future punishments or rewards in theories of deterrence (Fearon 1994b, 2002 Huth 1988, 1997), alliance formation and reliability (Leeds 2003a, 2003b), diplomacy (Sartori 2002, 2005), treaties (Downs & Jones 2002, Simmons & Hopkins 2005), civil conflicts (Walter 2006a, 2006b) and other social processes. In addition to ensuring a level of temporal "stickiness" or consistency, reputation also helps to explain the whole array of verbal or symbolic acts associated with foreign policy. Without reputation, much of what we think we know about world politics no longer makes sense.

Unfortunately, there is not much evidence that reputation exists. Skepticism about the empirical validity of the construct is reflected in several influential studies (Mercer 1996, 1997; Press 2004, 2005). While these challenges may be correct, it also seems likely that reputation is simply difficult to demonstrate. Knowing where to look for telltale evidence of reputation is far from trivial. The effects of reputation can often be attributed to other factors, such as power or proximity. Alternately, the consequences of reputation may be no consequences at all. In many cases, countries probably absorb the impact of reputation through tacit or anticipatory actions that leave no observable traces (Cetinyan 2002). Researchers can of course model these effects of reputation, but with no empirical comparison, we are left to suppose (or doubt) without knowing for sure.

If evidence of reputation is hard to find, its absence would be difficult to contemplate. Without the "dark matter" of reputation, much of orthodox international relations theory will need to be substantially reconsidered. Countries are said to honor alliance obligations out of fear that failure to do so will cause other allies or potential partners to doubt the veracity of the nation's commitments (Morrow 2000, Gartzke & Gleditsch 2004). Treaties are often said to influence compliance by making it easier to identify and sanction defectors (Simmons 1998). Similarly, deterrence replaces defense only to the extent that potential attackers perceive verbal statements to be potent (Wagner 1982, Betts 1985), while diplomacy hinges on the credibility of both threats and promises (George 1991). If states do not have (or cannot use) reputations, then countries need not honor agreements or act on their own promises or demands. Pronouncements (public or confidential) are then pointless; if reputation does not exist, we should not observe the dense activity of deliberative, diplomatic, or bellicose interstate dialogue. Indeed, if reputation suddenly vanished, students and practitioners of world politics would want to recreate it, since nothing appears capable of taking its place.

Contemporary alliance theory emphasizes the role of alliances as signals, reducing uncertainty about whether a state is likely to intervene in a contest (Fearon 1997), as commitment devices, stiffening the resolve to intervene (Smith 1995, 1996), or as coordination mechanisms, allowing more efficient deterrence or defense (Altfeld 1984, Morrow 1991). While all three rely in reputation, alliance signaling most clearly involves attempts to change beliefs under uncertainty. Imagine that we reset international politics so that nations have no track record of characteristic behavior. Under these conditions, existing explanations for alliance behavior work quite well. States with affinities—"friends" as Smith calls them (1995, page 416)—co-ally to demonstrate their intentions. As time passes and regular patterns of affinity and interaction become discernible, however, a new dynamic enters alliance theory. Some relationships become the subject of common conjecture, introducing a tension between what nations know, or believe, and the role of alliances in informing nations about what they do not know. Alliances and reputations are substitutes; to the degree that affinities are already understood, states with reputations as fast friends should have less need to ally. If reputations are exact maps of future behavior, alliance signals become. If instead reputations do not exist, then states intending to cooperate are generally better off signaling affinity through an alliance. The empirical world is "noisy," containing both reputation and uncertainty. As Morgenthau suggests, friends may not need to signal affinity through an alliance. To paraphrase Jervis, alliances convey two messages, that states are friends, and that their friendship is doubted.

2 Reputation, Perceptions, and International Politics

Students of international relations are increasingly aware of the relationship between what actors believe and how they behave. Research on deterrence, for example, treats the credibility of threats as the critical dynamic determining success or failure (Powell 1990, Kilgour & Zagare 1991). Alliances may be seen as an effort to strengthen the credibility of deterrence by signaling a state's willingness to intervene (Morrow 1994). Signaling involves efforts by one actor to influence the perceptions of another actor in an environment of uncertainty. While providing important insights, existing models of perceptions in international politics can be taken a step further by explicitly incorporating the impact of past perceptions (reputation) on subsequent signaling behavior. The need for nations to signal their intentions depends on what other nations already believe. If the information available to states varies in some systematic way, then so should signaling behavior.

2.1 Signaling

Informational asymmetries seem endemic in international affairs, in part because nations, like poker players, benefit from concealing both weakness and advantage, but also because of (exogenous) change and because the participants themselves may have doubts about their ideal actions. Jervis (1970) suggests that how states attempt to communicate has important empirical consequences. Formal theorists have begun to integrate psychological insights into the rationalist paradigm. In particular, the focus has been on strategic manipulation of beliefs. Studies have shown how signaling can lead to variation in dispute propensity by regime type (Fearon 1994a; Schultz 1998, 1999, 2001a), how interdependence can facilitate non-violent costly signaling (Gartzke, et al. 2001), how initial stages of a war act as a costly signal of resolve (Wagner 2000), how general deterrent threats shape subsequent challenges (Fearon 1994*b*), and how arms races inform competitors (Kydd 1997).

Given its centrality to international politics, it is tempting to seek to integrate reputation into dynamic models of cooperation and conflict. To some extent, this is beginning to occur. Sartori (2002, 2005) shows how reputation can make cheap talk diplomacy credible. As Nalebuff (1991) points out, however, reputation can have counter-intuitive effects, including providing incentives for strong types to feign weakness. States perceived as capable or resolved are less likely to be challenged and so have additional incentives to bluff. One remedy is to treat states as having reputations for strength or weakness largely only in the context of specific counterparts or specific issues or territory. The fact that Taiwan is seen as willing to defend itself against invasion from China may or may not mean that it is more likely to be believed should it assert interests in Central Africa. France and the United States may or may not be perceived as resolved in abstract, though France clearly lacks the same interest in defending Taiwan as does the United States.

Efforts to understand reputation are hampered by considerable conceptual ambiguity, even as the effects of reputation are difficult to observe. An absence of evidence makes it difficult to discipline speculation or inspire new insights, preventing the virtuous feedback that facilitates cumulation of knowledge. Precisely because reputation constitutes valuable information about partners or competitors, states will tend to anticipate, and often preempt, its effects. This difficulty in seeing reputation has even led to skepticism about its very existence. The research community will be best served by identifying those (relatively few) situations in which reputation should be observable. These can occur where different tendencies rub up against one another, forcing actors with different interests to respond to reputation differently and creating non-linear relationships that depart from the monotonic expectations of either preferences or uncertainty in isolation.

2.2 Models of Alliance Formation

A logic of alliance formation is a key dynamic of theories of international politics (Holsti, et al. 1973). According to Waltz (1979, page 125), alliances are attempts by nations to form "balances of power." States seek stability by tailoring commitments and pooling military resources to roughly equal the capabilities of other coalitions (Waltz 1979, pages 163-170). Walt (1987) sees states as focusing on threats rather than on capabilities. Christensen & Snyder (1990) extend Walt's theory, arguing that a perceived offensive advantage led to balancing on the eve of World War I, while a defensive advantage caused major powers to "pass the buck" prior to the Second World War.

Contemporary research on alliances has sought out the determinants of alliance duration (Bennett 1997), alliance reliability (Leeds, et al. 2000 Leeds 2003), the effects of regime type (Siverson & Emmons 1991, Gaubatz 1996, Simon & Gartzke 1996, Leeds 1999, Lai & Reiter 2000), or combinations of these factors (Reed 1997, Gartzke & Gleditsch 2004). Improvements in data have challenged previous convictions and helped to answer old questions (Leeds et al. 2000; Gibler & Sarkees 2004), while also demonstrating that alliances have observable effects (Leeds 2003*b*).

Other research broadens the study of alliance choice beyond capability aggregation. Altfeld (1984) characterizes alliances as involving tradeoffs between security and autonomy. He notes that weak states do little to increase the capabilities of powerful allies, but the weak can offer powerful states access to territory and pliant foreign policies. Morrow (1991) claims that asymmetric alliances between security-seeking and autonomy-seeking states are easier to form and are more durable than symmetric alliances. Most alliances form between militarily weak and strong states, and these alliances are more durable. Conybeare (1992, 1994) points out that states can minimize alliance tradeoffs through a diversified portfolio of allies. Smith (1995), following Bueno de Mesquita (1981), claims that alliances form between states with similar preferences. He and others (Morrow 1994) argue that the cost of alliances mean that states ally only when they possess some degree of mutual interest. These insights were further refined by signaling, which I discuss in the next section.

3 Goldie Locks and the Three Alliance Theories

Theories of alliance formation can be related to the children's tale *Goldilocks and the Three Bears*. Early theories assumed that actors were fully informed about all relevant aspects of play. Full information precluded using uncertainty as a potential cause of alliances (reputation is "too hot"). Asymmetric information models allow actors to be uncertain about salient characteristics of partners or competitors (Fearon 1995). Yet, use of asymmetric information also assumes that states treat partners and competitors as largely unknown entities. Actors facing uncertainty are said to possess beliefs represented as probability distributions over the values of strategic variables. An opponent "type" is drawn at random from this distribution. This approach implies uncertainty that is probably excessive when one considers that states interact or at least co-exist over long periods of time (reputation is "too cold"). States are bound to accumulate perceptions. Asymmetric information models themselves imply that this must be so, as states update beliefs when interaction reveals new information. Modeling the effects of reputation requires that information is randomly drawn from samples with different means. States facing uncertainty and a heterogeneous history of play construct perceptions tailored to specific counterparts or circumstances. To the extent that perceptions are public or can be inferred from observable behavior, perceptions become common conjecture and the basis for equilibria involving anticipated reactions (reputation is "just right").

Adding details to a causal argument invariably involves analytical costs. One potentially appealing compromise is to treat actors' beliefs as a summary of previous interactions. Rationality requires that perceptions be correct in equilibrium. Beliefs must generally be consistent with past behavior, though they can err in particular instances. Yet, if perceptions faithfully (if noisily) represent reality, why should perceptions be analytically interesting? In equilibrium, the stochastic component of beliefs must be randomly distributed about the "true" value while the non-stochastic component simply mirrors other variables. Given these conditions, research has concentrated on the dynamic process of informational change (signaling) and largely ignored ex ante beliefs.

The use of stochastic error in modeling uncertainty implies that the effect of beliefs are themselves stochastic. Informational models typically treat one actor as ignorant about some key attribute of another actor in the game. The sequence of play then affords the asymmetrically informed actor an opportunity to reduce the uncertainty of its counterpart through signaling. Yet, the value for the asymmetrically informed actor in seeking to manipulate the beliefs of an uninformed counterpart depends on that actor's ex ante beliefs. If players all have uniform ex ante perceptions, regardless of context and competitor, or if the asymmetrically informed actor is ignorant about the perceptions of the uninformed counterpart, then signaling should occur as a monotonic function of power, interests, etc. If instead, actors' perceptions of play differs from dyad to dyad, then whether the informed actor benefits from signaling depends on the uninformed player's prior perceptions.

Noisy information should matter most in situations where behavior is primarily intended to influence beliefs. Whether a state engages in costly signaling depends on the cost and expected impact of the signal. If observers already possess perceptions that benefit the potential signaler's objectives, then the value in signaling must be less than if the target possesses beliefs harmful to the sender. I next turn to alliance formation to discuss reputation in a more concrete context.

3.1 Reputation and Alliance Formation

An important function of military alliances is to communicate intent. Alliances reassure partners and warn adversaries. In terms of the informational mechanism, states can "tie hands" or "sink costs" (Fearon 1997). "Tying hands" addresses the commitment problem posed when partners face incentives to abandon alignments in times of crisis. Alliances bind states ex ante to costly behaviors ex post by linking a failure to intervene to some punishment. Alliances can also "sink costs." Costly formal commitments signal the willingness of states to intervene to assist partners.

