The Logic of Secret Alliances

Peter Bils Bradley C. Smith

Abstract

Alliances are typically understood as agreements intended to deter aggression from enemy states. By signaling an ally's commitment to a protégé state, a shared enemy may be deterred from attacking. In light of this signaling logic, secret alliances are puzzling. Because they are not observed, secret alliances by definition cannot achieve deterrence through signaling. So why do states enter into these secret agreements? We argue that alliances may also act as a signal of a state's intentions, potentially undermining deterrence. If a new alliance signals that the members' interests are unaligned with those of a shared enemy, the enemy may be provoked into war. As a consequence, states sometimes want to keep alliance agreements secret. To develop this argument, we analyze a formal model in which states may enter into secret or public alliance agreements. We find that the possibility of secret alliances has general effects on deterrence - possibly undermining the deterrent value of concurrent public alliances.

Introduction

A primary role of military alliances is to deter aggression from enemies. One important mechanism through which alliances influence these calculations is *signaling* (Morrow, 2000, p. 68-71). Alliances achieve this goal by serving as a costly signal of members' commitment to fight alongside one another against a shared enemy (Morrow, 1991, 1994). Through this mechanism, alliances alter the calculus of deterrence, reducing an enemy's incentive to attack.

In light of this signaling logic, secret alliances stand out as odd. For the signaling mechanism to operate, it is self-evident that an alliance, and the costs imposed by it, must be visible to adversaries. Secrecy undermines this mechanism by keeping these signals hidden from view, eliminating the deterrent value of alliances. This is consistent with empirical evidence, which finds that secret alliances are associated with an increased probability of conflict (Bas and Schub, 2016).

One possibility is that secret alliances are not intended for deterrence, but rather some other purpose. While this may be true in the case of some types of alliances, such as offense pacts, it is far from the norm. In fact, secret alliances have often been explicitly defensive in nature, intended to only be activated in the event of aggression from an adversary. According to the Alliance Treaty Obligations and Provisions (ATOP) data (Leeds et al., 2002), 63% of all secret alliances were defensive. Furthermore, in the period 1816-1918, 62% of all alliances were secret and, as Hinsley notes, in the Bismarckian era these secret alliances were "...without exception defensive, aimed at preserving the status quo..." (1967, p. 257-258).[1]

Another potential explanation is that many so-called "secret" alliances were not meaningfully secret, being commonly known in spite of their formal provisions requiring secrecy. Again, though this may explain some cases, many important alliances were meaningfully secret. For example, Ramm (1937) documents British knowledge about European alliances in the period 1879-1895, highlighting the significant role that uncertainty over relevant alliance commitments played in shaping British foreign policy in this era. Although British intelligence officers identified the possible existence of the Franco-Russian Alliance, they were uncertain of its purpose, believing that it possibly was designed to target and contain Russian aggression (Ramm, 1937, p. 87-133). In reality, however, Russia was a member of the alliance, rather than a target. This is an example of a more general pattern; even when the existence of an alliance was known or suspected, significant uncertainty often remained about the specific content or obligations contained in these alliances (Ritter, 2004, p. 38-39).

That secret alliances are often both defensive in nature and meaningfully secret presents a puzzle. If a primary goal of alliances is to deter enemies through signaling, then why would allies keep their

[1] Moreover, these estimates are a lower bound as there may be secret alliances that never become public and, thus, are not observed in the ATOP data.

commitments to one another secret?

In this paper, we propose that secret alliances are an attractive option when alliances may act as a signal of members' *alignment*. In our theory, alliances play two roles. First, they aggregate capabilities, allowing members to better resist aggression from a shared enemy. Second, and importantly for our argument, the presence of an alliance may endogenously reveal information about the constellation of interests in the international system. Consequently, a new alliance can indicate that a state previously thought to be friendly is, in reality, an adversary. We argue that if there is significant uncertainty about the alignment of states' interests then the second effect dominates and secret alliances emerge.

To develop this argument, we analyze a model in which two states, P and A, may form an alliance and, if so, may choose to publicly disclose its existence or not to a third state, E. Next, E can choose to attack P or not. A key ingredient of our modeling approach is incomplete information about how aligned state P is with A relative to E. This contrasts with existing work, which focuses on uncertainty over an ally's resolve or capabilities. We show that, indeed, it is possible for secret alliances to occur in equilibrium when there is uncertainty about whether P is aligned with A or E. Specifically, if E expects that P is an enemy but still believes there is some probability they have similar interests, then secret alliances are formed with positive probability in every equilibrium.

We also derive a number of results tying secret alliances to conflict. In our model, secret alliances arise due to P's desire to avoid provoking attack from an enemy. Ironically, however, we find that in equilibrium secret alliances are associated with war rather than peace. This occurs because P's incentive to keep the alliance secret is strongest precisely when E is most likely to believe an attack is attractive. Consequently, in equilibrium, making an alliance secret is not effective for avoiding conflict. The reason is that if E never attacks after failing to see a public alliance, then P and A could simply forgo forming the secret alliance. Moreover, our results suggest that the conditions which lead to secret alliances are also the conditions under which conflict is most likely. Thus, the model implies a correlation between secret alliances and conflict. This is consistent with previous work and we show this correlation holds specifically for secret defensive alliances. Most interestingly, we find that the existence of secret alliances also undermines the deterrent value of concurrent public alliances. If public alliances remained deterrent in a world of secret alliances, then states could profitably deviate from forming a secret alliance to forming a public one. Using data on alliances and military disputes, we find that when secret alliances are prevalent concurrent public alliances are less deterrent.

Finally, we discuss variation in secret alliances over time. In particular, there is a sharp decline in the number of secret alliances after World War II. Our theory suggests two possible mechanisms. First, the decline in uncertainty in international relations could have contributed to the decline in secret alliances. During the period of multipolarity, there was high uncertainty about alignments and intentions, whereas bipolarity reduced this uncertainty. Second, the increase in the deterrent value of alliances may have contributed to the decline in secret alliances. The emergence of nuclear weapons made states allied with nuclear-capable states unattractive targets for aggression, making highly deterrent alliances more prevalent and decreasing the need for secrecy.

Connections to the Literature

Our primary contribution is to advance a novel theoretical explanation for secret alliances. Previous work has explained secrecy as the product of intra-alliance dynamics. For example, Smith (2021) argues that secrecy can enhance the credibility of communication between military partners in an immediate crisis. Similarly, Kuo (2020) argues that secrecy begets more secrecy, as states want to conceal their commitment to one partner so that another partner does not question their credibility. In contrast to these studies, which point to intra-alliance dynamics, our model focuses on ally-enemy interactions. As such, this study is most closely related to Ritter (2004), who argues that a defending state may wish to keep its alliance commitments secret to prevent a shared enemy from taking "countermeasures" that would negate the deterrent value of an alliance.[2] While Ritter focuses on the military dimension of secrecy, our argument places politics at the forefront, treating alliances as informative of the member's *alignment*, rather than their capabilities.

By treating alliances as signals of members' interests, our theory builds on a long-standing line of research that connects alliances to the constellation of interests in the international system. The notion that alliances are indicative of a state's interests has also been a longstanding feature of quantitative empirical work on international conflict. Both Altfeld and Bueno De Mesquita (1979) and Signorino and Ritter (1999) use alliance portfolios to measure the similarity of states' alignment. Our work connects these empirical insights to theoretical work on secret alliances, incorporating the idea that alliances signal states' alignment into a formal model of secrecy and deterrence.

We also contribute broadly to the literature on alliance politics and security cooperation. This literature has long focused on the signaling value of alliances, arguing that alliances provide a venue for costly signaling that conveys members' commitment to fight alongside one another (Morrow, 1991, 1994). Scholars have applied this costly-signaling logic to connect alliance politics to a wide variety of empirical phenomena, including nuclear deterrence (Fuhrmann and Sechser, 2014), diplomatic visits (McManus, 2018), military conscription Horowitz et al. (2017), and domestic military spending (DiGiuseppe and Shea, 2021). In each case, an ally's desire to credibly signal support drives behavior. This contrasts with our study, in which an ally may want to conceal its support for fear of provoking a common enemy. In this way, our approach is related to recent theoretical and

[2] The logic of Ritter (2004) is similar in flavor to arguments about feigning weakness such as Slantchev (2010). These paper share a greater a similarity with our extension to offensive alliances we discuss later.

empirical work that considers whether defensive alliances provoke or deter. Addressing a debate in the empirical literature concerning whether alliances provoke (Kenwick et al., 2015) or deter (Leeds and Johnson, 2017), Morrow (2017) studies a model in which alliances may signal a conflict of interest. In contrast to Morrow's analysis, which treats all alliances as public knowledge, we consider endogenous secrecy - allowing states to conceal their alliances in an attempt to avoid provocation.

Further, alliances are known to vary along a variety of dimensions (Leeds et al., 2002). This variation is important in determining both the deterrent effect of alliances (Leeds, 2003b) and their reliability when challenged (Leeds and Long, 2000; Leeds, 2003a). By offering an explanation for secrecy, we contribute to this literature by detailing the strategic logic of an aspect of alliance design that is not currently well understood. In offering this explanation, we highlight the important role of state interests. Previous studies have shown that states' foreign policy interests shape the success of communication and signaling, whether it be cheap talk, as in Smith and Trager, or costly signals, as in Wolford. Our model builds on these studies by showing that if alliances act as signals of states' interests then members may want to conceal this information through secrecy provisions.

Our paper also relates to models in which a state may strategically generate ambiguity about its military capabilities by making unobserved arming investments. Meirowitz and Sartori (2008) show that meaningful uncertainty can be sustained in equilibrium when arming is unobserved. We instead focus on the decision to keep an alliance secret.[3] That is, the state in our model is allowed to form a public alliance and, thus, credibly reveal its increased military capacity.[4] Because secrecy arises in our model due to a desire to conceal one's type, the logic of equilibrium secrecy is distinct from that of Meirowitz and Sartori's, in which states use mixed strategies to endogenously generate uncertainty about their level of arming. More related is Baliga and Sjöström (2008), which studies a setting in which a state can acquire arms and chooses whether to reveal this information to an adversary. In Baliga and Sjöström's setting, if the state has armed or not it can credibly reveal this information. In contrast, in our model the state can only reveal it has a public alliance, but it can never credibly reveal it has not formed a secret alliance. This generates different strategic incentives for maintaining ambiguity in the two models. A final important difference between our work and the literature on arming is that the costs of arming are exogenous in previous work. In contrast, the costs for forming an alliance in our model depend endogenously on the alliance partner and equilibrium probability of war. This endogeneity of costs leads to different theoretical results and is a substantively crucial feature of alliance formation, relative to arming.

[3] Meirowitz and Sartori consider an extension in which states may choose to endogenously and verifiably reveal their arming. In contrast to our findings, they find that an equilibrium always exists in which the players always reveal their arming decisions.

[4] Similarly, this distinguishes our work from other models which assume arming is unobserved but focus on different issues, such as the effectiveness of communication (Baliga and Sjöström, 2004) or institutions (Meirowitz et al., 2019) for mitigating conflict, and nuclear proliferation (Bas and Coe, 2016).