These two logics of alliance signaling yield a unique implication of the informational approach. If there is some stability in national preferences, and if past behavior is at all indicative of interests, states believed to possess both common interests and abundant capabilities should be less willing to sink costs or tie hands. It is not enough that states are "friends" or that they have motives to cooperate. States can cooperate informally. As formal commitments to provide military assistance, or to coordinate in times of crisis, alliances are needed when informal alignments are inadequate, where doubt exists about the commitment of states to intervene. Alliances prove most valuable as signals where states have a particular need to warn or reassure. If it is already understood that one country will intervene to protect another—if an alignment is common conjecture—then signaling is redundant. Alliances will still form for other purposes (such as commitment or coordination), but the tendency to ally to signal common interests will be lost where affinity is already inferred.

Informational arguments in particular depend on reputation to motivate alliance formation, alliance reliability, and deterrence. States are said to use alliances to demonstrate a willingness to intervene, or to ensure reputational consequences for abandoning an ally (Morrow 1994). Failure to come to a partner's aid is thought to leave a state with a reputation as unreliable. Unreliability in turn makes it harder to be believed in subsequent communication, raising the country's costs for conducting an effective foreign policy. Some states may thus come to the assistance of partners in spite of incentives not to intervene. While the argument helps to explain the role of formal commitments in international politics, it is not clear why an alliance is assumed to be the *only* means by which states generate credibility. States often appear to find that their reputations are tied to intervention, even without formal commitment. Countries also appear to make informal promises to other states (Sartori 2005). If additional confidence building measures are needed, then they should be most necessary when verbal commitments are least likely to be believed.

If alliances involve some positive cost, then only states with similar security interests will be willing to ally. The notion that "friends" co-ally (Smith 1995) is incomplete, however, since this considers the cost of allying, but not the informational benefit. Potential aggressors must have beliefs about whether certain states will assist one another *regardless of whether these states are actually allied*. States with similar interests are more likely to act in concert. The tendency for states with similar interests to protect one another is then a subject of common conjecture. Given some positive cost, alliances will be more valuable when intervention is in doubt. It is with states that would not normally be expected to offer protection that alliances prove most informative.

Actors facing uncertainty form beliefs that summarize available knowledge. The history of play forms the basis for perceptions that condition subsequent attempts to influence beliefs. States already seen as possessing similar interests have less need of demonstrating a willingness to intervene.[1] Subsequent attempts to signal through alliances should then be conditioned on the history of play. The next section provides two formal models of alliance signaling, a standard asymmetric information model, and one where actors possess "noisy" information about preferences.

4 Two Models of Alliance Signaling

The main flaw in existing signaling models of alliance formation is that they treat the international environment as informationally pristine. Actors are assumed not to have interacted previously in any manner that might be revealing (such as allying). To begin to address the effects of previous behavior, and thus reputation, on the decision to ally, the noisy signaling model includes an initial stage in which states may or may not interact through force.

Both models have most elements in common. There are four actors. Player $\boldsymbol{A}$ is the protector, who makes an alliance decision ($\boldsymbol{a}$ "alliance" or $\sim\boldsymbol{a}$ "no alliance"). $\boldsymbol{A}$ also decides whether to intervene ($\boldsymbol{i}$) or not ($\sim\boldsymbol{i}$) if player $\boldsymbol{B}$ is attacked. Player $\boldsymbol{B}$ is the potential ally. $\boldsymbol{B}$'s alliance decision is redundant. There is no equilibrium where $\boldsymbol{A}$ seeks to ally and $\boldsymbol{B}$ does not. Further, if $\boldsymbol{A}$

[1] Noisy information can lead high-attribute types to pool with low-attribute types (Feltovich, Harbaugh & To 2001).

does not seek to ally, ***B***'s choice is mute. I simplify the game by making ***B*** a non-strategic actor. Player ***C*** is the potential attacker. ***C*** uses force (***f***) or not (∼***f***). The fourth actor, Nature (***N***), determines ***A***'s ideal point. Actors differ over goods or issues on a single dimension of unit domain ($\mathbf{x}$, where $[0 \leq \mathbf{x} \leq 1]$). For tractability, I normalize actors' ideal points so that player ***B*** prefers an outcome of zero (0), ***C***'s ideal point is one (1) and ***A***'s ideal point ($\mathbf{x}_A$) is between zero (0) and one half ($\frac{1}{2}$) inclusive ($\mathbf{x}_A \sim$U$[0,\frac{1}{2}]$).[2] ***A*** at least weakly prefers ***B***'s ideal point to ***C***'s.[3] The *status quo ante*, $\boldsymbol{x}_{\mathbf{SQ}}$, can be anywhere in the issue space, but for tractability, I assume a value of zero.

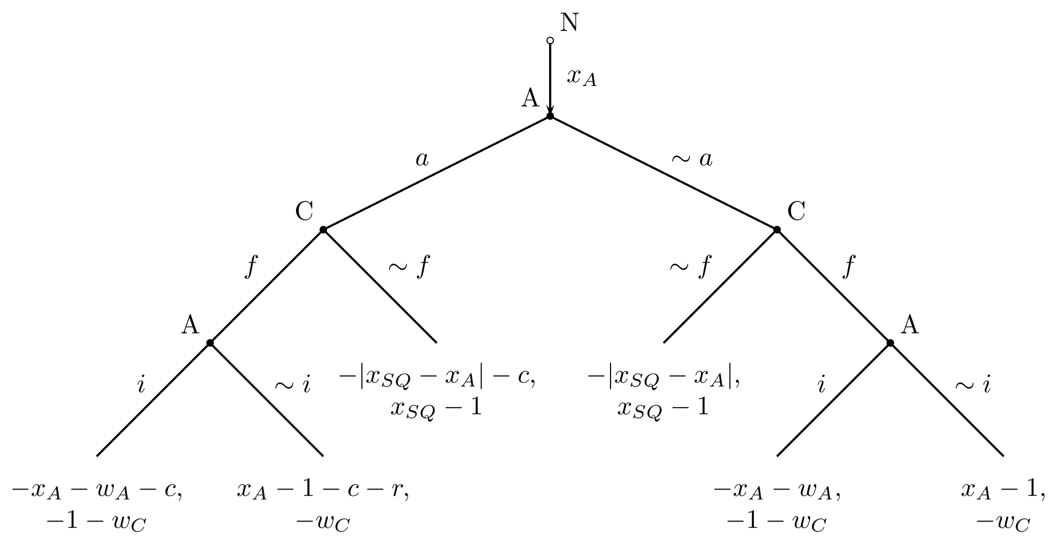

Figure 1: The Basic Alliance Game

Figure 1 represents the sequence of play in the basic alliance game. First, Nature (***N***) randomly assigns player ***A*** an ideal point. Second, ***A*** decides whether to ally with ***B*** (***a***) or not (∼***a***). Third, ***C*** decides to fight (***f***) or not (∼***f***). Finally, ***A*** intervenes (***i***) or not (∼***i***). The probability of victory for ***C*** is non-increasing in ***A***'s participation. In other words, ***C*** does better if it fights ***B*** alone than if it fights both ***A*** and ***B***. I assume that, if ***A*** intervenes after ***C*** has chosen to fight ***B***, ***B*** receives its preferred outcome (normalized to 0). Otherwise, ***C*** receives its preferred outcome.[4]

Alliances are costly. If ***A*** chooses to ally, then ***A*** and ***B*** pay some fee ***c*** (where $\boldsymbol{c} > 0$). The cost of fighting, ***w*** (where $\boldsymbol{w} > 0$), can differ for each actor. Reneging on commitments is also costly. ***A***

[2]Allowing ***B*** and ***C***'s ideal points to vary does not alter substantive conclusions, but does complicate the model.
[3]Pareto efficiency: for any $\acute{\mathbf{x}}$, $(-\infty \leq \acute{\mathbf{x}} < 0)$ or $(1 < \acute{\mathbf{x}} \leq \infty)$, there exists $(0 \leq \mathbf{x} \leq 1)$ *s.t.* ***B***($\mathbf{x}$ R $\acute{\mathbf{x}}$) and ***C***($\mathbf{x}$ P $\acute{\mathbf{x}}$).
[4]***A*** and ***B*** need merely exceed ***B***'s capabilities alone. Results are not changed by explicitly modeling the contest.

pays **r** (where $\mathbf{r} \geq 0$) if ***A*** abandons its ally. Figure 1 lists players' payoffs at each terminal node. Payoffs reflect outcomes and associated costs. Linear loss utility functions appear below:[5]

$$U_A = [(-|x - x_A|) - iw_A - a\,(c + (1 - i)\,r)] \tag{1}$$

$$U_B = [(-x) - ac - fw_B] \tag{2}$$

$$U_C = [(x - 1) - fw_C] \tag{3}$$

Payoffs are obtained by substituting 1 for affirmative decisions and 0 for negative decisions. For example, if player ***A*** allies (***a***) but does not intervene ($\sim\boldsymbol{i}$) while player ***C*** does not fight, ($\sim\boldsymbol{f}$), then players' utilities are as follows: $\mathrm{U}_A = [(-|x - x_A|) - c]$, $\mathrm{U}_B = [(-x) - c]$, $\mathrm{U}_C = [(x - 1)]$.

4.1 The Alliance Reputation Game with Asymmetric Information

Nature (***N***) first assigns ***A*** an ideal point ($0 \leq \mathbf{x}_A \leq \frac{1}{2}$), known only to player ***A***. ***C*** is said to possess beliefs (***b***) about **A**'s "type," which are in turn subject to common conjecture (***A*** can calculate ***C***'s beliefs, ***C*** knows that ***A*** knows, and so on). However, ***C*** makes its fight decision unaware of ***A***'s ideal point. A perfect Bayesian equilibrium requires that players' strategies be sequentially rational and that players update beliefs using Bayes' rule where possible.

Players' optimal strategies and the equilibria for the game appear in the appendix. Player ***A*** intervenes if the payoff for intervention exceeds war costs (w_A), ($x_A < \frac{1}{2}\,(1 - w_A + r)$, if allied, and $x_A < \frac{1}{2}\,(1 - w_A)$, if not). ***C*** knows ***A***'s decision rule, but ***C*** does not know ***A***'s type ($\boldsymbol{x}_{\mathbf{A}}$). Player ***C*** calculates its best response to each potential type ***A***, weighted by the probability of encountering a particular type. Given a uniform distribution, the probability of any behavior is just the proportion of types that will choose a given action in a given context. ***C*** fights if its expected utility for fighting exceeds its utility for the *status quo*, $\mathrm{U}_C^{\sim f} < EU_C^f = U_C^{f,i} \cdot p^i + U_C^{f,\sim i} \cdot (1 - p^i)$, where p^i is the proportion of players' ***A*** that intercede. ***C***'s decision rule is thus ($w_C < (1 - x_{SQ}) - p^i$).

In the no alliance subgame, no type ***A*** intervenes if war costs are high ($w_A \geq 1 \Rightarrow \boldsymbol{i} = 0$ $\forall\ x_A$), and all types ***A*** intervene ($\boldsymbol{i} = 1$) when war is costless. The probability that ***A*** intervenes if $0 < w_A < 1$ equals the portion of types ***A*** willing to intervene. In the no alliance subgame, ***C*** expects ***A*** to intervene with probability $p^i = (1 - w_A)\,, (0 \leq w_A \leq 1)$. ***C*** fights if $w_C < w_A - x_{SQ}$.

[5] Claims about non-linear utilities cannot be evaluated using ordinal alliance or dispute onset data (O'Neill 2001).