Secrecy in the Dual Alliance

Before moving to the presentation of our model, it is helpful to consider an example of a prominent secret alliance. This discussion aims to clarify and motivate our modeling decisions. Our underlying argument is that secrecy is motivated by a state's desire to conceal their interests, avoiding provocation of a potential enemy. The 1879 Dual Alliance between Germany and Austria-Hungary provides an illustration of the motivation for secrecy embodied in our theoretical model.

In 1873, Germany and Russia had been aligned alongside Austria-Hungary as members of the League of Three Emperors. The League was a formal alliance designed to foster the members' collective interest in quashing radical movements in the Balkans. The treaty served its purpose until 1878, when Russian victory in the recently concluded Russo-Turkish war gave rise to tensions between Russia and Austria-Hungary (Gildea, 2003, p. 235-238). The primary disagreement concerned the handling of Serbia in the wake of this conflict. These tensions came to a head at the Congress of Berlin, which had been organized by the leader of the League's third member: Otto von Bismarck. Given these tensions, Russia dissolved the League shortly after the Congress of Berlin in 1878. Bismarck moved quickly to negotiate an understanding that would preserve German interests in the Balkans (Andrew, 1966).

The result was a secret agreement between Germany and Austria-Hungary that was concluded on October 7, 1879 at Vienna. In negotiating the treaty, Bismarck sought to balance two objectives. First, he hoped to assure support from Austria-Hungary in order to counterbalance a potential threat from Russia, given the deterioration in relations between Russia and Germany in the wake of the Congress of Berlin. Believing that Russia would be unwilling to face both Germany and Austria-Hungary in an armed conflict simultaneously, Bismarck sought to arrange a mutual defense agreement. However, this goal was in tension with another of Bismarck's important goals: preventing further deterioration in relations with Russia. Consequently, the treaty remained secret. Though Bismarck hoped to make the treaty public to deter Russia, Austro-Hungarian diplomats repeatedly raised concerns about provocation, and the final text of the treaty included an article explicitly guaranteeing secrecy (Langer, 1951). Article IV of the alliance read: "This Treaty shall, in conformity with its peaceful character, and to avoid any misinterpretation, be kept secret by the two High Contracting Parties." This provides clear evidence of these competing goals. On one hand, the alliance was designed to shore up German defense in case of aggression from Russia. On the other hand, fear of provocation drove the members to keep the pact secret.

Model

Players: There are three states: a protégé P, a (potential) ally A, and a (potential) enemy E. State P has a type $\theta \in \Theta \subset \mathbb{R}$, where Θ is compact and contains at least two elements. Define $\overline{\theta} = \max \Theta$ and $\underline{\theta} = \min \Theta$. We assume that P and A know θ, while E has a commonly known prior belief about θ given by a distribution F with full support on Θ. The type θ can be interpreted as how aligned P is with A relative to E. Lower values of θ indicate greater alignment with A, while higher types have interests more aligned with E.

Timing and actions: First, P proposes either a public alliance, a secret alliance, or no alliance. Denote the type of alliance, if one is proposed, as $x \in \{Pub, Sec\}$. Furthermore, if P proposes an alliance it also chooses an amount $t \geq 0$ to transfer to A. Second, if an alliance is proposed, then state A decides whether to accept or reject the proposed alliance and transfer. If A accepts then an alliance is formed, denoted $a = 1$, and if A rejects, or P does not propose an alliance, then there is no alliance, $a = 0$.[5]

Next, State E observes if a public alliance was formed, otherwise, it observes nothing. Specifically, E observes an outcome determined by the function $\varphi : \{Pub, Sec\} \times \{0,1\} \to \{Pub, \emptyset\}$, where $\varphi(Pub, 1) = Pub$ and $\varphi(x, a) = \emptyset$ for any $(x, a) \neq (Pub, 1)$. Note that $\varphi \in \{Pub, \emptyset\}$ is the only information E observes, thus, it does not observe the transfer t or other details of the alliance bargaining process. Finally, State E decides to attack or not. If E attacks we say war occurs, if not, then we say the outcome is peace.

Payoffs: For $i \in \{P, A, E\}$ player i's payoffs from peace and war are determined, respectively, by (Borel measurable) functions $\pi_i : \Theta \to \mathbb{R}$ and $\omega_i : \Theta \times \{0,1\} \to \mathbb{R}$. Therefore, payoffs depend on P's type and, in the event of war, whether P and A formed an alliance. P and A also value the transfer t. E's final payoff is $\omega_E(\theta, a)$ if war occurs, and its payoff is $\pi_E(\theta)$ if peace prevails. P's payoff from war is $\omega_P(\theta, a) - at$, while its payoff if the outcome is peace is $\pi_P(\theta) - at$. On the other hand, A's payoff if war occurs is $\omega_A(\theta, a) + at$ and its peace payoff is $\pi_A(\theta) + at$.

We impose the following assumptions on the war and peace functions of each player. In the next section, we provide examples that satisfy these assumptions.

We assume an alliance increases P's payoff in the event of war, $\omega_P(\theta, 1) > \omega_P(\theta, 0)$ for all $\theta \in \Theta$. Further, we assume that P's payoff from war relative to peace is increasing in its type, $\omega_P(\theta, a) - \pi_P(\theta, a)$ is strictly increasing in θ for all $a \in \{0, 1\}$. Additionally, we assume that the alliance is more beneficial for lower types than for higher types, $\omega_P(\theta, 1) - \omega_P(\theta, 0)$ is weakly

[5] We assume P proposes the alliance, however, our results do not depend on this assumption. If A instead proposes the alliance it has the same incentives as P over the type of alliance, because this maximizes the size of the transfer P is willing to pay to A. We discuss this alternative set-up further in Appendix B.

decreasing in θ. Finally, to focus on understanding the emergence of defensive secret alliances, we assume that P always prefers peace to war, $\pi_P(\theta) \geq \omega_P(\theta, 1)$ for all $\theta \in \Theta$.

Although it is not necessary for our results, to make the alliance formation stage non-trivial we assume A prefers to not intervene on P's behalf absent a transfer, $\omega_A(\theta, 1) < \omega_A(\theta, 0)$ for all $\theta \in \Theta$. Additionally, we assume that A's payoff for conflict relative to peace is strictly increasing in P's type, $\omega_A(\theta, a) - \pi_A(\theta, a)$ is increasing in θ for $a \in \{0, 1\}$. Finally, we assume A is more willing to intervene on behalf of lower types, $\omega_A(\theta, 1) - \omega_A(\theta, 0)$ is weakly decreasing in θ.

For state E, we assume the alliance decreases E's payoff from fighting, $\omega_E(\theta, 1) < \omega_E(\theta, 0)$ for all $\theta \in \Theta$. Additionally, E's payoff from war relative to peace is decreasing in P's types, $\omega_P(\theta, a) - \pi_P(\theta, a)$ is strictly decreasing in θ for all $a \in \{0, 1\}$. Finally, E always prefers to fight the lowest type of state P, $\pi_E(\underline{\theta}) < \omega(\underline{\theta}, 1)$, and prefers to not fight the highest type, $\pi_E(\overline{\theta}) > \omega_E(\overline{\theta}, 0)$.

Special Cases

The payoffs and type space in our model are sufficiently general to admit several possible microfoundations. Below we describe two special cases.

Two-type model: Throughout we use a version of the model with just two types, $\Theta = \{\underline{\theta}, \overline{\theta}\}$, as an example to highlight equilibrium conditions. To fix notation, let E's prior belief about P be $Pr(\theta = \overline{\theta}) = \mu_0$ and $Pr(\theta = \underline{\theta}) = 1 - \mu_0$ for $\mu_0 \in (0, 1)$.

We interpret the $\overline{\theta}$ type as having interests aligned with E, the $\underline{\theta}$ type as having interests aligned with A, and A and E as having opposed interests. Thus, E does not want to attack P if it has aligned interests with E, but does want to attack if P is aligned with A. To further reinforce this interpretation, we assume:

$$\omega_P(\overline{\theta}, 1) - \omega_P(\overline{\theta}, 0) < \omega_A(\overline{\theta}, 0) - \omega_A(\overline{\theta}, 1).$$

This assumption implies the $\overline{\theta}$ type is sufficiently aligned with E that it is prohibitively costly for P to obtain A as an ally in the event of war. Similarly, we assume that the $\underline{\theta}$ type is sufficiently aligned with A that it is able to form an alliance with A if it anticipates conflict. Specifically, $\omega_P(\underline{\theta}, 1) - \omega_P(\underline{\theta}, 0) > \omega_A(\underline{\theta}, 0) - \omega_A(\underline{\theta}, 1)$.

Alignment model: One way to microfound the general payoffs in our model is to conceive of the interaction as one about settling a disputed international policy. While we use the term policy to describe the issue under disagreement, these payoffs can be interpreted generally as representing differences in preferences over many types of actions a state might take. Assume if E does not attack then P chooses a policy $y \in [0, 1]$. If E does attack then war occurs, which we model as the standard costly lottery. In particular, P wins with probability $p(a)$, with $p(1) > p(0)$, and E wins with probability $1 - p(a)$. In this case, the winner of the war chooses a policy $y \in [0, 1]$. Additionally, player i incurs a cost of war $c_i(a) > 0$ for $a \in \{0, 1\}$. A's preferred action is given by $\hat{y}_A = 0$, E's preferred action is $\hat{y}_E = 1$, and P's preferred action is $\hat{y}_P = \theta$, with $\theta \in \Theta = [0, 1]$. Therefore, θ captures how aligned P's interests are with A versus E. Assume each player i has quadratic preferences over policies, i.e., $u_i(y) = -(y - \hat{y}_i)^2$.

Clearly, whoever eventually chooses the policy y will set $y = \hat{y}_i$. This yields

$$\begin{aligned}
&\pi_i(\theta) = -(\theta - \hat{y}_i)^2 \\
&\omega_i(\theta, a) = -p(a)(\theta - \hat{y}_i)^2 - (1 - p(a))(1 - \hat{y}_i)^2 - c_i(a).
\end{aligned}$$

Under some additional assumptions on the costs of war, $c_i(a)$, this set-up satisfies our assumptions on payoffs.

Comments on the Model

An important feature of our model is that P offers a transfer, t, to A in exchange for their alliance commitment. This captures the idea that alliance formation involves allies making concessions to one another (Snyder, 2007). Empirically, these transfers take a wide variety of forms including direct payments that compensate an ally for the costs of fighting and are concessions on unrelated political or economic issues valued by an ally (Henke, 2019a,b). Previous work has also shown that the possibility of such transfers play an important role in shaping communication among potential allies (Smith, 2021), as well as in shoring up alliance commitments when moral hazard is a concern (Benson et al., 2014).