However, some types $\boldsymbol{A}$ exit the sample by allying. $\boldsymbol{C}$'s rule appears as equation (4), where ($\boldsymbol{t}$, $0 \leq \boldsymbol{t} \leq \frac{1}{2}$) is the type $\boldsymbol{A}$ that is just indifferent between allying and not allying. If all types of $\boldsymbol{A}$ ally, $\boldsymbol{C}$ fights if $w_C < w_A - x_{SQ} - r$. Equation (5) describes $\boldsymbol{C}$'s decision rule for the alliance subgame.

$$w_C < (1 - x_{SQ}) - p_A^{\sim a,i},\ p_A^{\sim a,i} = 1 - \left(\frac{w_A}{1-2t}\right),\ s.t.\ 0 \leq p_A^{\sim a,i} \leq 1,\ 0 \leq t < \frac{1}{2}\ ^{6} \tag{4}$$

$$w_C < (1 - x_{SQ}) - p_A^{a,i},\ p_A^{a,i} = \left(\frac{1 - w_A + r}{2t}\right),\ s.t.\ 0 \leq p_A^{a,i} \leq 1,\ 0 \leq t < \frac{1}{2} \tag{5}$$

$\boldsymbol{A}$ allies if $EU_A^a > EU_A^{\sim a}$. Equation (6) describes $\boldsymbol{A}$'s alliance calculation where $\boldsymbol{p}$ is the probability $\boldsymbol{C}$ does not fight, $\boldsymbol{q}$ is the probability that $\boldsymbol{C}$ fights and $\boldsymbol{A}$ intervenes and (1-$\boldsymbol{p}$-$\boldsymbol{q}$) is the probability that $\boldsymbol{C}$ fights and $\boldsymbol{A}$ does not intervene ($\boldsymbol{A}$ and $\boldsymbol{C}$ play pure strategies):

$$\begin{aligned} &U_A^{a,\sim f} \cdot p^a + U_A^{a,f,i} \cdot q^a + U_A^{a,f,i} \cdot (1 - p^a - q^a) > \\ &\quad U_A^{\sim a,\sim f} \cdot p^{\sim a} + U_A^{\sim a,f,i} \cdot q^{\sim a} + U_A^{\sim a,f,i} \cdot (1 - p^{\sim a} - q^{\sim a}) \end{aligned} \tag{6}$$

Alliance costs mean that $\boldsymbol{A}$ only allies when there is a positive expected benefit of the alliance. If $\boldsymbol{A}$ intervenes in both subgames, then $\boldsymbol{A}$ prefers not to form an alliance, since $c > 0 \Rightarrow U_A^{a,f,i} < U_A^{\sim a,f,i}\ \forall x_A,\ c, w_A$. Three sets of conditions satisfy equation (6), so that an alliance is formed:

$$\begin{aligned} &U_A^{a,\sim f} > U_A^{a,f,i} \Rightarrow (-\left|x_{SQ} - x_A\right| - c) > -x_A - w_A; \\ &\quad \text{if } x_{SQ} \leq x_A \Rightarrow c < (x_{SQ} + w_A), \\ &\quad \text{or if } x_{SQ} > x_A \Rightarrow x_A > \frac{1}{2}(x_{SQ} - w_A + c) \end{aligned} \tag{7}$$

$$U_A^{a,f,i} > U_A^{\sim a,f,\sim i} \Rightarrow (-x_A - w_A - c) > (x_A - 1) \Rightarrow x_A < \frac{1}{2}(1 - w_A - c) \tag{8}$$

$$\begin{aligned} &U_A^{a,\sim f} > U_A^{\sim a,f,\sim i} \Rightarrow (-\left|x_{SQ} - x_A\right| - c) > (x_A - 1); \\ &\quad \text{if } x_{SQ} \leq x_A \Rightarrow x_A < \frac{1}{2}(1 + x_{SQ} - c), \\ &\quad \text{or if } x_{SQ} > x_A \Rightarrow c < (1 - x_{SQ}) \end{aligned} \tag{9}$$

[6] $p^{\sim a,i} = \left[\frac{\frac{1}{2}(1-w_A)-t}{\frac{1}{2}-t}\right] = 1 - \frac{w_A}{(1-2t)};\ p^{a,i} = \left[\frac{\frac{1}{2}(1-w_A+r)}{t}\right] = \frac{(1-w_A+r)}{2t}$

If equation (7) holds, then $x_A < \frac{1}{2}(1 - w_A)$ (because of $U_A^{\sim a,f,i}$). If $x_{SQ} \leq x_A$, then $c < (x_{SQ} + w_A)$. This means that, $c - x_{SQ} < w_A < (1 - 2x_A) \Rightarrow c - x_{SQ} < (1 - 2x_A) \Rightarrow x_A < \frac{1}{2}(1 + x_{SQ} - c)$. Thus, for all types $\mathbf{A}$ for which $x_{SQ} \leq x_A$ and $x_A < \frac{1}{2}(1 - w_A)$, and $c < (x_{SQ} + w_A)$, equation (7) is a special case of equation (9). Similarly, equation (9) is a special case of equation (7) if $x_{SQ} > x_A$. $U_A^{\sim a,f,\sim i}$ implies $x_A \geq \frac{1}{2}(1 - w_A)$ and $x_{SQ} > x_A$ and $U_A^{a,\sim f} > U_A^{\sim a,f,\sim i}$, implies $c < (1 - x_{SQ})$. So, $c < (1 - x_{SQ}) \Rightarrow 1 > x_{SQ} + c \Rightarrow 1 - w_A > x_{SQ} - w_A + c$, and $x_A \geq \frac{1}{2}(1 - w_A) \Rightarrow x_A \geq \frac{1}{2}(1 - w_A) > \frac{1}{2}(x_{SQ} - w_A + c) \Rightarrow x_A > \frac{1}{2}(x_{SQ} - w_A + c)$. Finally, $U_A^{a,f,i} > U_A^{\sim a,f,\sim i}$ implies $x_A < \frac{1}{2}(1 - w_A - c)$ but $U_A^{\sim a,f,\sim i} \Rightarrow x_A < \frac{1}{2}(1 - w_A)$. Since, by definition, $(1 - w_A - c) < (1 - w_A)\ \forall\ c$, there are *no* values of x_A such that equation (8) holds true. For every combination of x_{SQ} and x_A, there is a unique type (t) such that $\boldsymbol{A}$ is just indifferent between subgames if $t = \frac{1}{2}(1 + x_{SQ} - c)$ and $x_{SQ} \leq x_A$ or if $t = \frac{1}{2}(x_{SQ} - w_A + c)$ and $x_{SQ} > x_A$. Put a bit more prosaically, $\boldsymbol{A}$ tends to ally when x_A is close to zero ($\boldsymbol{B}$'s ideal point). "Friends" co-ally.

4.2 The Alliance Reputation Game with "Noisy" Information

Assumptions of the second game are the same as in the standard asymmetric information game, except that Nature now publicly reveals noisy information about $\boldsymbol{A}$'s ideal point. Imagine that past action or common knowledge has lead to some differentiation among types of player $\boldsymbol{A}$. States are now said to have reputations that affect other players' expectations about $\boldsymbol{A}$'s type.

Nature first identifies the range $\boldsymbol{x}_\mathbf{N}$ ($0 \leq \boldsymbol{x}_\mathbf{N} \leq \frac{1}{2}$ - ϵ) to $\boldsymbol{x}_\mathbf{N} + \epsilon$ ($\epsilon \leq \boldsymbol{x}_\mathbf{N} + \epsilon \leq \frac{1}{2}$), where $0 \leq \epsilon \leq \frac{1}{2}$. Nature then randomly selects $\boldsymbol{A}$'s ideal point from this domain ($\boldsymbol{x}_\mathbf{A} \in \boldsymbol{x}{\sim}\text{U}[\boldsymbol{x}_\mathbf{N}, \boldsymbol{x}_\mathbf{N} + \epsilon]$). $\boldsymbol{C}$ thus has imprecise, but better, information about what $\boldsymbol{A}$ wants.

Once again, players evaluate strategies in terms of payoffs and the sequence of play. $\boldsymbol{A}$'s decision to intervene is the same. Given an alliance, $\boldsymbol{A}$ intervenes if $x_A < \frac{1}{2}(1 - w_A + r)$ and $\boldsymbol{A}$ intervenes with no alliance if $x_A < \frac{1}{2}(1 - w_A)$. $\boldsymbol{C}$'s fight decision rule is nominally, $(w_C < (1 - x_{SQ}) - p^i\ )$, but this is modified by the domain for $\boldsymbol{A}$'s type. Since $\frac{1}{2}(1 - w_A + r) \geq \frac{1}{2}(1 - w_A)$ and $(\boldsymbol{x}_\mathbf{N} + \epsilon)$ $\geq (\boldsymbol{x}_\mathbf{N})$, six sets of conditions relate Nature's revelation and other parameters to players' actions. The first condition set is when all types $\boldsymbol{A}$ permitted by Nature's revelation intervene, regardless of alliance status, since $(\boldsymbol{x}_\mathbf{N} + \epsilon) < \frac{1}{2}(1 - w_A) \leq \frac{1}{2}(1 - w_A + r)$. Given these conditions, $\boldsymbol{C}$ prefers not to fight. The second condition set occurs when no player $\boldsymbol{A}$ intervenes, regardless of alliance status. If all types of player $\boldsymbol{A}$ prefer not to intervene, then $\boldsymbol{C}$ fights if the cost of fighting is less than the difference between $\boldsymbol{C}$'s ideal point and the *status quo*, $w_C < (1 - x_{SQ})$ (because $p^i = 0$).

In the four remaining condition sets, player $\boldsymbol{A}$'s decision is contingent on type. In the third condition set, no types of $\boldsymbol{A}$ intervene without an alliance, but all types intervene with an alliance. If $\frac{1}{2}(1 - w_A) \leq (\boldsymbol{x_N})$ and $(\boldsymbol{x_N} + \epsilon) < \frac{1}{2}(1 - w_A + r)$, then $\boldsymbol{C}$ fights if there is no alliance and $w_C < (1 - x_{SQ})$. The last three condition sets require $\boldsymbol{C}$ to rely on probability. $\boldsymbol{C}$'s decision problem is analogous to $\boldsymbol{C}$'s problem in the asymmetric information game, except that more is known about $\boldsymbol{A}$'s type within the space. In the fourth condition set, only some types $\boldsymbol{A}$ permitted by Nature's revelation intervene with an alliance. No types intervene without the alliance. Without an alliance, since $\boldsymbol{A}$ with not intervene, $\boldsymbol{C}$ fights if $w_C < (1 - x_{SQ})$. If $\boldsymbol{A}$ allies, $\boldsymbol{C}$ knows that $x_A < t = (1 - |x_{SQ} - x_A| - c)$ and fights if $w_C < (1 - x_{SQ}) - p^i$, where p^i equals the portion of types allowed by Nature's revelation that ally and intervene $\frac{1}{2}(1 - w_A + r) - (x_N)$, divided by the portion of types that ally $t - (x_N + \epsilon)$. Given an alliance, $\boldsymbol{C}$ fights if equation (11) holds.

In condition set 5, $\boldsymbol{A}$ always intervenes with an alliance and sometimes intervenes without an alliance. $\boldsymbol{C}$ never fights if $\boldsymbol{A}$ and $\boldsymbol{B}$ ally. Without an alliance, $\boldsymbol{C}$ determines the probability that $\boldsymbol{A}$ will intervene as the portion of types in Nature's revelation that intervene $(\frac{1}{2}(1 - w_A) - t)$, divided by the portion of types that do not ally $((x_N + \epsilon) - t)$. $\boldsymbol{C}$ fights if equation (10) holds. The last set of conditions simply combines the rules in condition sets 4 and 5. Without an alliance, $\boldsymbol{C}$ fights if equation (10) holds. Given an alliance, $\boldsymbol{C}$ fights provided equation (11) is true.

$$w_C < (1 - x_{SQ}) - p_A^{\sim a,i}, p_A^{\sim a,i} = 1 - \left(\frac{(x_N + \epsilon) - \frac{1}{2}(1 - w_A)}{(x_N + \epsilon) - t} \right),$$
$$s.t.\ 0 \leq p_A^{\sim a,i} \leq 1, x_N \leq t < (x_N + \epsilon)^7 \tag{10}$$

$$w_C < (1 - x_{SQ}) - p_A^{a,i}, p_A^{a,i} = \left(\frac{\frac{1}{2}(1 - w_A + r) - (x_N)}{t - (x_N)} \right),$$
$$s.t.\ 0 \leq p_A^{a,i} \leq 1, x_N \leq t < (x_N + \epsilon) \tag{11}$$

A allies with ***B*** if $EU_A^a > EU_A^{\sim a}$. Given the discussion in the asymmetric information game, we know that $EU_A^a > EU_A^{\sim a}$ if $x_A < t$ (where $t = \frac{1}{2}(1 + x_{SQ} - c)$ if $x_{SQ} \leq x_A$ and $t = \frac{1}{2}(x_{SQ} - w_A + c)$ if $x_{SQ} > x_A$) and only if ***C*** fights without an alliance but not with an alliance.

Noisy information leads part of the alliance equilibrium to be replaced by the *status quo* equilibrium [Node 4]. While lacking precise knowledge of ***A***'s type, ***C*** sometimes learns enough to infer that all available types ***A*** will intervene. States with the most similar preferences do not need to co-ally. If war costs are moderate, states at either extreme of the issue space do not ally. Alliances between "friends" occur less often than alliances between states with less similar interests. In contrast, high war costs can force even friends to question intervening. The effect of noisy signaling should diminish for states facing major threats, multiple capable opponents, or a nuclear adversary.

Similar preferences imply alignments. Alignments may in turn invite alliances if intentions are underestimated and must be advertised. If on the other hand, states have reputations as "friends," alignments are the subject of common conjecture, and signaling affinity by allying is not necessary.

5 Tests of Reputation and Alliance Formation

This section assesses the external validity of the reputational argument. I first describe the hypotheses and data before detailing the research design and reporting the statistical findings.

5.1 Hypotheses

Figure 2 describes two conceptions of the relationship between preferences and alliance formation. The horizontal axis measures preference similarity while the vertical axis indicates the probability of alliance. The preferences argument from the basic signaling model anticipates a positive monotonic

[7] $p^{\sim a,i} = \left[\frac{\frac{1}{2}(1-w_A)-t}{(x_N+\epsilon)-t}\right] = \left[\frac{\frac{1}{2}(1-w_A)-(x_N+\epsilon)}{(x_N+\epsilon)-t}\right] + \left[\frac{(x_N+\epsilon)-t}{(x_N+\epsilon)-t}\right] = 1 - \left[\frac{(x_N+\epsilon)-\frac{1}{2}(1-w_A)}{(x_N+\epsilon)-t}\right]$

relationship between preference similarity and alliance formation. "Friends" form alliances. The reputational hypothesis resulting from noisy information claims that alliance formation is a function of both preferences and reputation. Similar interests encourage alignments, but common conjecture about similar interest reduces the need to ally. The second model predicts a non-linear relationship between similarity and alliance formation. Multiple allies relax these tendencies by degrees.

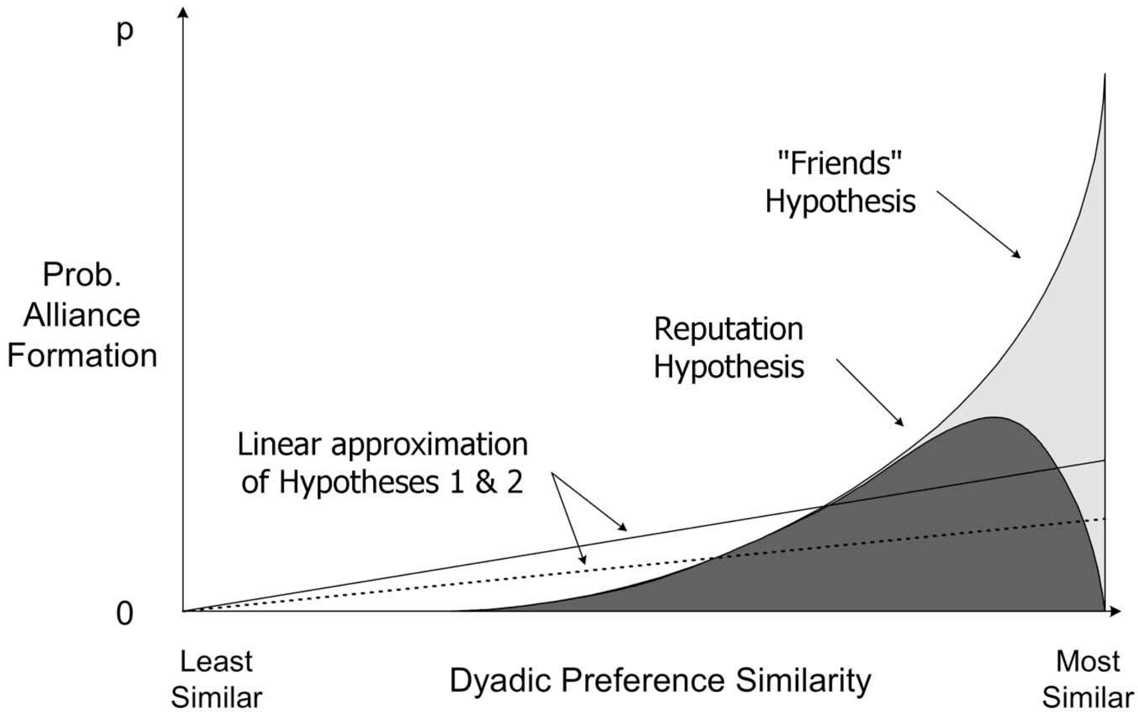

Figure 2: Hypothesized Relationships Between Preference Similarity and Alliance Formation.

Hypothesis 1 *"Friends": The probability of allying increases in the similarity of dyadic interests.*

Hypothesis 2 *Reputation: The probability of allying first increases and then declines in the similarity of dyadic interests.*

Hypothesis 3 *"Threat × Reputation": The effect of reputation in decreasing alliances among "friends" diminishes in the size of the threat (insecurity tends to weaken the Reputation hypothesis).*

Note that linear estimates of the hypotheses will be very similar, allowing the arguments to differ only in magnitude. The "friends" hypothesis anticipates a positive monotonic relationship while reputation implies a second or third-order function where the highest term is negative. While the overall relationship is positive, the function is initially increasing and later decreasing, implying a non-linear relationship. I use two estimators, logit with multiple preference terms and the General Additive Model (GAM). Before proceeding, I review releant data and discuss the research design.

5.2 Data

With the exception of the dependent variable, the Correlates of War (COW) alliance data, and the index of affinity, most data in the analyses are generated using EUGene (Bennett & Stam 2001). The unit of analysis is the dyad year. The spatial-temporal domain involves all pairwise combinations of states (dyads) listed by the COW Inter-state System Membership Dataset for the period 1946–1996.

The dependent variable comes from two versions of the COW Alliance Data (Small & Singer 1990). The COW project code alliances for the period 1815 to 1984. I rely on two "unofficial" sources, though I also verify my findings using the official dataset.[8] First, Bennett and Lebrun offer a revised version of the COW data, updated to 1992. Though newer than the official release, I refer to these as the "old" alliance data. Second, the COW project is revising the alliance data. I use a "beta" version of these data (the "new" alliance data) with alliances coded to 1996.[9] COW data code three types of commitment. I report results using both all alliances and defense pacts.[10]

Initial alliance years are coded "1." Subsequent years of alliance and dyad years without alliances are coded "0." I lead the dependent variable (alliance onset) on the independent variables by placing alliance values at time $t-1$ in the subsequent year (time t). Thus, all of the independent variables are lagged on the dependent variable, limiting endogeneity and providing more stringent tests of the hypotheses. Alliances involve four phases: formation, maintenance, termination, and non-alliance. The hypotheses refer to formation and non-alliance. Factors leading states to ally may differ from those encouraging states to maintain existing commitments (Bennett & Tarry 1996). The distinction between subsequent alliance dyad years and no-alliance dyad years poses problems

[8]Leeds, et al. (2000) offer alliance data with important advances in coding. Unfortunately, there is currently no overlap between these data and the affinity measure used here.

[9]Use of the beta version of the "new" alliance data has not been officially authorized by the COW project, but the dataset has been circulating informally for several years.

[10]The argument does not apply to non-aggression pacts (sample does not affect results).

in estimation. I analyze a sample of the population of all dyad years composed of all initial alliance years and randomly selected cases of non-alliance years (Siverson & Starr 1994). The approach offers several advantages. First, sampling reduces temporal dependence, since observations are no longer temporally linked (Beck, Katz & Tucker 1998). Second, sampling reduces the inflation of significance tests that occurs in large samples (I also use Huber/White robust estimates and correct for dyadic clustering). One concern when sampling is that it can bias coefficient estimates (King & Zeng (2001*a*, 2001*b*)). I sample alliance onset and a small portion of non-alliance years equal to the percentage of total alliance years in the population (8.31% for the old alliance data, 5.46% in the new data) so that the probability of a dyad containing an alliance in the samples is comparable to that in the population.[11] Finally, GAM is computationally intensive. I need to analyze a smaller number of cases than the population of post-World War II dyad years. I use identical samples for each estimator to facilitate comparison and to corroborate results.

The analysis requires a measure of state preference similarity. Preferences are a theoretical construct related to "utility" in rational theory. Preferences are not observable, but proxies exist that arguably capture some of what preferences represent. I examined two different measures in preliminary research, though I report results only from the AFFINITY index. Bueno de Mesquita's alliance portfolio index (τ_B) is available for the entire period covered by the alliance data (Bueno de Mesquita 1981, pages 109–118) (Bueno de Mesquita & Lalman 1992, pages 286-294).[12] The alliance portfolio measure is widely used and referenced. Nevertheless, alliance ties may constitute a biased indicator of the similarity of state preferences. First, the cost of outcomes distorts observers' perceptions of actors' preference orderings. One may prefer Ferraris to Fords, but the differential cost of exercising this preference means that observers will not know whether observing a subject at the Ford dealership is indicative of ranking or of budget constraint. *Ceteris paribus*, a preference index based on costly behavior biases toward inexpensive acts. Second, if the perceptions argument posed here is correct, then a preference indicator based on alliance patterns bias tests of the hypothe-

[11] I over-sample non-alliance cases to compensate for heavier listwise deletion of non-alliance dyad years. Results using all dyad years and coding for alliance status are basically the same.

[12] The τ_B measure is not a utility index. A utility function maps an actor's valuation over outcomes. τ_B and AFFINITY relate two actors' choice of outcomes to outcomes arrived at by other pairings, an interpersonal comparison that cannot recover individual utilities. Note also that the construction of AFFINITY differs from τ_B (Bueno de Mesquita 1981).

ses. Alliances indicate preferences but also highlight shortcomings in international perceptions of preferences. Allies are not necessarily states with the most similar preferences so that a measure based on which states ally may not accurately attribute states' interests. Third, using alliance ties to predict alliance formation appears empirically problematic.[13]

For these reasons, I adopt a measure of preference similarity based on roll-call voting in the United Nations General Assembly. AFFINITY uses Signorino and Ritter's "S" indicator to calculate the association between states' annual voting records in the UN (2001).[14] Use of UN voting constrains analysis to the period 1946-1996, respectively the inaugural year of the United Nations and the most recent year for which data has been coded. Raw UN data code votes as positive, negative, absent, abstain, or present and non-voting. One version of the index uses only "yes" or "no" votes. Another version codes absences, abstentions and non-votes as a category between positive and negative votes. I examined both versions of the index, but report results for only the latter, since results are equivalent. Values of AFFINITY range from –1 (least similar interests) to 1 (most similar interests). Finally, I add 1 and divide by 2 to produce a scale of 0 to 1, identical to that of the formal models.

Other independent variables seek to address competing influences on state alliance decisions. DISTANCE reports the metric distance between state capitals, or between major cities for large countries. I log the variable, replacing negative values with "0." POL_REL codes for dyadic contiguity. States that border one another or that are closer than 150 miles by sea are coded as "1", otherwise "0." Power is an important potential determinant of alliance formation. Capable states can overcome distance while unequal capabilities in a dyad encourage patronage. EUGene codes COW Composite Index of National Capability scores (CINC) for all states for which there exists data. I recode CINC scores into two dyadic variables representing higher (CINCHI) and lower (CINCLO) values. COW also provides a list of major power states. Some argue that major powers behave differently from other states. I construct dummy variables for the presence of one (MAJPOWR1) or two major powers (MAJPOWR2) in the dyad. Systems arguments also suggest that alliance ties

[13] Construction of τ_B removes much of the correlation with individual alliances. Both indices (τ_B and AFFINITY) produce similar results, though the latter has greater variance.