We also assume the alliance is formed prior to conflict. Thus, if A accepts then it is committed to fighting in the event of conflict and P is committed to making the transfer even if a conflict does not arise. By assuming that A is committed to fighting if an alliance is formed we abstract from issues of alliance credibility and create conditions favorable to deterrence. We could instead assume that transfers and support are contingent on whether conflict occurs, as in, e.g., Benson et al. (2014). This change would have minimal impact on our results because transfers under pure strategy equilibria are already chosen correctly anticipating the outbreak of war, and pure strategy equilibria always exist in our model.

Overall, we abstract from any differences between types of alliances in order to clearly isolate the effects of secrecy. As such, our theoretical approach is experimental in the sense described by Paine et al. (2020), aiming to hold all else constant to focus on explicating a particular causal mechanism. We now discuss some of these possible differences.

First, we assume that the increased capabilities from forming an alliance and the cost function for transfers are the same if the alliance is public or secret. In reality, there may be frictions that make it more or less difficult to form a secret alliance and thus alter the effectiveness of alliance. However, it is unclear whether these differences favor secret or public alliances, and including such idiosyncratic features in the model would complicate our findings without adding new insights.

Second, we assume that public and secret alliance commitments are similarly credible. Surveying the literature, Morrow (2000) shows that allies might make their commitments credible through a variety of mechanisms. For example, alliance commitments may be credible due to the interaction being part of a larger repeated game in which states value a reputation for reliability. In such a setting, secret alliances may be more difficult to enforce because they are unobserved. However, given our interest in providing an explanation for why states may form secret alliances in the first place, we abstract from this question. Moreover, states presumably enter into secret alliances anticipating the alliance is credible with some probability. In terms of our model, commitments may be (reasonably) credible because alliances are formed between states with similar interests. We could also model credibility by assuming the transfer is contingent on the ally joining in the event of war, which would not alter our results.[6] Finally, the empirical record suggests that alliance credibility was at its lowest post-1945, a period in which secret alliances have been relatively rare (Berkemeier and Fuhrmann, 2018). As such, it is unclear empirically whether secret alliances are more or less reliable. There is also reason to believe that private diplomacy can enhance credibility in general (Kurizaki, 2007) as well as in the specific context of alliance and coalition formation (Smith, 2021). An interesting avenue for future research is to study how and when secret alliance commitments are credible relative to public alliances.

Results

We now characterize equilibrium behavior in our model. Our solution concept is perfect Bayesian equilibrium (PBE). We add the additional refinement that if P is indifferent between forming an alliance or not then it chooses to not form the alliance. This refinement captures behavior that is robust to a small exogenous cost to forming an alliance.[7] Henceforth, we refer to PBE that satisfy this refinement as "equilibrium."[8]

We begin by analyzing E's decision to attack. After observing either a public alliance or nothing, E updates its beliefs about both P's type and whether P and A have formed an alliance (using

[6] We could also extend the model to allow for some probability that if a secret alliance is formed there is some probability the alliance partner does not follow through on its commitment. While this would shift states' incentives, it would not undermine the core logic of our mechanism.

[7] Specifically, this refinement could be formalized by assuming there is an exogenous cost for forming an alliance, and selecting equilibria that are limit points of the equilibrium correspondence as this cost goes to 0.

[8] Note, the statements below should be taken as holding for all but a measure zero set of parameter values.

Bayes' rule whenever possible), and then chooses to attack or not. We can represent a (mixed) strategy for E as a mapping $\rho : \{\emptyset, Pub\} \to [0,1]$. Let $\rho(\varphi)$ be the probability E attacks after seeing outcome $\varphi \in \{\emptyset, Pub\}$.

Therefore, E strictly prefers to attack, $\rho(\varphi) = 1$, if

$$\mathbb{E}_\theta\Big[\mathbb{E}_a[\omega(\theta, a)|\theta]\Big|\varphi\Big] > \mathbb{E}_\theta[\pi_E(\theta)|\varphi]. \tag{1}$$

If (1) is reversed, then E must not attack in equilibrium, $\rho(\varphi) = 0$. Finally, if (1) holds with equality then E is indifferent and can attack with any probability, $\rho(\varphi) \in [0,1]$. Note that after observing a public alliance, $\varphi = Pub$, state E is certain an alliance has formed, which simplifies (1) to $\mathbb{E}_\theta[\omega(\theta, 1)|Pub] > \mathbb{E}_\theta[\pi_E(\theta)|Pub]$.

Moving back from E's decision to attack, our first lemma characterizes the equilibrium transfer from P to A when the states form an alliance.

Lemma 1. *In any equilibrium, if P proposes an alliance of type $x \in \{Pub, Sec\}$ and A accepts, then the transfer is given by:*

$$t^*(\theta) = \max\Big\{\rho(\emptyset)(\omega_A(\theta, 0) - \pi_A(\theta)) - \rho(\varphi(x, 1))(\omega_A(\theta, 1) - \pi_A(\theta)), 0\Big\}$$

The transfer must compensate A for fighting in the event of conflict. However, the magnitude of the transfer is adjusted by the risk of war, which is endogenously determined by E's equilibrium behavior. For secret alliances, this adjustment is mitigated because E observes nothing whether such an alliance is formed or not. However, if, for example, a public alliance is fully deterrent in equilibrium then $t^* = 0$ because there is no risk of war. On the other hand, if E is more likely to attack after observing a public alliance than after observing nothing, then this increases the payment that P must make to A.

We see from Lemma 1 that different types of P face different costs for forming an alliance. Additionally, the relative costs for fighting vary for P depending on its type θ. Consequently, P's willingness to form an alliance depends on its type. In particular, given a fixed probability of war, the types of P that prefer to *not* form an alliance are defined by the following set:

$$\Theta^* = \Big\{\theta \in \Theta \Big| \omega_P(\theta, 1) - [\omega_A(\theta, 0) - \omega_A(\theta, 1)] \leq \omega_P(\theta, 0)\Big\}. \tag{2}$$

This set is clearest in the context of the alignment model, which we depict in Figure 1. Recall, in this setting, E and A have opposite interests, with E's preferred action at 1 and A's preferred action at 0. When θ is close to 1 it is very costly for P to obtain A's support and there is not much downside to P for losing a conflict to E. Thus, such a type of P prefers to forgo forming an alliance. The opposite logic holds when θ is close to 0, in this case P and A are relatively aligned. This implies there is a θ^* such that all types $\theta > \theta^*$ do not form an alliance and all types $\theta < \theta^*$ would choose to make the transfer $t^*(\theta)$ and form an alliance. Of course, Θ^* is constructed assuming that forming an alliance does not alter the probability of conflict, which may not be true in equilibrium. However, as we show, the set Θ^* still plays an important role in our equilibrium analysis.

Figure 1: Alliance Decision

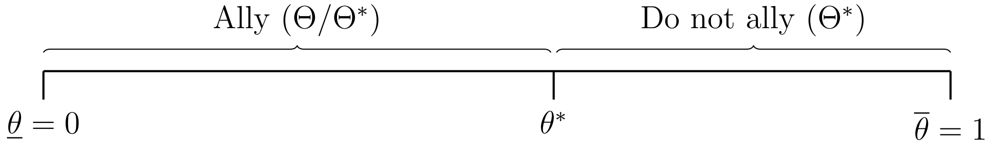

Figure 1 depicts the decision to form an alliance or not for each type θ, given a fixed probability that E attacks.

Before proceeding to study equilibrium alliances, we first introduce several definitions that distinguish different types of equilibria that can arise.

Definition 1. *An equilibrium is a secret alliance equilibrium if there is a positive probability that a secret alliance is formed on the path of play.*

Definition 2. *An equilibrium is a public alliance equilibrium if there is a positive probability that a public alliance is formed and a secret alliance is formed on the path of play with probability* 0.

Definition 3. *A secret alliance equilibrium is non-trivial if there does not also exist a public alliance equilibrium that is payoff equivalent for all types $\theta \in \Theta$.*

Definitions 1 and 2 differentiate between equilibria in which secret alliances occur versus those in which the only alliances are public. Definition 3 further refines secret alliance equilibrium to consider equilibrium outcomes that can only be explained using secret alliances. For example, this definition serves to rule out equilibria in which all types form a secret alliance on the path of play. In this case, though the alliance is nominally secret, E's equilibrium conjecture tells it that there is guaranteed to be an alliance. Therefore, we could also construct an equivalent equilibrium in which every type forms a public alliance. Thus, this serves to restrict our attention to equilibria in which secret alliances are associated with meaningful uncertainty on the path of play. More generally, for our theory to provide an explanation for the emergence of secret alliances this behavior should not be equally rationalizable using public alliances.

With these definitions in hand, Lemma 2 shows that if secret alliances occur in our setting it must be due to signaling concerns.

Lemma 2. *If there is complete information then a non-trivial secret alliance equilibrium does not exist.*

If E knows θ then it is not possible to sustain (meaningful) secret alliances in equilibrium. In our setting, P is (generically) never indifferent between no alliance and a secret alliance. This is because the cost of an alliance is endogenous to the probability of conflict. In particular, if E attacks with a higher probability then the increased gain from forming an alliance is offset by the need to provide a larger transfer to A.[9] Consequently, with complete information E can always perfectly predict whether P and E have an alliance in equilibrium and any behavior driven by secret alliances can also be explained using only public alliances.

Existence of Secret Alliances

We now provide a series of results characterizing equilibrium. We describe the conditions under which secret alliances cannot occur, when they can be sustained in equilibrium and, finally, when only secret alliance equilibrium exist. Throughout, we include results for the two-type case of the model to clearly highlight the parameters under which these different behaviors emerge.

To start, we introduce the following condition:

$$\mathbb{E}_\theta[\omega_E(\theta, 0)] \le \mathbb{E}_\theta[\pi_E(\theta)]. \tag{3}$$

If (3) holds then E prefers not to attack P even when every type of P does not form alliance. In this case, E expects that P's interests are aligned with its own. Thus, when condition (3) holds we say that P is *friendly.* In the two-type model, this condition can be expressed as:

$$\mu_0 \ge \frac{\omega_E(\underline{\theta}, 0) - \pi_E(\underline{\theta})}{[\omega_E(\underline{\theta}, 0) - \pi_E(\underline{\theta})] - [\omega_E(\overline{\theta}, 0) - \pi_E(\overline{\theta})]} \equiv \overline{\mu}. \tag{3*}$$

Recall that μ_0 is P's prior belief that $\theta = \overline{\theta}$ in the two-type model. Thus, condition (3*) holds when E believes P is sufficiently likely to be the friendly type.

This leads to our next lemma, which highlights when secret alliances cannot be sustained in equilibrium.

Lemma 3. *If state P is friendly then a secret alliance equilibrium does not exist. There exists an equilibrium in which all types do not form an alliance and E never attacks on the path of play.*

Intuitively, it is not worth it to E to incur the costs of fighting when E is sufficiently certain that P is friendly. Consequently, there exists a peaceful equilibrium where no type of P has an incentive to deviate and form an alliance. Moreover, because P is friendly there is never an equilibrium in which E attacks on the path of play. Thus, any strategy profile where some types form a

[9] This differentiates secret alliances from models where states make unobserved arming decisions, because the costs of investing in arms is exogenous to the risk of war (e.g., Meirowitz and Sartori, 2008).

secret alliance is ruled out by our equilibrium refinement, because such a type could instead form no alliance without changing its equilibrium payoff. Further, any such strategy profile would be a trivial secret alliance equilibrium, as all types could instead form a public alliance and not be attacked, and this would be payoff equivalent.