[14] $S = 1 - 2*(\boldsymbol{d}/\boldsymbol{dmax})$, where $\boldsymbol{d} = \sum$(metric distance between votes) and $\boldsymbol{dmax}$ = the largest value of $\boldsymbol{d}$.

are more likely to change when international system structure changes. MAJCHANG is a dummy variable equal to “1” for a five year period (two year lead, two year lag) surrounding any change in the number of major powers, and “0” otherwise. The coding of MAJCHANG is admittedly arbitrary, but the variable seems to capture the effects of systemic change.[15]

There is considerable interest in regime effects in the literature. Regime type may influence states’ alliance decisions (Gaubatz 1996, Leeds 1999, Reed 1997, Simon & Gartzke 1996, Siverson & Emmons 1991). Using Polity data, I construct variables for higher and lower regime scores in the dyad, where REGIME = (DEMOC − AUTOC + 10)/2.[16]

Finally, I include variables to control for the effects of war on alliance formation. Wars can influence states’ alliance decisions in at least three ways. First, contests within the dyad arguably affect alliance ties. I use a dummy variable (WAR) to code for the presence or absence of war in a dyad year (1999). I lag values five years to capture follow-on effects. Second, contests involving one, but not both, states in a dyad could induce or deter alliances before or after contests. I construct leading and lagging dummy variables using Militarized Interstate Dispute data (MIDs) to attempt to capture confounding third-party effects. I remove disputes involving both dyad members. The resulting third-party variables equal “0” if neither state in a dyad experiences a MID, “1” if one state experiences an outside MID, and “2” if both states participate in (separate) disputes. Finally, conflict external to a dyad might spread. Contagion effects could lead states to be more aggressive in seeking alliance partners (Siverson & Starr 1990, Siverson & Starr 1991). Alliance formation may be affected by expectations about future conflict with other states. I therefore include a dummy variable for the current year and a lagged year to identify whether MID level 5 events (“wars”) precipitate alliance formation. While certainly not exhaustive, this list of variables addresses major influences on alliance behavior beyond interests.

5.3 Analysis

Tests in social science are seldom irrefutable. Assessing signaling theories is doubly difficult because the process is unobservable (Schultz 2001*b*). One cannot see communication. At best, one can

[15]Other measures of structural change yielded no noticeable alteration in key results.

[16]The variable differs from of Oneal and Russett (1997) in that it ranges from 0 to 10.

identify unique implications of signaling. The relationships I examine are bound to be muted by the nature of the tests. Nevertheless, key findings appear to be unanticipated by alternative explanations even as the results appear robust. I offer several analyses to reduce the possibility that the results are due to a particular method or approach.

Table 1 lists multivariate logit estimates of six models of dyadic alliance formation.[17] Models 1 through 3 include alliance formation dyad years and a random sample of dyad years without alliance. The dependent variable in models 4 through 6 includes only defense pacts. Models 1, 4, and 6 use the "old" alliance data while models 2, 3, and 5 report results using the "new" alliance data. The size of each sample is determined by first calculating the proportion of alliance dyad years in the population. Second, I randomly select cases from the population of dyad years that are not alliance dyad years. For example, model 1 uses the old alliance data. There are 38,773 alliance dyad years out of 466,770 dyad years (or 8.31% of the sample) between 1946 to 1992. After dropping cases of subsequent alliance dyad years, I randomly select 8.31% of the non-alliance dyad years or 30,039 cases and add the 995 initial alliance formation cases. This process is repeated with the new alliance data.[18]

I use Huber/White estimates of variance to compensate for panel effects in cross-sectional data. I also control for clustering in dyads. Temporal dependence is less of a problem (and less easily remedied) than in other cross-sectional time-series analyses (Beck, Katz & Tucker 1998). Sampling interrupts periodicity in the time-series. Consecutive cases are separated by an average of 12 and 18.3 years, for old and new samples respectively.

Construction of variables representing state preference similarity is central to the analysis. I use LINEAR, SQUARE and CUBE functions of AFFINITY. Again, the baseline "friends" hypothesis (H1) suggests that preference similarity is positively and monotonically related to the likelihood of alliance formation. Second and third power exponents of AFFINITY should be positive or insignificant. The perceptions hypothesis (H2) implies a non-monotonic function—one that is first

[17] Results are representative of those examined in preparing this study. The new alliance data codes for wartime alliances. Using only "peacetime" alliances does not alter results.

[18] The approach slightly over samples non-alliance dyad years (0's) to compensate for the higher proportion of non-alliance dyad years deleted listwise in the regressions. I replicated model 1, Table 1 using a true representative sample and found comparable results.

positive-increasing, then positive-decreasing. In both two term and three term models, the highest-power coefficient is expected to be negative.

Models 1 through 6 in Table 1 estimate the probability of alliance (or defense pact) formation. In Model 1, the coefficient for the linear term measuring affinity is positive and significant while the squared term is significant and negative. This is also true for Model 2 with the "new" alliance data. Model 3 estimates three coefficients for preference similarity. The three-term model "over-fits" the data. Still, the highest power coefficient is negative. Models 4 through 6 report equivalent results for defense pacts. Preference similarity seems to have a specific non-linear effect on the probability of alliance formation.

Most of the other independent variables perform as expected. DISTANCE is significant and negative; the more proximate two states, the more likely they are to ally. The more democratic the dyad, the more likely the states are to ally. The greater the difference in regime types, the less likely is an alliance. LOWER DYADIC CAPABILITIES is negative but is at times only marginally significant. The more powerful the less powerful state in the dyad, the less likely is an alliance. HIGHER DYADIC CAPABILITIES shows that larger disparities in capabilities are associated with a significant increase in the probability of alliance formation. The presence of a MAJOR POWER in the dyad significantly increases the likelihood of an alliance. Two major powers in a dyad further increase the chances of alliance formation. Dyadic WAR reduces the probability of an alliance, but this is only significant in the old alliance data. Third party effects are generally insignificant, though THIRD PARTY LAG is significant and negative when predicting defense pacts. States appear less likely to form defense pacts if they suspect a looming conflict. SYSTEM war does not appear significant. Political relevancy alone offers a surprise. POL. REL. DYAD is negative and significant.

Figure 3 offers six graphs, one each for the six models in Table 1. Predicted values are generated using the method of recycled predictions (StataCorp 1995). I replace all values of the preference variables in the data with a given value (say 0). Values for all independent variables are then run back through the coefficients to generate the predicted probabilities. Results are stored as a new variable and the process is repeated with another substitute value of the preference variables (say 0.05). Mean predicted probabilities are plotted in Figure 3. All alliance types appear in the left column while defense pacts are on the right. The top two rows of plots use a linear and a squared term. The bottom row lists results with three terms (linear, square, and cube). All plots appear to support the perceptions hypothesis.

Another way to examine the argument is to note that states are most likely to err in their perceptions of interest when interests are in flux. The perceptions argument emphasizes the discrepancy between state interests and perceptions of interest. Observers may seek to extrapolate from current relations. Extrapolations are bound to err in proportion to the dynamism of preference relations. States that recognize that observers are likely to underestimate their true preference similarity should be more likely to ally.

Measuring international perceptions is tricky. One cannot observe leaders' perceptions. However, one can use AFFINITY to characterize a rational guess by states. Contemporary indicators of state preferences give observers a crude reference with which to extrapolate future trends. This information is "noisy," but it is bound to be more informative than no information at all. States use current behavior to predict where other states are headed.

To measure discrepancies between state preferences and perceptions of preference similarity, I construct a variable that codes the difference between interest similarity as measured by AFFINITY and an estimate of the interests based on a linear extrapolation of prior AFFINITY scores. I calculate a baseline estimate of the similarity of state interests using a three-year running average. I then move the measure two time periods into the future and subtract it from actual 'future' values of AFFINITY. The final step is to lag the discrepancy between the 'future' AFFINITY score and the estimate by three periods (back to the past) to represent a discrepancy between expected and actual future AFFINITY scores. I assume that states correctly anticipate their own alignments. Dyad members anticipate future values of their own AFFINITY index and can calculate the leading estimate based on past AFFINITY values. The dyad knows whether it is being underestimated. Positive values of the constructed discrepancy variable should thus indicate an increased likelihood of alliance formation. This measure is certainly far from ideal. However, the variable does indicate where signaling is most valuable. The measure is bound to miss much of what it is intended to capture, but appears much more likely to commit errors of omission than of commission.

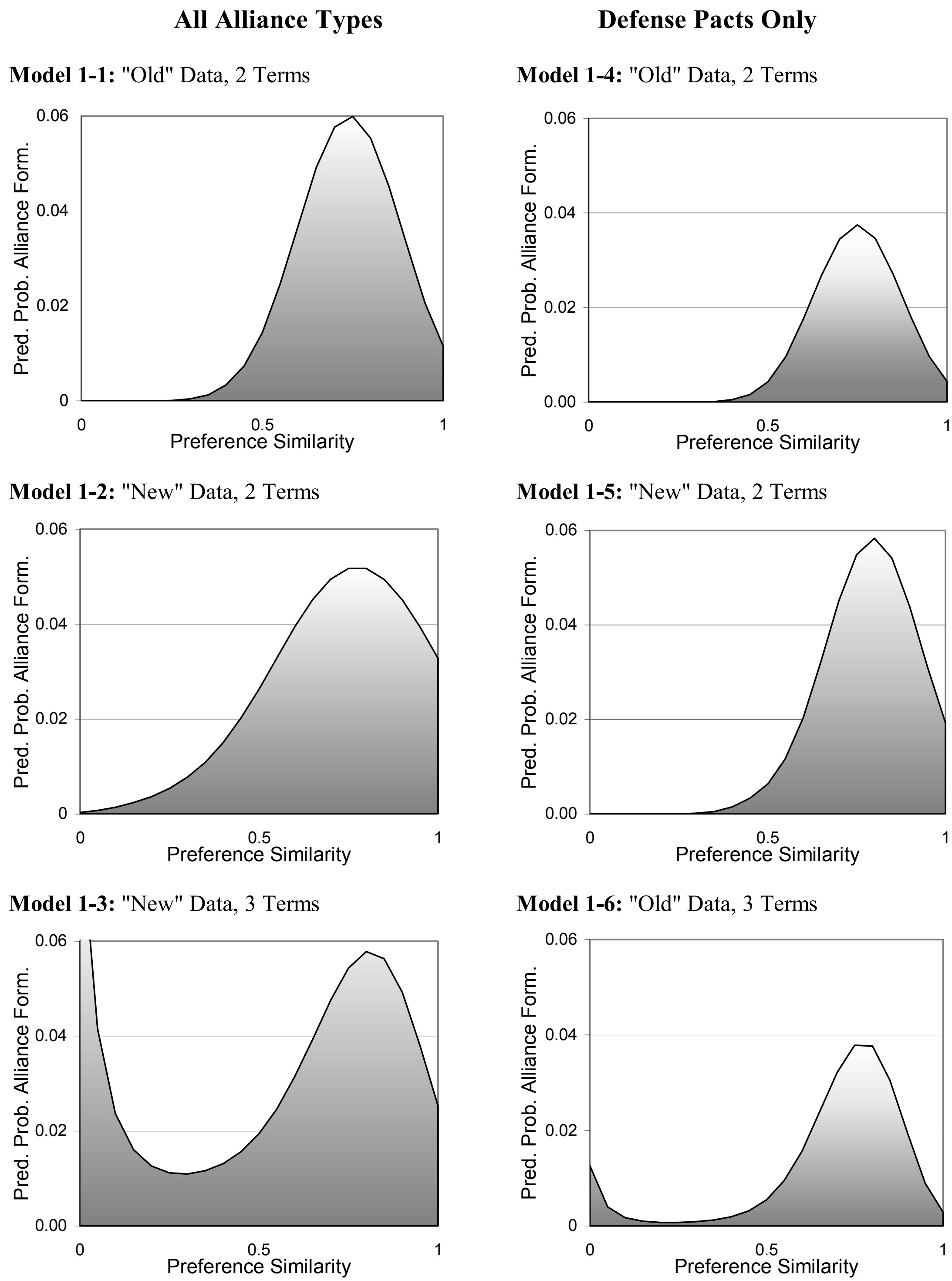

Figure 3: Predicted Probabilities of Alliance Formation for Dyads.

I append the discrepancy variable to Model 1 from Table 1. Since the regression already includes AFFINITY, DISCREPANCY captures only the effect of perceptions. DISCREPANCY is positive and significant at the 0.001 level. Even controlling for interests and other factors, underestimating preference change increases the likelihood of an alliance. Results for other variables are similar to Model 1. Rather than report the results, Figure 4 lists predicted probabilities of alliance formation for values of the linear and squared AFFINITY variables and for DISCREPANCY. Examining marginal effects facilitates a direct comparison of changes in the probability of alliance formation associated with changes in DISCREPANCY.

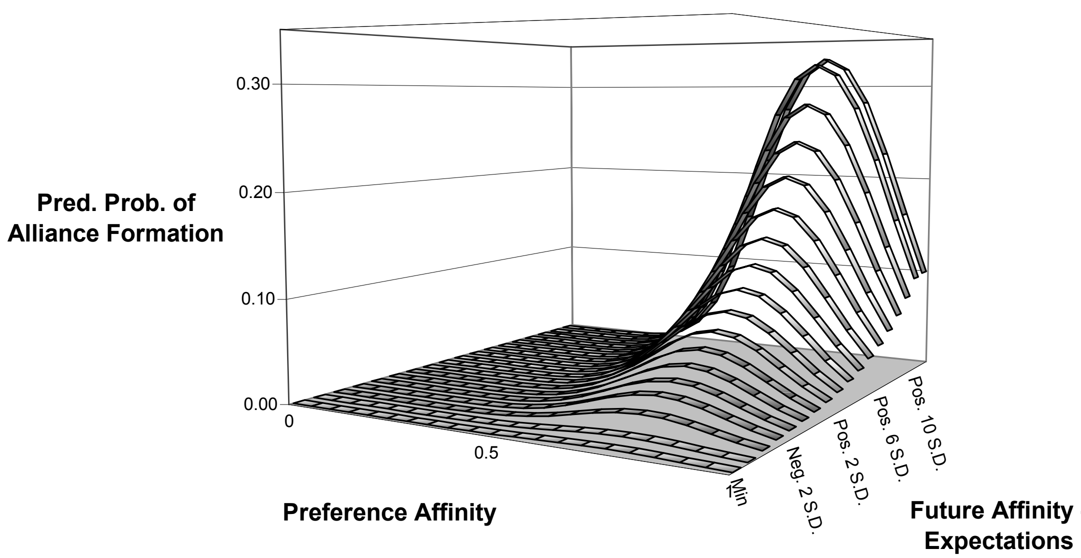

Figure 4: Effect of Discrepancies Between Expectations and Future Preference Affinity.

Two effects are apparent. First, changes in DISCREPANCY alter the probability of alliance. For example, for a state with an AFFINITY score of 0.75, increasing DISCREPANCY from its mean value by two standard deviations increases the odds of an alliance by roughly 50%. States that have closer interests than observers realize are more likely to ally than states that are not underestimated by international observers. Second, the effect is non-monotonic. The largest effect occurs among states whose preferences are similar but not identical.

These results are informative, but logit regressions with multiple terms are an inelegant way to model a non-linear process. Table 2 reports four examples of regressions using the General Additive Model (GAM). GAM allows for a non-linear relationship between the dependent variable and key independent variables. I use loess (locally smoothed regression) to estimate the relationship between the textscaffinity variable and state alliance decisions. Models 1 and 2 use the old alliance data, while models 3 and 4 report the new alliance data. Models 1 and 3 use just the single preference term, while models 2 and 4 include the DISCREPANCY variable. Results for other variables are as expected.

Perhaps the best way to assess the GAM results is to examine plots of the loess smoothed functions linking AFFINITY with alliance formation. Figure 5 provides six such plots, one for each AFFINITY variable in Table 2 and two plots of the (linear) relationship between DISCREPANCY and alliance formation. Plots are laid out as in Figure 3. Once again, the effect of AFFINITY on alliance behavior is first positive and then negative, while the probability of an alliance increases in the discrepancy between actual and expected values of AFFINITY.

6 Conclusion

Applications of game theory to international politics have tended to treat actors as either fully informed or largely ignorant about strategic features of partners or competitors. While reasonable in some contexts, a more plausible assumption is that states possess both knowledge and uncertainty. A history of interaction should lead to beliefs that are noisy, that are founded on fact but also subject to error. The task here has been to begin to examine where such a refinement in the treatment of perceptions matters, where different claims about initial beliefs lead to different empirical predictions. Tests of the perceptions argument seem to show that noisy information can

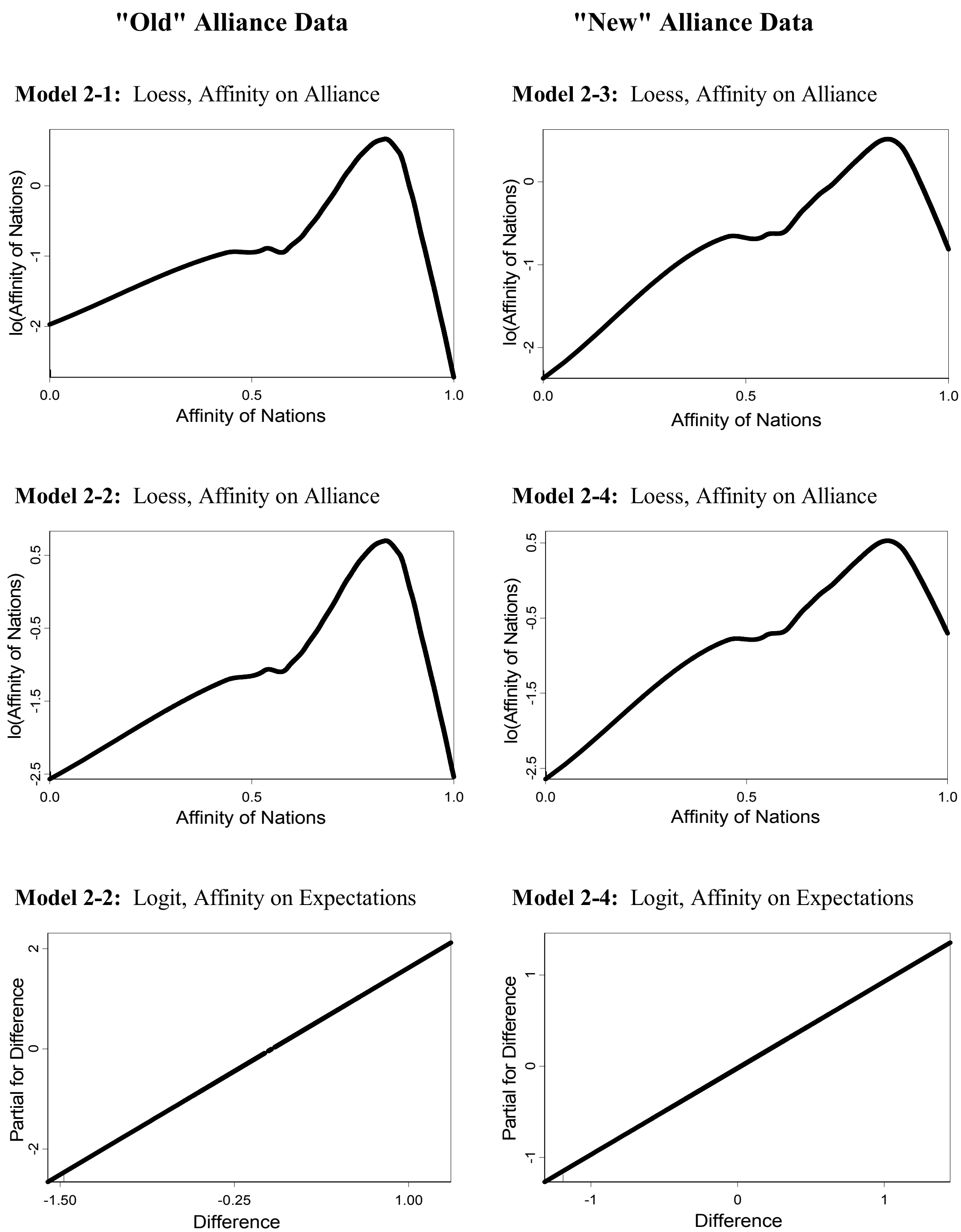

Figure 5: Plots of AFFINITY and DISCREPANCY using GAM (from Table 2).

be a significant determinant of behavior, at least when actions are primarily directed at the manipulation of beliefs. It appears that alliances are as much about perceptions as about power or preferences. States are most likely to formalize security partnerships when common interests are not recognized by the world community.

More generally, the results presented here lend support to the informational approach. Signaling theories are persuasive and deductively powerful, but their attention to communication makes them difficult to evaluate empirically (Schultz 2001*b*). The challenge for researchers is to identify situations where signaling theories make novel, testable predictions.

Finally, this study endorses the linking of insights from political psychology and rational theory. A better understanding of international relations involves relaxing assumptions that privilege material variables. At the same time, we must define areas where perceptions are, and are not, likely to matter. Modal behavior in rational theory is bound to reflect interests and material variables. Actors may misperceive in particular instances, but correct equilibrium perceptions means that aggregate behavior will tend to cancel out perceptual errors, leading rational theory to discount misperception as an explanation and leading proponents of psychological explanations to discount rational theory. I provide an example of a situations where stochastic variation in perception produces novel and counterintuitive empirical consequences. It is these unanticipated characteristics that deserve attention and which help to validate and extend informational theories in international relations.

A Equilibria and Optimal Strategies for the Games

The appendix provides details of players' optimal strategies and equilibria for both the asymmetric information and noisy information games discussed in the text.

A.0.1 Players' Optimal Strategies for Asymmetric Information Variant:

$$\text{Where } t = \frac{1}{2}(1 + x_{SQ} - c) \text{ if } x_{SQ} \leq x_A,$$
$$\text{and } t = \frac{1}{2}(x_{SQ} - w_A + c) \text{ if } x_{SQ} > x_A, \; s.t. \; (0 \leq \boldsymbol{t} \leq \frac{1}{2}).$$

$$\begin{aligned}
\mathbf{A}: \quad \boldsymbol{a} &= 1 \quad \text{if} \quad (1 - x_{SQ}) - p_A^{a,i} \leq w_C < (1 - x_{SQ}) - p_A^{\sim a,i}, \\
&\quad \text{and} \quad x_A \geq \frac{1}{2}(1 - w_A), \text{ and } x_A < (1 - |x_{SQ} - x_A| - c) \\
&\quad \text{or if} \quad (1 - x_{SQ}) - p_A^{a,i} \leq w_C < (1 - x_{SQ}) - p_A^{\sim a,i}, \\
&\quad \text{and} \quad x_A < \frac{1}{2}(1 - w_A), \text{ and } (-|x_{SQ} - x_A| - c) > (-x_A - w_A) \\
&= 0 \quad \text{if else.} \\
\boldsymbol{i} &= 1 \quad \text{if} \quad \boldsymbol{a} = 1, \; \boldsymbol{f} = 1, \text{ and } w_A < (1 - 2x_A) + r \\
&\quad \text{or if} \quad \boldsymbol{a} = 0, \; \boldsymbol{f} = 1, \text{ and } w_A < (1 - 2x_A) \\
&= 0 \quad \text{if else.} \\
\mathbf{C}: \quad \boldsymbol{f} &= 1 \quad \text{if} \quad \boldsymbol{a} = 1, \text{ and } w_C < (1 - x_{SQ}) - p_A^{a,i}, \\
&\quad \text{and where} \quad p_A^{a,i} = \left(\frac{1 - w_A + r}{2t}\right), \; s.t. \; 0 \leq p_A^{a,i} \leq 1 \\
&\quad \text{or if} \quad \boldsymbol{a} = 0, \text{ and } w_C < (1 - x_{SQ}) - p_A^{\sim a,i}, \\
&\quad \text{and where} \quad p_A^{\sim a,i} = 1 - \left(\frac{w_A}{1 - 2t}\right), \; s.t. \; 0 \leq p_A^{\sim a,i} \leq 1 \\
&= 0 \quad \text{if else.} \\
\boldsymbol{b} &= (x_A | a = 1, x_A \in x \sim \text{U}[0, t]) \\
&\quad \text{if} \quad (1 - x_{SQ}) - p_A^{a,i} \leq w_C < (1 - x_{SQ}) - p_A^{\sim a,i} \\
&= \left(x_A | a = 0, x_A \in x \sim \text{U}[t, \frac{1}{2}]\right) \\
&\quad \text{if} \quad (1 - x_{SQ}) - p_A^{a,i} \leq w_C < (1 - x_{SQ}) - p_A^{\sim a,i} \\
&= \left(x_A | a = 0, x_A \in x \sim \text{U}[0, \frac{1}{2}]\right) \quad \text{if else.}
\end{aligned}$$