We next provide conditions under which secret alliances exist. To do so, we say that P is *antagonistic* if the following inequality holds:

$$E_\theta[\omega_E(\theta, 1)] \geq E_\theta[\pi_E(\theta)]. \tag{4}$$

Under this condition, E prefers to fight even if it knows with certainty that P has formed an alliance. Substantively, this means that the expected conflict of interest between E and P is so great that an alliance is not sufficient to deter aggression from E. In our two-type model, this implies E believes that P is ex-ante unlikely to be the friendly $\overline{\theta}$ type. Specifically,

$$\mu_0 \leq \frac{\omega_E(\underline{\theta}, 1) - \pi_E(\underline{\theta})}{[\omega_E(\underline{\theta}, 1) - \pi_E(\underline{\theta})] - [\omega_E(\overline{\theta}, 1) - \pi_E(\overline{\theta})]} \equiv \underline{\mu}. \tag{4*}$$

If the antagonistic condition (4) does not hold then alliances are strongly deterrent, or P and E are somewhat likely to be aligned. The following result shows that, in this case, an equilibrium exists in which all types join an alliance and war does not occur.

Lemma 4. *If P is not antagonistic, then there exists an equilibrium in which all types receive a public alliance and the probability of war is 0.*

This finding is intuitive: if P is not antagonistic, then an alliance does not send a negative signal to E. As a result, it is possible to construct an equilibrium in which all types form a public alliance. In this equilibrium, upon observing a public alliance E does not attack, and so the alliance is "free" as no transfer is required to induce A to accept an alliance with P. However, if P is antagonistic, then such an equilibrium is not possible. Instead, there exists a non-trivial secret alliance equilibrium. The following result establishes the claim that secret alliances can occur under these conditions.

Lemma 5. *If P is antagonistic then there exists a non-trivial secret alliance equilibrium. Moreover, there exists such an equilibrium in which all $\theta \in \Theta^*$ do not form an alliance, all $\theta \notin \Theta^*$ form a secret alliance, and E always attacks after seeing nothing. That is, all alliances formed on the path of play are secret.*

To see the logic for why secret alliances can be sustained under this condition, conjecture that all types of P choose either to form a secret alliance or no alliance. When P is antagonistic E is incentivized to attack after seeing nothing because it expects that P has strongly opposed interests. Anticipating conflict, some types of P want to form a secret alliance even though this does nothing to deter E. However, the types that form an alliance are those most opposed to E, thus, the existence of secret alliances does not dissuade E from attacking. To understand why this equilibrium is non-trivial consider a strategy profile where all the types who form a secret alliance instead form a public alliance. Although the types who form an alliance are still always attacked, now seeing nothing sends too good of a signal and E never attacks. Consequently, those forming a public alliance can profitably deviate to no alliance.

Lemma 5 and the preceding discussion further highlight that the equilibrium emergence of secret alliances is driven by uncertainty about P's type. In particular, there is a secret alliance equilibrium in which, with probability 1, each type forms an alliance or not. Thus, although we allow for mixing by P and by E, strategic uncertainty generated by mixing is not what drives secret alliance equilibria.[10]

Although Lemma 5 demonstrates that meaningful secret alliances can arise, it does not rule out the possibility that public alliance equilibria also exist. As such, it does not give a sharp characterization for when we should most expect the emergence of secret alliances. Our first proposition shows when secret alliances are a pervasive feature of the conflict environment. In particular, we show that it depends on the following condition:

$$\mathbb{E}_\theta[\omega_E(\theta, 0) | \theta \in \Theta^*] \leq \mathbb{E}_\theta[\pi_E(\theta) | \theta \in \Theta^*]. \tag{5}$$

If inequality (5) holds we say that P and E are *conditionally mutually friendly*. In this case, conditional on E knowing that $\theta \in \Theta^*$, E prefers peace to war. That is, if P is a type that does not want to form an alliance against E, then E also does not want to fight P. Thus, under this condition E's friendliness towards P is conditional on knowledge that P is a type that does not wish to form an alliance against it. In this sense, (5) indicates that E is peaceful, conditional on its knowledge that this friendly orientation is mutual with P.[11]

With this inequality defined, we can now characterize the conditions under which every equilibrium must feature secret alliances.

Proposition 1. *Every equilibrium is a secret alliance equilibrium if and only if P is antagonistic and P and E are conditionally mutually friendly.*

To understand why all equilibria involve secret alliances under these conditions it is useful to consider why there cannot be an equilibrium that does not involve secret alliances. Suppose that

[10] Indeed, as noted in the discussion following Lemma 2, for any probability of attack (almost) all types θ strictly prefer to form an alliance or no alliance. Thus, any mixing in equilibrium by a type θ must be over the type of alliance, rather than the forming of an alliance.

[11] Note this condition is always satisfied under the assumptions of the two-type model.

P is antagonistic, and that P and E are conditionally mutually friendly. Now suppose there is an equilibrium in which some types form public alliances and some types do not form an alliance, but no type forms a secret alliance. If this were the case, then a public alliance sends a strong negative signal to E and observing nothing sends a relatively positive signal, under the conditions in the proposition. Consequently, E attacks after observing a public alliance but does not attack after observing nothing. However, this incentivizes types who are forming a public alliance to deviate, as such a deviation means they are not attacked and do not have to pay a transfer to A. Thus, such a profile cannot be an equilibrium under these conditions. As a consequence, in order to avoid sending a strong negative signal, some types must be forming a secret alliance in equilibrium.

To further clarify the conditions under which secret alliances emerge in equilibrium we conclude this section with an analysis of the two-type model. Proposition 2 characterizes all equilibria of the two-type model.

Proposition 2. *There exists $\mu^* \in (\underline{\mu}, \overline{\mu})$ such that:*

- *There is an equilibrium in which the $\underline{\theta}$ type forms a secret alliance and the $\overline{\theta}$ does not form an alliance if and only if $\mu_0 \leq \mu^*$. On the path of play, E attacks with probability 1 after seeing nothing.*
- *There is an equilibrium in which both types form a public alliance if and only if $\mu_0 \geq \underline{\mu}_0$. On the path of play, E never attacks after seeing a public alliance.*
- *There is an equilibrium in which neither type forms an alliance if and only if $\mu_0 \geq \overline{\mu}$. On the path of play, E never attacks after seeing nothing.*

No other equilibria exist.

Proposition 2 highlights that secret alliances arise when E is highly suspicious that P is aligned with A. Otherwise, if E believes that P is likely to have shared interests, then either both types form a public alliance or no type forms an alliance. Notice there is an intermediate region in which there exists both a secret and a public alliance equilibrium. Figure 2 depicts the characterization of equilibrium behavior described in the proposition.

Conflict and Deterrence in Equilibrium

We now study the relationship between secret alliances and conflict in equilibrium. Although the multiplicity of equilibria makes crisp comparative statics difficult to obtain, our results suggest an association between the existence of secret alliances and a high probability of conflict.

To start, we study the probability of war within an equilibrium. That is, we examine the probability E attacks after observing nothing compared to seeing a public alliance.

Figure 2: Equilibrium Characterization Two-types

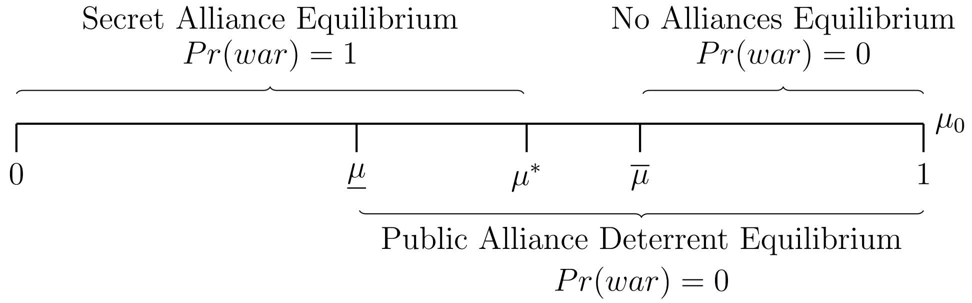

Figure 2 characterizes equilibria of the two-type model as a function of E's prior belief that P is the friendly $\overline{\theta}$ type, $Pr(\theta = \overline{\theta}) = \mu_0$.

Proposition 3. *Assume the P antagonistic condition holds. In every equilibrium $\rho(\emptyset) \geq \rho(Pub)$. In every secret alliance equilibrium $\rho(\emptyset) = \rho(Pub) > 0$.*

Proposition 3 demonstrates that the probability E attacks after seeing nothing must be at least as high as the probability it attacks if P and A form a public alliance. If seeing nothing deterred E, relative to seeing a public alliance, then any types who formed a public alliance would do strictly better by instead forming a secret alliance (or no alliance). Therefore, although secret alliances only exist because P wants to avoid signaling misalignment with E, forming a secret rather than public alliance cannot decrease the probability of war in equilibrium. Indeed, in any secret alliance equilibrium the probability of war must be strictly positive, otherwise P could profitably deviate to forming no alliance. Moreover, secret alliances also undermine the ability of public alliances to deter. If not, then a type choosing to form a secret alliance would instead form a deterrent public alliance.

Our final result studies the probability of conflict across different regions of the parameter space and equilibria. In particular, we compare the case where secret alliances do not exist to the case where public alliances do not exist.

Proposition 4.

1. *If a secret alliance equilibrium does not exist, then there exists a public alliance equilibrium where the probability of war is* 0.
2. *If a public alliance equilibrium does not exist, then there exists a secret alliance equilibrium where the probability of war is* 1.

Proposition 4 shows there is always an equilibrium in which war occurs with certainty, under the conditions where secret alliances are most prevalent. To see why, consider a strategy profile where all the relatively friendly types of P, $\theta \in \Theta^*$, do not form an alliance, and all the relatively opposed types, $\theta \notin \Theta^*$ form a secret alliance. By Proposition 1, if a public alliance equilibrium does not exist then the P is antagonistic condition must hold. Thus, after seeing nothing E attacks P with certainty. By construction, for types in Θ^* deviating to form an alliance is not worth the cost, while types outside Θ^* are willing to make the transfer to A. In contrast, in the cases where secret alliances never form, there is always an equilibrium that is peaceful. By Lemma 5, in this case the antagonistic condition must fail. Thus, if all types form a public alliance then E is deterred and does not attack. Moreover, because E does not attack, the alliance is free for all types of P and, hence, no type has an incentive to forming no alliance and (potentially) being attacked off the path of play.