A.0.2 Equilibria for Asymmetric Information Variant:

$$
\begin{array}{lll}
\textbf{[Node 3]} & \text{if} & x_A \geq \frac{1}{2}(1-w_A)\,,\ \text{and } x_A < (1-|x_{SQ}-x_A|-c)\,, \\
& \text{and} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
& \text{or if} & x_A < \frac{1}{2}(1-w_A)\,,\ \text{and } (-|x_{SQ}-x_A|-c) > (-x_A-w_A)\,, \\
& \text{and} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
\textbf{[Node 5]} & \text{if} & x_A < \frac{1}{2}(1-w_A)\,,\ \text{and } w_C < (1-x_{SQ}) - p_A^{a,i} \\
& \text{or if} & x_A < \frac{1}{2}(1-w_A)\,,\ \text{and } (-|x_{SQ}-x_A|-c) \leq (-x_A-w_A)\,, \\
& \text{and} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
\textbf{[Node 6]} & \text{if} & \frac{1}{2}(1-w_A) \leq x_A < \frac{1}{2}(1-w_A+r)\,, \\
& \text{and} & w_C < (1-x_{SQ}) - p_A^{a,i},\ \text{and } x_A \geq (1-|x_{SQ}-x_A|-c) \\
& \text{or if} & x_A \geq \frac{1}{2}(1-w_A+r)\,,\ \text{and } w_C < (1-x_{SQ}) - p_A^{a,i} \\
& \text{or if} & x_A \geq \frac{1}{2}(1-w_A)\,,\ \text{and } x_A \geq (1-|x_{SQ}-x_A|-c)\,, \\
& \text{and} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
\textbf{[Node 4]} & \text{if else.} &
\end{array}
$$

A.0.3 Players' Optimal Strategies for "Noisy" Information Variant:

$$
\begin{aligned}
\text{Where } t &= \frac{1}{2}(1+x_{SQ}-c)\ \text{ if } x_{SQ} \leq x_A, \\
\text{and } t &= \frac{1}{2}(x_{SQ}-w_A+c)\ \text{ if } x_{SQ} > x_A,\ s.t.\ x_N \leq t \leq (x_N+\epsilon)\,.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A}: \quad \boldsymbol{a} &= 1 \quad \text{if} \quad (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
&\quad \text{and} \quad (x_N+\epsilon) > \frac{1}{2}(1-w_A), \text{ and } x_A \geq \frac{1}{2}(1-w_A) \\
&\quad \text{and} \quad x_A < (1-|x_{SQ}-x_A|-c) \\
&\quad \text{or if} \quad (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
&\quad \text{and} \quad (x_N+\epsilon) > \frac{1}{2}(1-w_A), \\
&\quad \text{and} \quad x_A < \frac{1}{2}(1-w_A) \text{ and } (-|x_{SQ}-x_A|-c) > (-x_A-w_A) \\
&= 0 \quad \text{if else.} \\
\boldsymbol{i} &= 1 \quad \text{if} \quad \boldsymbol{a}=1,\ \boldsymbol{f}=1,\ \text{and } w_A < (1-2x_A)+r \\
&\quad \text{or if} \quad \boldsymbol{a}=0,\ \boldsymbol{f}=1,\ \text{and } w_A < (1-2x_A) \\
&= 0 \quad \text{if else.} \\
\mathbf{C}: \quad \boldsymbol{f} &= 1 \quad \text{if} \quad \boldsymbol{a}=1,\ \text{and } w_C < (1-x_{SQ}) - p_A^{a,i}, \\
&\quad \text{and} \quad (x_N+\epsilon) > \frac{1}{2}(1-w_A+r), \\
&\quad \text{and where } p_A^{a,i} = \left(\frac{\frac{1}{2}(1-w_A+r)-x_N}{t-x_N}\right),\ s.t.\ 0 \leq p_A^{a,i} \leq 1 \\
&\quad \text{or if} \quad \boldsymbol{a}=0,\ \text{and } w_C < (1-x_{SQ}) - p_A^{\sim a,i},\ \text{and } (x_N+\epsilon) > \frac{1}{2}(1-w_A), \\
&\quad \text{and where} \quad p_A^{\sim a,i} = 1 - \left(\frac{(x_N+\epsilon)-\frac{1}{2}(1-w_A)}{(x_N+\epsilon)-t}\right),\ s.t.\ 0 \leq p_A^{\sim a,i} \leq 1 \\
&= 0 \quad \text{if else.} \\
\boldsymbol{b} &= (x_A | a=1, x_A \in x \sim \mathrm{U}[x_N, t]) \\
&\quad \text{if} \quad (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
&= (x_A | a=0, x_A \in x \sim \mathrm{U}[t, (x_N+\epsilon)]), \\
&\quad \text{if}\, (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i} \\
&= (x_A | a=0, x_A \in x \sim \mathrm{U}[x_N, (x_N+\epsilon)]) \text{ if else.}
\end{aligned}
$$

A.0.4 Equilibria for "Noisy" Information Variant:

$$
\begin{array}{lll}
\textbf{[Node 3]} & \text{if} & x_A \geq \frac{1}{2}(1-w_A), \text{ and } x_A < (1-|x_{SQ}-x_A|-c), \\
& \text{and} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A) \\
& \text{or if} & x_A < \frac{1}{2}(1-w_A), \text{ and } (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A), \text{ and } (-|x_{SQ}-x_A|-c) > (-x_A-w_A) \\
\textbf{[Node 5]} & \text{if} & x_A < \frac{1}{2}(1-w_A), \text{and } w_C < (1-x_{SQ}) - p_A^{a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A) \\
& \text{or if} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A), \text{ and } (-|x_{SQ}-x_A|-c) \leq (-x_A-w_A) \\
\textbf{[Node 6]} & \text{if} & \frac{1}{2}(1-w_A) \leq x_A < \frac{1}{2}(1-w_A+r), \text{ and } w_C < (1-x_{SQ}) - p_A^{a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A), \text{ and } x_A \geq (1-|x_{SQ}-x_A|-c) \\
& \text{or if} & x_A \geq \frac{1}{2}(1-w_A+r), \text{ and } w_C < (1-x_{SQ}) - p_A^{a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A) \\
& \text{or if} & (1-x_{SQ}) - p_A^{a,i} \leq w_C < (1-x_{SQ}) - p_A^{\sim a,i}, \\
& \text{and} & (x_N+\epsilon) > \frac{1}{2}(1-w_A), \text{ and } x_A \geq (1-|x_{SQ}-x_A|-c) \\
& \text{and} & x_A \geq \frac{1}{2}(1-w_A) \\
\textbf{[Node 4]} & \text{if else.} &
\end{array}
$$

References

Altfeld, Michael F. 1984. "The Decision to Ally: A Theory and Test." *Western Political Quarterly* 37(4):523–544.

Beck, Neal, Jonathan Katz & Richard Tucker. 1998. "Taking Time Seriously: Time-series–Cross-section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816–1984." *American Journal of Political Science* 41(3):846–878.

Bennett, D. Scott & Allan C. Stam. 2001. "EUGene: Expected Utility and Data Management Program, version 2.250." Computer Program. The Pennsylvania State University and Dartmouth University.

Bennett, D. Scott & Scott E. Tarry. 1996. "Self-Perpetuation or Rational Choices?: A Model of Rationality and Hysteresis in International Alliances." Presented at the Annual Meetings of the Midwest Political Science Association, Chicago, IL.

Betts, Richard K. 1985. "Conventional Deterrence: Predictive Uncertainty and Policy Confidence." *World Politics* 37(2):153–179.

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

Bueno de Mesquita, Bruce & David Lalman. 1992. *War and Reason: Domestic and International Imperatives.* New Haven: Yale University Press.

Cetinyan, Rupen. 2002. "Not Seeing Is Believing: What Game Theory Says about Reputation and Deterrence." University of Pittsburgh. Typescript.

Christensen, Thomas J. & Jack Snyder. 1990. "Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity." *International Organization* 44(2):137–168.

Conybeare, John A.C. 1992. "A Portfolio Diversification Model of Alliances: The Triple Alliance and Triple Entente, 1879-1914." *Journal of Conflict Resolution* 36(1):53–85.

Conybeare, John A.C. 1994. "A Portfolio Benefits of Free Riding in Military Alliances." *International Studies Quarterly* 38(3):405–419.

Downs, George & Michael A. Jones. 2002. "Reputation, Compliance, and International Law." *Journal of Legal Studies* 31:95–114.

Fearon, James D. 1994*a*. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Fearon, James D. 1994*b*. "Signaling Versus the Balance of Power and Interests: An Empirical Test of the Crisis Bargaining Model." *Journal of Conflict Resolution* 38(2):236–269.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414.

Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs." *Journal of Conflict Resolution* 41(1):68–90.

Fearon, James D. 2002. "Selection Effects and Deterrence." *International Interactions* 28(1):5–29.

Feltovich, Nick, Rich Harbaugh & Ted To. 2001. "Too Cool for School?: Signaling and Countersignaling." University of Houston. Typescript.

Gartzke, Erik & Kristian S. Gleditsch. 2004. "Why Democracies May Actually be Less Reliable Allies." *American Journal of Political Science* 48(4):775–795.

Gartzke, Erik, Quan Li & Charles Boehmer. 2001. "Investing in the Peace: Economic Interdependence and International Conflict." *International Organization* 55(2):391–438.

Gaubatz, Kurt Taylor. 1996. "Democratic States and Commitment in International Relations." *International Organization* 50(1):109–139.

George, Alexander L. 1991. *Forceful Persuasion: Coercive Diplomacy as an Alternative to War.* Washington DC: United States Institute of Peace Press.

Gibler, Douglas & Meredith Reid Sarkees. 2004. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816–2000." *Journal of Peace Research* 41(2):211–222.

Holsti, Ole R., P. Terrence Hopmann & John D. Sullivan. 1973. *Unity and Disintegration in International Alliances: Comparative Studies.* New York: John Wiley & Sons.

Huth, Paul K. 1988. *Extended Deterrence and the Prevention of War.* New Haven: Yale University Press.

Huth, Paul K. 1997. "Reputations and Deterrence: A Theoretical and Empirical Assessment." *Security Studies* 7(1):72–99.

Jervis, Robert. 1970. *The Logic of Images in International Relations.* Princeton, NJ: Princeton University Press.

Jervis, Robert. 2001. Signaling and Perception: Drawing Inferences and Projecting Images. In *Political Psychology.*, ed. Kristen Renwick Monroe. Mahwah, NJ: Lawrence Erlbaum Associates pp. 293–312.

Kilgour, Marc D. & Frank Zagare. 1991. "Credibility, Uncertainty, and Deterrence." *American Journal of Political Science* 35(2):305–334.

King, Gary & Langche Zeng. 2001*a*. "Explaining Rare Events in International Relations." *International Organization* 55(3):693–715.

King, Gary & Langche Zeng. 2001*b*. "Logistic Regression in Rare Events Data." *Political Analysis* 9(2):137–163.

Kydd, Andrew. 1997. "Game Theory and the Spiral Model." *World Politics* 49(3):371–383.

Lai, Brian & Dan Reiter. 2000. "Democracy, Political Similarity, and International Alliances, 1816–1992." *Journal of Conflict Resolution* 44(2):203–227.

Leeds, Brett Ashley. 1999. "Domestic Political Institutions, Credible Commitments, and International Cooperation." *American Journal of Political Science* 43(4):979–1002.

Leeds, Brett Ashley. 2003*a*. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57:801–827.