Taken together, Propositions 3 and 4 imply a positive correlation between secret alliances and conflict. This implication is derived from both a within-equilibrium comparison and an across-equilibrium comparison. First, as Proposition 3 demonstrates, the correlation holds when comparing secret and public alliances *within* a particular equilibrium. In any equilibrium, the probability of war will be weakly higher following a secret alliance than a public alliance. Further, the existence of non-trivial secret alliances precludes the possibility of concurrent peaceful, deterrent public alliances. Second, as shown in Proposition 4, the correlation holds when comparing *across* equilibria. To do this, we focus on cases where only public and secret alliance equilibria exist, respectively. If the only alliances that can be supported in equilibrium are public, then a peaceful, deterrent equilibrium is guaranteed to exist. In contrast, if every equilibrium involves a secret alliance, then every equilibrium has a positive probability of war and there exists an equilibrium in which war is guaranteed. Thus, drawing on both across and within-equilibrium comparisons, our results imply that secret alliances are positively associated with conflict.

In sum, although states keep alliances secret because of an incentive to conceal their interests and avoid provocation, equilibrium forces render this ineffective at avoiding conflict. Consequently, in equilibrium, secret alliances are associated with war, rather than peace. In the following section, we discuss this pattern and others in the context of the historical record.

Discussion

We now turn to discuss the implications of our theory for understanding patterns of secret alliances. Throughout, we use the ATOP data (Leeds et al., 2002) to assess these patterns. Of course, we may not observe all the secret alliances that occur in a given year as, by their nature, secret alliances are secret. Thus, the data should be taken with this caveat. However, we believe these patterns are still informative for several reasons, which we discuss in Appendix C.

The Decline of Secret Alliances

We begin by analyzing the prevalence of secret alliances. Figure 3 plots the percent of alliances that are secret in each year. We see there is a stark difference in the frequency of secret alliances across time periods. In particular, there is a steep decline in secret alliances after World War I and they virtually disappear after World War II. Our model points to two factors that align with this observed drop-off.

Figure 3: Secret Alliances over Time

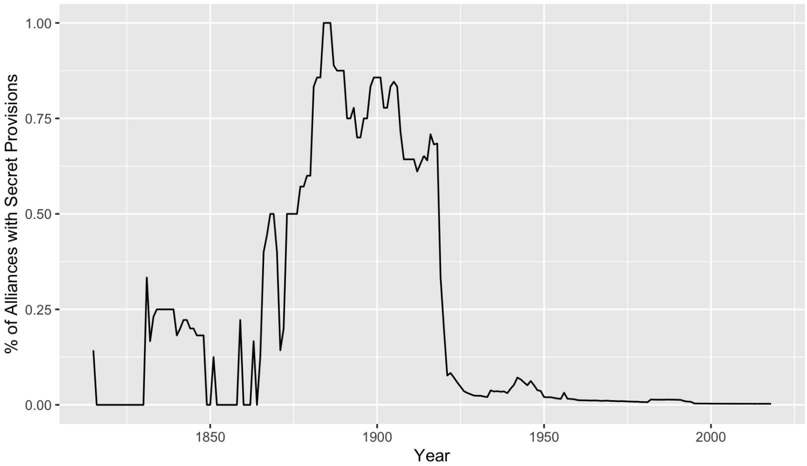

Figure 3 plots the percent of alliances each year with active secret provisions.

First, our theory suggests a connection between uncertainty and secret alliances. As Lemma 2 illustrates, secret alliance equilibria do not exist under complete information. Thus, our theory suggests that a potential mechanism driving the decline in secret alliances over time is a corresponding decline in uncertainty over the same time period. What could account for a reduction in uncertainty? One possibility is the broad shift in the polarity of the international system. During the period in which secret alliances were most prevalent, international relations were characterized by multipolarity, with the distribution of power being spread relatively evenly across a relatively large number of states. This changed in the era after World War I, and was solidified by the shift to bipolarity in the wake of World War II, in which the preponderance of military power was concentrated in two states: the U.S. and Soviet Union. Polarity has a natural connection to the kind of uncertainty embedded in our model.

Uncertainty about a state's alignment often amount to uncertainty about which "side" of a given issue a state is on. Under bipolarity, it is relatively easy to identify where each state stands because power is concentrated in two states - typically in alignment with one and only one of the major powers (Beres, 1972). This pattern emerged in the period after World War II, where the vast majority of states came into alignment with either the U.S. or the Soviet Union. Theorists have previously argued that bipolar systems may be more stable and less war-prone than multipolar systems (Waltz, 1964). In contrast, in the period before World War I, the international system was characterized by high uncertainty about alignments, motives, and intentions. Scholars have previously argued that the multipolarity of the international system during this time fostered such uncertainty and instability (Waltz, 1979).

Our findings, specifically Lemma 4 also point to a second mechanism that may have contributed to the decline in secret alliances over time: an increase in the deterrent value of alliances. Looking to the historical record, the emergence of nuclear weapons coincided with the virtual disappearance of secret alliances after the second world war. There is a wealth of theoretical and empirical evidence that nuclear weapons contribute significantly to extended deterrence (Huth, 1990; Fuhrmann and Sechser, 2014). Given these findings, it is clear that states allied with nuclear-capable states are extremely unattractive targets for aggression. To see the connection to our model, recall condition 4. When this condition fails to hold, as is the case when alliances are highly deterrent, there exists an equilibrium in which all alliances are public and the probability of war is 0. Thus, the emergence of nuclear weapons, and the corresponding increase in the deterrent value of nuclear-capable alliances, is another mechanism that plausibly accounts for the demise of secret alliances under our theory.

Secret Alliances and Conflict

Propositions 3 and 4 study the equilibrium probability of conflict in our model. These results make predictions about the empirical relationship between alliances and the onset of conflict. We now discuss two of the implications of these propositions and show they are consistent with patterns in the data.

First, our theory is consistent with a positive association between the overall number of secret alliances that exist in a given year and the number of disputes in the international system. Taken as a whole, our results indicate that equilibria involving secret alliances generally have higher rates of conflict than equilibria that involve no secret alliances. To see this most clearly, recall Proposition 4. This result shows that if a secret alliance equilibrium does not exist, then there always exists a peaceful public alliance equlibrium. Further, the result also shows that if a public alliance equilibrium does not exist, then there exists a secret alliance equilibrium in which the probability of war is 1. This implies that historical periods with a high prevalence of secret alliances should have a high incidence of conflict.

To evaluate this implication, we turn to the data. We begin by gathering data on the total number of defensive secret alliances in a given year according to the Alliance Treaties, Obligations, and Provisions (ATOP) data (Leeds et al., 2002). With this variable in hand, we turn to the Militarized Interstate Dispute (MID) data to capture conflict (Palmer et al., 2022). Using these data, we sum the total number of MIDs in a given year. Consistent with our theory, we find a positive and statistically significant relationship between these two variables, with an estimated correlation coefficient of 0.05132 and a 95% confidence interval of $[0.0363, 0.0662]$.

This finding is consistent with existing research pointing to a positive correlation between secret alliances and conflict onset (Bas and Schub, 2016). This previous work has found a relationship between secret alliances and conflict at the dyadic level. In Bas and Schub (2016)'s framework, the mechanism connecting secret alliances and conflict is asymmetric information. Secrecy generates inconsistent estimates of strength, which leads to conflict via the standard arguments derived from bargaining models, exemplified by Fearon (1995). As such, a dyadic specification is appropriate for evaluating their causal argument that posits a directional relationship that flows from secret alliances to observed conflict.

In contrast, our theoretical argument does not advance such a unidirectional, causal claim. Rather, our argument is that *both* secret alliances *and* conflict are the result of a third, system-level factor: uncertainty about the constellation of interests in the international system. As such, our theory does not generate a causal claim about the relationship between secret alliances and war. Rather, it simply points to a system-level correlation, which we have found. To be clear, we do not view our argument and that of Bas and Schub (2016) as mutually exclusive. Rather, we believe that it is likely that both mechanisms operate in the data. For this reason, for our next test we turn to evaluation of a pattern in the data that is not consistent with dyadic-level explanations for the connection between secret alliances and war.

Second, our theory also points to a subtle connection between the existence of secret alliances and the deterrent value of concurrent public alliances. Proposition 4 indicates that when secret alliance equilibria do not exist, there always exists a public alliance equilibrium in which the probability of war is 0. In contrast, combining Propositions 1 and 3 reveals that, under conditions in which only secret alliance equilibria exist, the probability of war following a public alliance on the path of play is strictly positive. Taken as a whole, our results allow us to form a prediction about the correlation between (i) the *total* number of secret alliances in the international system and (ii) the probability of attack against any *particular* public alliance. In particular, these two results suggest an empirical correlation between the total number of secret alliances and the probability that any given public alliance experiences conflict.

To assess this correlation, we again turn to both the ATOP and MIDs data. Our unit of analysis is the country-year, and we restrict our sample to country-years in which the given country was a member of a public defensive alliance according to ATOP. For each such country-year, we calculate two quantities. First, we calculate the *total* number of active secret alliances in the international system in the year corresponding to each country-year observation. Second, for each country-year we calculate the total number of militarized interstate disputes initiated against that country in the given year, with dispute data drawn from the MIDs 5.0 data.

Consistent with our expectation, we find a positive and statistically significant correlation, with an estimated correlation coefficient of 0.1628, and 95% confidence interval of $[0.1317, 0.1935]$. This correlation is consistent with the implications of our theory, which predict an association between the total number of secret alliances and the deterrent value of concurrent public defensive alliances. Before we continue, it is worth emphasizing the correct interpretation of this finding. To be clear, we do not argue that this relationship is causal. To the contrary, our model suggests that the relationship is *not* causal. Our model instead suggests that a third factor, prevailing beliefs and suspicions about the arrangement of interests in the international system, drives *both* the prevalence of secret alliances *and* a concurrent rise in attacks against members of public defensive alliances. During times of high suspicion, an analyst should observe *both* high rates of secret alliances *and* higher than typical rates of attacks against members of public defensive alliances. Thus, while this does not provide confirmatory evidence of our model, this pattern is consistent with the predictions of our theory. Further, we believe that this empirical pattern is not straightforward to explain without the guidance of our theoretical arguments.

Finally, returning to the late 19th century Bismarckian system of secret alliances which we discussed previously, our interpretation of this correlational evidence is consistent with historians' interpretation of the relevant diplomatic motives. For example, in 1883 Romania entered into a secret defensive alliance with Austria-Hungary. Romania desired such an alliance to shore up its defenses against Russia. Otto von Bismarck, who brokered the alliance deal, emphasized secrecy during the negotiation of the pact. Oldson describes the motivation for secrecy as driven by Bismarck's desire to "...fashion an alliance system based on the desire to avoid war, and at the same time, to prepare for war." (1977[p. 232]). Anticipating that concrete knowledge of this alliance might alert Russia to Romanian aims, Bismarck urged Romania and Austria-Hungary to keep the alliance secret, which they did by including a provision for secrecy in Article VI of the Austro-Romanian Treaty of 1883 (Pribram, 1921)[p. 78-82]. This motivation for secrecy was not isolated to this alliance. Characterizing Bismarck's secret correspondence with German diplomats during this period, and emphasizing the role of secrecy in the overall system of alliances, Zerner writes that "Thus, between 1879 and 1887 the German chancellor managed to fashion an alliance system based on the desire to avoid war and, at the same time, to prepare for war." (1968[p. 232]). Thus, our understanding for the motivation behind secret alliances is broadly consistent with historians' interpretations of diplomatic motives during the period in which secret alliances were most prevalent. Furthermore, while these alliance were kept secret to try and avoid provocation, they were formed with the understanding that the environment was characterized by a high level of suspicion. Thus, these alliances anticipated and prepared for war, consistent with our theoretical findings on secret alliances and conflict.

Robustness

We now briefly discuss the robustness of our findings to several extensions of the model. Appendix B contains formal results and additional discussion of the extensions. For our extensions, we focus on the two-type version of the model.

To start, we show that our main results are robust to two different changes in the informational environment. First, we let the transfer t be observed by E when a public alliance is formed. Here, the public transfer may signal additional information about P's alignment. Second, we consider an extension where A does not observe θ, and thus is uncertain about P's alignment. Recall that A is willing to enter an alliance with the $\underline{\theta}$ type for a lower transfer than it is with the $\overline{\theta}$ type, as the $\underline{\theta}$ type is more aligned with A. Consequently, now the $\overline{\theta}$ type may be tempted deviate, mimic the $\underline{\theta}$ type, and form an alliance. However, for both extensions we show that neither of these new incentives undermine our main insights. For both extensions we show: First, if the friendly condition holds then secret alliances never form in equilibrium. Second, if the antagonistic condition holds then a secret alliance equilibrium exists.[12] Third, that the probability of war in a secret alliance equilibrium is higher than in a public alliance equilibrium.

Next, we allow for bargaining between E and P. Bargaining significantly alters the structure of the model. Nevertheless, we show that secret alliances emerge in equilibrium when there is significant uncertainty about P's intentions and are more associated with conflict compared to public alliances.

Finally, we consider secret alliances that are explicitly offensive in nature. That is, E, rather than attacking, may choose to shore up its defenses or not, after which P can choose whether to attack E. We show that if E is uncertain about whether P is motivated to attack then secret alliances always emerge. However, unlike the non-offensive alliances considered in the main model, here they are most prevalent when E believes P is likely to be friendly.

Conclusion

In this paper, we have argued that alliances may act as a signal of members' alignment, revealing information about the constellation of interests in the international system, rather than just a signal of commitments and capabilities. Thus, a state may want to maintain secrecy if a public alliance would send an antagonistic signal about its interests to a potential enemy. By analyzing a model in which states may form an alliance and choose to publicly disclose its existence or not, we find that secret alliances may occur in equilibrium. We also derive results tying secret alliances to conflict and explain the decline in secret alliances over time. Overall, understanding the logic of secret alliances

[12] For the latter extension this result also imposes that war is sufficiently costly for A.

is crucial for understanding the functions and roles of military alliances in general. Furthermore, even beyond the phenomenon of secret alliances, conflict scholars should take the implications of signaling alignment seriously in the study of alliances.

References

Altfeld, Michael F and Bruce Bueno De Mesquita, "Choosing sides in wars," *International Studies Quarterly*, 1979, *23* (1), 87–112.

Andrew, Christopher, "German World policy and the Reshaping of the Dual Alliance," *Journal of Contemporary History*, 1966, *1* (3), 137–151.

Baliga, Sandeep and Tomas Sjöström, "Arms races and negotiations," *The Review of Economic Studies*, 2004, *71* (2), 351–369.

_ and _ , "Strategic ambiguity and arms proliferation," *Journal of political Economy*, 2008, *116* (6), 1023–1057.

Bas, Muhammet A and Andrew J Coe, "A dynamic theory of nuclear proliferation and preventive war," *International Organization*, 2016, *70* (4), 655–685.

Bas, Muhammet and Robert Schub, "Mutual optimism as a cause of conflict: Secret alliances and conflict onset," *International Studies Quarterly*, 2016, *60* (3), 552–564.

Benson, Brett V, Adam Meirowitz, and Kristopher W Ramsay, "Inducing deterrence through moral hazard in alliance contracts," *Journal of Conflict Resolution*, 2014, *58* (2), 307–335.

Beres, Louis René, "Bipolarity, multipolarity, and the reliability of alliance commitments," *Western Political Quarterly*, 1972, *25* (4), 702–710.

Berkemeier, Molly and Matthew Fuhrmann, "Reassessing the fulfillment of alliance commitments in war," *Research & Politics*, 2018, *5* (2), 2053168018779697.

DiGiuseppe, Matthew and Patrick E Shea, "Alliances, signals of support, and military effort," *European Journal of International Relations*, 2021, *27* (4), 1067–1089.

Fearon, James D, "Rationalist explanations for war," *International organization*, 1995, *49* (3), 379–414.

Fuhrmann, Matthew and Todd S Sechser, "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence," *American Journal of Political Science*, 2014, *58* (4), 919–935.

Gildea, Robert, *Barricades and borders: Europe 1800-1914*, OUP Oxford, 2003.

Henke, Marina E, "Buying allies: payment practices in multilateral military coalition-building," *International Security*, 2019, *43* (4), 128–162.

— , *Constructing allied cooperation: Diplomacy, payments, and power in multilateral military coalitions*, Cornell University Press, 2019.

Hinsley, Francis Harry, *Power and the Pursuit of Peace: Theory and Practice in the History of Relations between States*, Cambridge University Press, 1967.

Horowitz, Michael C, Paul Poast, and Allan C Stam, "Domestic signaling of commitment credibility: Military recruitment and alliance formation," *Journal of Conflict Resolution*, 2017, *61* (8), 1682–1710.

Huth, Paul K, "The extended deterrent value of nuclear weapons," *Journal of Conflict Resolution*, 1990, *34* (2), 270–290.

Kenwick, Michael R, John A Vasquez, and Matthew A Powers, "Do alliances really deter?," *The Journal of Politics*, 2015, *77* (4), 943–954.

Kuo, Raymond, "Secrecy among friends: covert military alliances and portfolio consistency," *Journal of Conflict Resolution*, 2020, *64* (1), 63–89.

Kurizaki, Shuhei, "Efficient secrecy: Public versus private threats in crisis diplomacy," *American Political Science Review*, 2007, *101* (3), 543–558.

Langer, William Leonard, *The diplomacy of imperialism, 1890-1902*, New York, Knopf, 1951.

Leeds, Brett Ashley, "Alliance reliability in times of war: Explaining state decisions to violate treaties," *International Organization*, 2003, *57* (4), 801–827.

— , "Do alliances deter aggression? The influence of military alliances on the initiation of militarized interstate disputes," *American Journal of Political Science*, 2003, *47* (3), 427–439.

— **and Andrew G Long**, "Reevaluating alliance reliability," *Journal of Conflict Resolution*, 2000, *44* (5), 686–699.

— **and Jesse C Johnson**, "Theory, data, and deterrence: A response to Kenwick, Vasquez, and Powers," *The Journal of Politics*, 2017, *79* (1), 335–340.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell, and Andrew Long, "Alliance treaty obligations and provisions, 1815-1944," *International Interactions*, 2002, *28* (3), 237–260.

McManus, Roseanne W, "Making it personal: The role of leader-specific signals in extended deterrence," *The Journal of Politics*, 2018, *80* (3), 982–995.

Meirowitz, Adam and Anne E Sartori, "Strategic uncertainty as a cause of war," *Quarterly Journal of Political Science*, 2008, *3* (4), 327–352.

_ , Massimo Morelli, Kristopher W Ramsay, and Francesco Squintani, "Dispute resolution institutions and strategic militarization," *Journal of Political Economy*, 2019, *127* (1), 378–418.

Morrow, James D, "Alliances and asymmetry: An alternative to the capability aggregation model of alliances," *American journal of political science*, 1991, pp. 904–933.

_ , "Alliances, credibility, and peacetime costs," *Journal of Conflict Resolution*, 1994, *38* (2), 270–297.

_ , "Alliances: Why write them down?," *Annual Review of Political Science*, 2000, *3* (1), 63–83.

_ , "When do defensive alliances provoke rather than deter?," *The Journal of Politics*, 2017, *79* (1), 341–345.

Oldson, William O, "Bismarck Looks East: The Austro-Romanian Treaty of 1883," *Il Politico*, 1977, pp. 290–300.

Paine, Jack, Scott A Tyson, Dirk Berg-Schlosser, Bertrand Badie, and Leonardo Morlino, "Uses and abuses of formal models in political science," 2020.

Palmer, Glenn, Roseanne W McManus, Vito D'Orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter, and Michael J Soules, "The MID5 Dataset, 2011–2014: Procedures, coding rules, and description," *Conflict Management and Peace Science*, 2022, *39* (4), 470–482.

Pribram, Alfred Francis, *The Secret Treaties of Austria-Hungary, 1879-1914: Negotiations leading to the treaties of the Triple alliance, with documentary appendices. Tr. by JC D'Arcy Paul and Denys P. Myers*, Vol. 2, Harvard University Press, 1921.

Ramm, Agatha, *European alliances and ententes 1879-85: A study of contemporary British information*, University of London, Bedford College (United Kingdom), 1937.

Ritter, Jeffrey Munro, *"Silent partners" and other essays on alliance politics*, Harvard University, 2004.

Shlaim, Avi, "The Protocol of Séevres, 1956: anatomy of a war plot," *International Affairs*, 1997, *73* (3), 509–530.

Signorino, Curtis S and Jeffrey M Ritter, “Tau-b or not tau-b: Measuring the similarity of foreign policy positions,” *International Studies Quarterly*, 1999, *43* (1), 115–144.

Slantchev, Branislav L, “Feigning weakness,” *International Organization*, 2010, *64* (3), 357–388.

Smith, Bradley C, “Military coalitions and the politics of information,” *The Journal of Politics*, 2021, *83* (4), 1369–1382.

Snyder, Glenn Herald, *Alliance politics*, Cornell University Press, 2007.

Trager, Robert F, “Diplomatic signaling among multiple states,” *The Journal of Politics*, 2015, *77* (3), 635–647.

Waltz, Kenneth N, “The stability of a bipolar world,” *Daedalus*, 1964, pp. 881–909.

_ , *Theory of international politics*, McGraw Hill, 1979.

Wolford, Scott, “Showing restraint, signaling resolve: Coalitions, cooperation, and crisis bargaining,” *American Journal of Political Science*, 2014, *58* (1), 144–156.

Zerner, Ruth, “Bismarck’s Views on the Austro-German Alliance and Future European Wars: a Dispatch of October 26, 1887,” *Austrian History Yearbook*, 1968, *4*, 227–244.