Leeds, Brett Ashley. 2003*b*. "Do Alliances Deter Aggression?: The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47(3):427–439.

Leeds, Brett Ashley, Andrew G. Long & Sara McLaughlin Mitchell. 2000. "Reevaluating Alliance Reliability: Specific Threats, Specific Promises." *Journal of Conflict Resolution* 44(5):686–699.

Maoz, Zeev. 1999. "Dyadic Militarized Interstate Disputes (DYMID1.1) Dataset, version 1.1." Computer File. Tel-Aviv University.

Mercer, Jonathan. 1996. *Reputation and International Politics.* Ithaca, NY: Cornell University Press.

Mercer, Jonathan. 1997. "Reputation and Rational Deterrence Theory." *Security Studies* 7(1):100–113.

Morgenthau, Hans J. 1973 [1948]. *Politics Among Nations: The Struggle for Power and Peace.* New York: Knopf.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35(4):904–933.

Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38(2):270–294.

Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3:63–83.

Nalebuff, Barry. 1991. "Rational Deterrence in an Imperfect World." *World Politics* 43(3):313–335.

Oneal, John R. & Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly* 41(2):267–293.

O'Neill, Barry. 2001. "Risk Aversion in International Relations Theory." *International Studies Quarterly* 45(4):617–640.

Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility.* Cambridge: Cambridge University Press.

Press, Daryl G. 2004. "The Credibility of Power." *International Security* 29(3):135–169.

Press, Daryl G. 2005. *Calculating Credibility: How Leaders Assess Military Threats.* Ithaca, NY: Cornell University Press.

Reed, William. 1997. "Alliance Duration and Democracy: An Extension and Cross-Validation of 'Democratic States and Commitment in International Relations'." *American Journal of Political Science* 41(3):1072–1078.

Sartori, Anne. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 55(3):121–149.

Sartori, Anne. 2005. *Deterrence by Diplomacy.* Princeton, NJ: Princeton University Press.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 94(4):829–844.

Schultz, Kenneth A. 1999. "Do Domestic Institutions Constrain or Inform?: Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53(2):233–266.

Schultz, Kenneth A. 2001*a*. *Democracy and Coercive Diplomacy.* Cambridge: Cambridge University Press.

Schultz, Kenneth A. 2001*b*. "Looking for Audience Costs." *Journal of Conflict Resolution* 45(1):32–60.

Signorino, Curtis S. & Jeffrey M. Ritter. 2001. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43(1):115–144.

Simmons, Beth. 1998. "Compliance with International Agreements." *Annual Review of Political Science* 1:75–93.

Simmons, Beth A. & Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99(4):623–631.

Simon, Michael W. & Erik Gartzke. 1996. "Political System Similarity and the Choice of Allies: Do Democracies Flock Together or Do Opposites Attract?" *Journal of Conflict Resolution* 40(4):617–635.

Siverson, Randolph M. & Harvey Starr. 1990. "Opportunity, Willingness, and the Diffusion of War." *American Political Science Review* 84(1):47–67.

Siverson, Randolph M. & Harvey Starr. 1991. *The Diffusion of War: A Study of Opportunity and Willingness.* Ann Arbor: University of Michigan Press.

Siverson, Randolph M. & Harvey Starr. 1994. "Regime Change and the Restructuring of Alliances." *American Journal of Political Science* 38(1):145–161.

Siverson, Randolph M. & Juliann Emmons. 1991. "Birds of a Feather: Democratic Political Systems and Alliance Choices in the Twentieth Century." *Journal of Conflict Resolution* 35(2):285–306.

Small, Melvin & J. David Singer. 1990. Formal Alliances, 1816–1965: An Extention of the Basic Data. In *Measuring the Correlates of War*, ed. J. David Singer & Paul Diehl. University of Michigan Press pp. 159–190.

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39(4):405–425.

Smith, Alastair. 1996. "To Intervene or Not to Intervene: A Biased Decision." *Journal of Conflict Resolution* 40(1):16–40.

StataCorp. 1995. *Stata Reference Manual 4, vols. 1-3.* College Station, TX: Stata Press.

Wagner, R. Harrison. 1982. "Deterrence and Bargaining." *Journal of Conflict Resolution* 26(2):329–358.

Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science* 44(3):469–484.

Walt, Stephen M. 1987. *The Origins of Alliances.* Ithaca, NY: Cornell University Press.

Walter, Barbara F. 2006*a*. "Building Reputation: Why Governments Fight Some Separatists but Not Others." *American Journal of Political Science* 50(2):313–330.

Walter, Barbara F. 2006*b*. "Information, Uncertainty, and the Decision to Secede." *International Organization* 60(1):xxx–xxx.

Waltz, Kenneth N. 1979. *Theory of International Politics.* New York: McGraw-Hill.

Table 1: Logit Estimates of Alliance Formation (Robust S.E., Adj. Clustering in Dyads)

	ALL ALLIANCES			DEFENSE PACTS		
	Model 1–1 “Old” data	Model 1–2 “New” data	Model 1–3 “New” data	Model 1–4 “Old” data	Model 1–5 “New” data	Model 1–6 “Old” data
AFFINITY						
Linear	45.50 ***	19.16 **	– 22.93 **	65.70 ***	60.75 ***	– 32.94 ***
	(8.98)	(5.58)	(7.01)	(9.92)	(8.91)	(6.92)
Square	– 30.69 ***	– 12.37 ***	53.99 ***	– 43.77 ***	– 38.08 ***	92.43 ***
	(5.61)	(3.51)	(11.54)	(6.19)	(5.44)	(13.96)
Cubed			– 32.75 ***			– 61.29 ***
			(6.24)			(8.13)
DISTANCE [1]	– 0.66 ***	– 0.91 ***	– 0.88 ***	– 0.69 ***	– 1.10 ***	– 0.70 ***
	(0.11)	(0.16)	(0.15)	(0.09)	(0.16)	(0.09)
POL. REL. DYAD [2]	– 3.03 ***	– 4.87 ***	– 4.63 ***	– 3.10 ***	– 6.35 ***	– 3.16 ***
	(0.87)	(1.24)	(1.21)	(0.76)	(1.26)	(0.76)
DEMOCRACY						
Lower Dyadic [3]	0.11 ***	0.17 ***	0.16 ***	0.16 ***	0.13 ***	0.15 ***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Higher Dyadic [3]	– 0.16 ***	– 0.16 ***	– 0.16 ***	– 0.16 ***	– 0.13 ***	– 0.17 ***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
CAPABILITIES						
Lower Dyadic [4]	– 40.92	– 73.84 **	– 70.56 **	– 21.11	– 105.44 **	– 22.63
	(21.24)	(21.59)	(21.69)	(15.93)	(30.58)	(15.30)
Higher Dyadic [4]	9.36 ***	9.98 ***	9.80 ***	12.75 ***	13.17 ***	12.80 ***
	(2.14)	(1.91)	(1.96)	(2.15)	(2.62)	(2.17)
MAJOR POWER						
One [2]	1.97 *	4.09 **	3.84 **	1.50	4.36 **	1.54 *
	(0.92)	(1.31)	(1.28)	(0.77)	(1.51)	(0.78)
Both [2]	5.08 **	4.93 *	5.04 *	3.73 **	7.29 **	3.62 **
	(1.59)	(1.97)	(1.95)	(1.25)	(2.26)	(1.23)
Δ # [2]	1.60 ***	1.86 ***	1.93 ***	2.93 ***	2.03 ***	2.98 ***
	(0.10)	(0.12)	(0.12)	(0.19)	(0.13)	(0.19)
WAR						
Dyadic [2]	– 3.20 **	– 1.26	– 1.41	– 2.65 ***	– 1.70	– 2.50 **
	(0.92)	(0.98)	(0.97)	(0.75)	(1.16)	(0.74)
3rd Party, Lag [2]	0.025	– 0.09	– 0.08	– 0.69 ***	– 0.39 **	– 0.68 ***
	(0.09)	(0.10)	(0.10)	(0.14)	(0.12)	(0.14)
3rd Party, Lead [2]	– 0.24 *	0.07	0.07	– 0.07	0.07	– 0.08
	(0.10)	(0.09)	(0.09)	(0.14)	(0.11)	(0.14)
System [2]	1.36 *	– 0.06	– 0.09	0.94	– 0.35	0.94
	(0.62)	(0.49)	(0.50)	(1.04)	(0.51)	(1.04)
CONSTANT	– 15.62 ***	– 3.37	4.54 *	– 24.12	– 18.23	– 0.84
	(3.45)	(2.15)	(2.04)	(3.94)	(3.52)	(1.71)
N	20407	14176	14176	20407	14176	20407
Wald Chi^2	565.45	701.21	728.41	594.47	638.83	604.51
Prob > Chi^2	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R^2	0.21	0.27	0.27	0.36	0.31	0.36
Log Likelihood	– 2120.74	– 1650.25	– 1640.16	– 1051.69	– 1410.70	– 1040.83

Values in parentheses are standard errors. All significance tests are two-tailed.

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$
[1] logged variable
[2] dummy variable
[3] democracy variable = [(Dem – Aut + 10)/2]
[4] capability variable = [CINC/total]

Table 2: GAM Estimates of Alliance Formation

	ALL ALLIANCES			
	Model 2–1 "Old" data	Model 2–2 "Old" data	Model 2–3 "New" data	Model 2–4 "New" data
AFFINITY				
Similarity of Preferences	-40.89 **	20.79 *	12.63	21.53 **
	(16.09)	(11.54)	(8.87)	(9.45)
Difference btw Affinity & Exp.		1.65 ***		0.95 ***
		(0.32)		(0.35)
DISTANCE [1]	-0.69 ***	-0.68 ***	-0.87 ***	-0.87 ***
	(0.05)	(0.04)	(0.05)	(0.05)
POL. REL. DYAD [2]	-3.16 ***	-3.17 ***	-4.53 ***	-4.56 ***
	(0.45)	(0.36)	(0.45)	(0.45)
DEMOCRACY				
Lower Dyadic [3]	0.15 ***	0.10 ***	0.16 ***	0.16 ***
	(0.02)	(0.02)	(0.02)	(0.02)
Higher Dyadic [3]	-0.16 ***	-0.14 ***	-0.15 ***	-0.15 ***
	(0.03)	(0.02)	(0.02)	(0.02)
CAPABILITIES				
Lower Dyadic [4]	-26.53	-41.26 **	-70.16 **	-69.97 **
	(15.72)	(13.49)	(14.27)	(14.30)
Higher Dyadic [4]	12.82 ***	9.65 ***	9.97 ***	10.15 ***
	(1.74)	(1.54)	(1.54)	(1.55)
MAJOR POWER				
One [2]	1.57 **	2.08 ***	3.70 ***	3.73 ***
	(0.44)	(0.39)	(0.48)	(0.48)
Both [2]	3.57 **	4.23 ***	4.62 ***	4.57 **
	(1.47)	(0.98)	(1.24)	(1.27)
Δ # [2]	2.99 ***	1.73 ***	1.96 ***	1.97 ***
	(0.15)	(0.10)	(0.11)	(0.11)
WAR				
Dyadic [2]	-2.43	-2.87	-1.24	-1.30
	(1.42) *	(1.20) **	(1.91)	(1.96)
3rd Party, Lag [2]	-0.67 ***	0.04	-0.09	-0.09
	(0.14)	(0.08)	(0.09)	(0.09)
3rd Party, Lead [2]	-0.12	-0.29 ***	0.06	0.05
	(0.13)	(0.09)	(0.09)	(0.09)
System [2]	1.05 *	1.31	-0.06	-0.04
	(1.00)	(0.58)	(0.40)	(0.40)
CONSTANT	-0.14	0.87	3.45 ***	3.41 ***
	(1.07)	(0.67)	(0.59)	(0.59)
N	20407	20234	14176	14040
Wald Chi^2	144.37	326.76	83.27	79.10
Prob > Chi^2	0.00	0.00	0.00	0.00

Values in parentheses are standard errors. All significance tests are two-tailed.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

[1] logged variable

[2] dummy variable

[3] democracy variable = [(Dem – Aut + 10)/2]

[4] capability variable = [CINC/total]