Online Appendix for "The Logic of Secret Alliances"

Table of contents:

A Proofs

Lemma 1. *In any equilibrium, if P proposes an alliance of type $x \in \{Pub, Sec\}$ and A accepts, then the transfer is given by:*

$$t^*(\theta) = \max\Big\{\rho(\emptyset)(\omega_A(\theta,0) - \pi_A(\theta)) - \rho(\varphi(x,1))(\omega_A(\theta,1) - \pi_A(\theta)), 0\Big\}$$

Proof. A's expected utility for accepting alliance (x,t) is

$$\rho(\varphi(x,1))\omega_A(\theta,1) + (1 - \rho(\varphi(x,1)))\pi_A(\theta) + t$$

and rejecting yields

$$\rho(\emptyset)\omega_A(\theta,0) + (1 - \rho(\emptyset))\pi_A(\theta).$$

Thus, A accepts any t such that

$$t \geq \rho(\emptyset)(\omega_A(\theta,0) - \pi_A(\theta)) - \rho(\varphi(x,1))(\omega_A(\theta,1) - \pi(\theta)),$$

and P will obviously not propose a larger value of t if it proposes an alliance. □

Lemma 2. *If there is complete information then a non-trivial secret alliance equilibrium does not exist.*

Proof. Assume P is known to be type θ. For a strategy profile to be a non-trivial secret alliance equilibrium there must be a positive probability that P chooses a secret alliance and chooses nothing. Assume $\rho(\emptyset) > 0$, Thus, P must be indifferent between the two action:

$$\begin{aligned} &\rho(\emptyset)\omega_P(\theta,1) + (1 - \rho(\emptyset))\pi_P(\theta) - t^* = \rho(\emptyset)\omega_P(\theta,0) + (1 - \rho(\emptyset))\pi_P(\theta) \\ &\Leftrightarrow \omega_P(\theta,1) - \omega_P(\theta,0) = \omega_A(\theta,0) - \omega_A(\theta,1) \end{aligned}$$

Which does not hold for almost all parameters. If $\rho(\emptyset) = 0$ then such an equilibrium is ruled out by assumption that P chooses no alliance when indifferent. □

Lemma 3. *If state P is friendly then a secret alliance equilibrium does not exist. There exists a PBE in which all types do not form an alliance and E never attacks on the path of play.*

Proof. To prove the first statement assume there is an equilibrium in which P only chooses no alliance or a secret alliance. Then E's payoff from attacking after $\varphi = \emptyset$ is bound above by

$\mathbb{E}_\theta[\omega_E(\theta, 0)] \leq \mathbb{E}_\theta[\pi_E(\theta)]$ and so $\rho(\emptyset) = 0$. But then any type forming a secret alliance is indifferent to deviating to no alliance, and so this is ruled out by our refinement.

Second, assume there is an equilibrium in which P also chooses a public alliance with positive probability. We first show that $\rho(\emptyset) > 0$ and $\rho(\text{Pub}) > 0$. First, for a contradiction suppose that $\rho(\emptyset) = \rho(\text{Pub}) = 0$. Such an equilibrium would not survive our refinement that alliances are subject to a small exogenous cost. Next for a contradiction suppose that $\rho(\emptyset) = 0$. Given this, at least one type θ' that is forming a public alliance would prefer to deviate to forming no alliance, a contradiction. An analogous argument shows that there cannot be such an equilibrium if $\rho(\text{Pub}) = 0$.

Because $\rho(\varphi) > 0$ for $\varphi \in \{\emptyset, Pub\}$ for this to be an equilibrium we must have that the following conditions hold:

$$\mathbb{E}_\theta[\pi_E(\theta)|Pub] \leq \mathbb{E}_\theta[\omega_E(\theta, 1)|Pub] \tag{6}$$

$$\mathbb{E}_\theta[\pi_E(\theta)|\emptyset] \leq \mathbb{E}_\theta\Big[E_a[\omega_E(\theta, a)|\theta]\Big|\emptyset\Big]. \tag{7}$$

If A and P form an alliance, by assumption this decreases E's war payoff. Thus, $\mathbb{E}_\theta[\omega_E(\theta, 1)|Pub] < \mathbb{E}_\theta[\omega_E(\theta, 0)|Pub]$ and $\mathbb{E}_\theta\Big[E_a[\omega_E(\theta, a)|\theta]\Big|\emptyset\Big] < \mathbb{E}_\theta[\omega_E(\theta, 0)|\emptyset]$. Together with the above inequalities, this implies that for this to be an equilibrium we must have

$$\mathbb{E}_\theta[\pi_E(\theta)|Pub] < \mathbb{E}_\theta[\omega_E(\theta, 0)|Pub] \tag{8}$$

$$\mathbb{E}_\theta[\pi_E(\theta)|\emptyset] < \mathbb{E}_\theta[E_a[\omega_E(\theta, 0)|\emptyset]. \tag{9}$$

If both inequalities hold then

$$\mathbb{E}_\theta[\pi_E(\theta)|Pub] + \mathbb{E}_\theta[\pi_E(\theta)|\emptyset] < \mathbb{E}_\theta[\omega_E(\theta, 0)|Pub] + \mathbb{E}_\theta[\omega_E(\theta, 0)|\emptyset]. \tag{10}$$

By the law of total expectation, (10) reduces to $\mathbb{E}_\theta[\pi_E(\theta)] < \mathbb{E}_\theta[\omega_E(\theta, 0)]$, a contradiction.

Finally, the existence of a no alliance peaceful equilibrium is immediate from the friendly condition. □

Lemma 4. *If P is not antagonistic, then there exists an equilibrium in which all types receive a public alliance and the probability of war is* 0.

Proof. Suppose that P is not antagonistic. We construct an equilibrium with the desired features. Suppose that all types of P choose a public alliance with $t = 0$. After observing a public alliance, let E's belief be equal to the prior, and after observing $\varphi = \emptyset$ E's belief places probability 1 on $\theta = \theta$. Suppose that $\rho(\emptyset) = 1$ and $\rho(\text{Pub}) = 0$. Now, we verify that no player has a profitable deviation. Clearly no type of P can profitably deviate, as a deviation results in a payoff of $\omega_E(\theta, 0) < \pi_E(\theta)$. As

$\rho(\text{Pub}) = 0$, A cannot profitably deviate from joining a public alliance after $t = 0$. Finally, given E's beliefs and the assumption that P is not antagonistic, it cannot profitably deviate from its strategy of attacking after observing nothing and choosing peace after observing a public alliance. □

Lemma 5. *If P is antagonistic then there exists a non-trivial secret alliance equilibrium. Moreover, there exists such an equilibrium in which all alliances formed on the path of play are secret.*

Proof. We prove the result by construction. Assume that if $\theta \in \Theta^*$ then P does not propose an alliance. For all types $\theta \notin \Theta^*$ P proposes a secret alliance with transfer $t^*(\theta)$, thus, A accepts and an alliance is formed by Lemma 1. By the antagonistic condition, E's expected utility for attacking after seeing $\varphi = \emptyset$ is greater than its expected utility for not attacking. By construction of Θ^* no type choosing to form a secret alliance has an incentive to deviate to no alliance and vice versa. Off the path of play if E observes $\varphi = Pub$ assume E believes the deviation came from a type θ such that $\omega_E(\theta, 1) > \pi_E(\theta)$, which always exists by assumption for type $\underline{\theta}$. Thus, $\rho(Pub) = 1$ and no type will deviate to a public alliance by assumption that $\pi_P(\theta) \geq \omega_P(\theta, 1)$ for all θ.

□

Proposition 1. *Every equilibrium is a secret alliance equilibrium if and only if P is antagonistic and P and E are conditionally mutually friendly.*

Proof. To prove the if direction, assume that both the antagonistic and conditionally mutually friendly conditions hold. We show there does not exist a public alliance equilibrium.

To start, assume there is an equilibrium where all types form a public equilibrium. By the antagonistic condition we must have $\rho(Pub) = 1$. Thus, any beliefs of E off the path of play must yield $\rho(\emptyset) \leq \rho(Pub)$. However, by construction, any type $\theta \in \Theta^*$ can profitably deviate to doing nothing, a contradiction.

Next, notice there cannot exist an equilibrium in which no type forms an alliance. If all types do nothing, then $\rho(\emptyset) = 1$ by the antagonistic condition. Consequently, all $\theta \notin \Theta^*$ have a profitable deviation to form a secret alliance. Additionally, if there does not exist a $\theta \notin \Theta^*$ then the conditionally mutually friendly condition implies that E prefers not to attack after $\varphi = \emptyset$, which would contradict the antagonistic condition.

Therefore, if a public alliance equilibrium exists under these conditions it must be that some types do nothing and some types form a public alliance. Clearly, there cannot be an equilibrium such that $\rho(\emptyset) < \rho(Pub)$, as any type forming a public alliance can profitably deviate to a secret alliance. Assume there is a public alliance equilibrium in which both actions are played with positive probability and $\rho(\emptyset) \geq \rho(Pub)$. A type θ prefers to do nothing if

$$\begin{aligned}\rho(\emptyset)\Big(\omega_P(\theta,0) - &\pi_P(\theta) + \omega_A(\theta,0) - \pi_A(\theta)\Big) \\ &> \rho(Pub)\Big(\omega_P(\theta,1) - \pi_P(\theta) + \omega_A(\theta,1) - \pi_A(\theta)\Big)\end{aligned} \tag{11}$$

and otherwise prefers to form an alliance. We show that if (11) holds for a type θ' then it holds for any type $\theta'' > \theta'$. A sufficient condition for this to be true is that:

$$\begin{aligned}&\rho(\emptyset)\Big(\omega_P(\theta'',0) - \pi_P(\theta'') + \omega_A(\theta'',0) - \pi_A(\theta'')\Big) - \rho(Pub)\Big(\omega_P(\theta'',1) - \pi_P(\theta'') + \omega_A(\theta'',1) - \pi_A(\theta'')\Big) \\ &\geq \rho(\emptyset)\Big(\omega_P(\theta',0) - \pi_P(\theta') + \omega_A(\theta',0) - \pi_A(\theta')\Big) - \rho(Pub)\Big(\omega_P(\theta',1) - \pi_P(\theta') + \omega_A(\theta',1) - \pi_A(\theta')\Big),\end{aligned}$$

which rearranges to

$$\begin{aligned}&\rho(\emptyset)\Big([\omega_P(\theta'',0) - \pi_P(\theta'')] - [\omega_P(\theta',0) - \pi_P(\theta')] + [\omega_A(\theta'',0) - \pi_A(\theta'')] - [\omega_A(\theta',0) - \pi_A(\theta')]\Big) \\ &\geq \rho(Pub)\Big([\omega_P(\theta'',1) - \pi_P(\theta'')] - [\omega_P(\theta',1) - \pi_P(\theta')] + [\omega_A(\theta'',1) - \pi_A(\theta'')] - [\omega_A(\theta',1) - \pi_A(\theta')]\Big)\end{aligned} \tag{12}$$

By assumption that $\omega_i(\theta, a) - \pi(\theta)$ is increasing in θ for $i \in \{P, A\}$ and $a \in \{0, 1\}$ we have that each side of (12) is positive. Because $\rho(\emptyset) \geq \rho(Pub)$, this implies that a sufficient condition for (12) to hold is that the following two inequalities hold:

$$[\omega_P(\theta'',0) - \pi_P(\theta'')] - [\omega_P(\theta',0) - \pi_P(\theta')] \geq [\omega_P(\theta'',1) - \pi_P(\theta'')] - [\omega_P(\theta',1) - \pi_P(\theta')] \tag{13}$$

$$[\omega_A(\theta'',0) - \pi_A(\theta'')] - [\omega_A(\theta',0) - \pi_A(\theta')] \geq [\omega_A(\theta'',1) - \pi_A(\theta'')] - [\omega_A(\theta',1) - \pi_A(\theta')]. \tag{14}$$

Simplifying, the above inequalities reduce to:

$$\omega_P(\theta',1) - \omega_P(\theta',0) \geq \omega_P(\theta'',1) - \omega_P(\theta'',0) \tag{15}$$

$$\omega_A(\theta',1) - \omega_A(\theta',0) \geq \omega_A(\theta'',1) - \omega_A(\theta'',0). \tag{16}$$

Note that both inequalities hold by assumption that $\omega_i(\theta, 1) - \omega_i(\theta, 0)$ is decreasing in θ for $i \in \{P, A\}$. Therefore, the set $\widehat{\Theta}^*$ is characterized by a $\hat{\theta}^*$ such that $\theta \in \widehat{\Theta}^*$ if and only if $\theta \geq \hat{\theta}^*$. For $\rho(\emptyset) = \rho(Pub)$ we have $\widehat{\Theta}^* = \Theta^*$ and let θ^* be the corresponding cutoff. Finally, the RHS of (12) is increasing in $\rho(Pub)$. Therefore, if (12) holds at θ for $\rho(Pub) < \rho(\emptyset)$ then it also holds at θ for $\rho(Pub) = \rho(\emptyset)$. Thus, $\rho(Pub) \leq \rho(\emptyset)$ implies $\hat{\theta}^* \leq \theta^*$. Consequently, the assumption that

$\omega_E(\theta, a) - \pi_E(\theta)$ is decreasing in θ together with the conditionally mutually friendly condition yields

$$0 \leq \mathbb{E}_\theta[\pi_E(\theta)|\theta \in \Theta^*] - \mathbb{E}_\theta[\omega_E(\theta, 0)|\theta \in \Theta^*] \leq \mathbb{E}_\theta[\pi_E(\theta)|\theta \in \widehat{\Theta}^*] - \mathbb{E}_\theta[\omega_E(\theta, 0)|\theta \in \widehat{\Theta}^*].$$

In the conjectured public alliance equilibrium all types $\theta \in \widehat{\Theta}^*$ must choose no alliance and all types $\theta \notin \widehat{\Theta}^*$ must be choosing an alliance. Therefore, after seeing $\varphi = \emptyset$ E must prefer not to attack and by the antagonistic condition after seeing $\varphi = Pub$ E must prefer to attack. However, then any type $\theta \notin \widehat{\Theta}^*$ has a profitable deviation to doing nothing, which contradicts that there is a PBE in which $\rho(\emptyset) \geq \rho(Pub)$ under the antagonistic and conditional mutually friendly conditions.

Next, to prove the only if direction, assume that every equilibrium is a secret alliance equilibrium. We show this implies that the antagonistic and conditionally mutually friendly conditions both must hold. We proceed in two steps. First, to derive a contradiction, suppose that the antagonistic condition does not hold. If this is true, then by Lemma 4 there exists a public alliance equilibrium. Therefore, if every alliance equilibrium is a secret alliance equilibrium the antagonistic condition must hold. To complete the proof, for a contradiction assume that every equilibrium is a secret alliance equilibrium but that conditional mutual friendliness does not hold. To derive a contradiction, we construct a public alliance equilibrium. Suppose that all $\theta \in \Theta^*$ choose no alliance and that all $\theta \notin \Theta^*$ choose a public alliance. Further, suppose that $\rho(\emptyset) = \rho(\text{Pub}) = 1$. To see that no type of any player has a profitable deviation, note that by the definition of Θ^* no type of P can profitably deviate. Further, as conditional mutual friendliness does not hold and the antagonistic condition does hold, E cannot profitably deviate from attacking after observing either a public alliance or nothing. Of course, as P is offering t^*, A cannot profitably deviate. Therefore, if every equilibrium is a secret alliance then the antagonistic condition and conditional mutual friendliness must both hold. □

Proposition 2. *There exists $\mu^* \in (\underline{\mu}, \overline{\mu})$ such that:*

- *There is an equilibrium in which the $\underline{\theta}$ type forms a secret alliance and the $\overline{\theta}$ does not form an alliance if and only if $\mu_0 \leq \mu^*$.*
- *There is an equilibrium in which both types form a public alliance if and only if $\mu_0 \geq \underline{\mu}_0$.*
- *There is an equilibrium in which neither type forms an alliance if and only if $\mu_0 \geq \overline{\mu}$.*

No other equilibria exist.

Proof. Because the transfer t^* depends on the equilibrium probability of war, it is straightforward that with two types a mixed strategy equilibrium cannot exist for almost all values of the parameters. The result the follows from considering pure strategy equilibria and applying the previous results for the general Θ case. □

Proposition 3. *Assume the P antagonistic condition holds. In every equilibrium $\rho(\emptyset) \geq \rho(Pub)$. In every secret alliance equilibrium $\rho(\emptyset) = \rho(Pub) > 0$.*

Proof. We show that if in equilibrium $\rho(\emptyset) < \rho(Pub)$ then at least one type is choosing a public alliance. For a contradiction suppose not. By assumption that $\omega_E(\theta, 1) < \omega_E(\theta, 0)$ we have that E's expected utility for attacking is bounded below by $\mathbb{E}_\theta[\omega_E(\theta, 1)]$, which is strictly greater than $\mathbb{E}_\theta[\pi_E(\theta)]$ by the antagonistic condition. This implies $\rho(\emptyset) = 1 \geq \rho(Pub)$, a contradiction.

To finish proving the first part of the proposition, towards a contradiction suppose that there exists an equilibrium in which $\rho(\emptyset) < \rho(Pub)$. By the above, we know that at least one type, θ', chooses a public alliance. Type θ' has an equilibrium payoff of

$$\rho(Pub)\omega_P(\theta', 1) + (1 - \rho(Pub))\pi_P(\theta) - t^*.$$

Now consider a deviation to (Sec, t^*). By the above, A accepts the offer of t^* and a secret alliance is formed. Thus, if P deviates in this way its payoff is

$$\rho(\emptyset)\omega_P(\theta, 1) + (1 - \rho(\emptyset))\pi_P(\theta) - t^* > \rho(Pub)\omega_P(\theta', 1) + (1 - \rho(Pub))\pi_P(\theta) - t^*,$$

a contradiction.

For the second part, assume there is a secret alliance in which $\rho(\emptyset) > \rho(Pub)$. This cannot be an equilibrium, as any type forming a secret alliance can profitably deviate to a public alliance. □

Proposition 4.

1. *If a secret alliance equilibrium does not exist, then there exists a public alliance equilibrium where the probability of war is* 0.
2. *If a public alliance equilibrium does not exist, then there exists a secret alliance equilibrium where the probability of war is* 1.

Proof. If a secret alliance equilibrium does not exist then by Lemma 5 the antagonistic condition must fail. Conjecture that all types choose to form a public alliance. Then $\mathbb{E}_\theta[\pi_E(\theta)|Pub] = \mathbb{E}_\theta[\pi_E(\theta)] > \mathbb{E}_\theta[\omega_E(\theta, 1)] = \mathbb{E}_\theta[\omega_E(\theta, 1)|Pub]$ because the antagonistic condition holds. Thus, $\rho(Pub) = 0$ and $t^*(\theta) = 0$ for all types θ and no type has an incentive to deviate.

Next, for the second part, assume a public alliance equilibrium does not exist. By Proposition 1 the antagonistic and conditionally mutually friendly conditions must hold. Because the antagonistic condition holds such an equilibrium exists by the proof of Lemma 5. □

B Robustness

Here we consider a number of extensions to the baseline model. For simplicity, we focus on the two-type case throughout, $\Theta = \{\underline{\theta}, \overline{\theta}\}$.

Public transfers

We now alter the model so if a public alliance is formed then E observes the value of the transfer t. Now E makes inferences about P's type from both the existence of a public alliance and the size of the transfer. Although this creates additional signaling incentives for P when forming public alliances, it does not have a significant impact on the formation of secret alliances as similar incentive constraints as before must still hold.

Lemma B1. *If state P is friendly then a secret alliance equilibrium does not exist.*

Proof. The proof there is not a secret alliance equilibrium in which no type forms a public alliance is the same as for the proof for Lemma 3. When some types form public alliances the proof is similar to the proof for Lemma 3, but with expectations conditioned on both *Pub* and t. □

Lemma B2. *If P is antagonistic then there exists a non-trivial secret alliance equilibrium. Moreover, there exists such an equilibrium in which all alliances formed on the path of play are secret.*

Proof. The proof is the same as for the proof of Lemma 5. □

Lemma B3. *The probability of war in any public alliance equilibrium is* 0. *If P is antagonistic, then the probability of war in any secret alliance equilibrium is* 1.

Proof. We start by proving the result for public alliances.

First, note that there cannot be a public alliance equilibrium in which the types separate. If such an equilibrium existed then E strictly prefers not to attack following one signal $\varphi \in \{\emptyset, 1\}$, and strictly prefers to attack after the other signal. Thus, one type has a profitable deviation.

Second, assume both types form a public alliance. For almost all parameters E has a strictly optimal decision after observing $\varphi = 1$. If E prefers not to attack, then the probability of war is 0, as required. If E prefers to attack, then the $\overline{\theta}$ type can deviate to not forming an alliance, which yields at most the same probability of war, and which is profitable by assumption that $\overline{\theta} \in \Theta^*$.

Now we prove the second part of the result. Suppose that P is antagonistic. We show that every secret alliance equilibrium has probability of war 1. First, suppose that both types are forming a secret alliance or no alliance. Then it follows immediately from the assumption that P is antagonistic that E will attack with probability 1. Next, suppose that a public alliance occurs on the path of play. E must attack following a public alliance with probability 1. Applying arguments from the baseline model, the probability of war following a public alliance and nothing must be equal. □

A does not observe θ

Now alter the baseline model so only P observes θ, while A and E have a commonly known prior belief that θ is drawn according to distribution F.

Lemma B4. *If state P is friendly then a secret alliance equilibrium does not exist.*

Proof. The proof is analogous to the proof of Lemma 3. □

Lemma B5. *In equilibrium, if types θ' and θ'' both propose an alliance of type $x \in \{Pub, Sec\}$ with transfers t' and t'', respectively, then $t' = t''$.*

Proof. Assume not. Without loss of generality assume $t' < t''$. Then θ'' strictly benefits by deviating to t'. □

Lemma B6. *If P is antagonistic and $\omega_A(\underline{\theta}, 0) - \omega_A(\underline{\theta}, 1) > \omega_P(\overline{\theta}, 1) - \omega_P(\overline{\theta}, 0)$ then a secret alliance equilibrium exists.*

Proof. Assume the $\underline{\theta}$ type proposes a secret alliance with transfer $t^*(\underline{\theta})$ with $\rho(\emptyset) = 1$, as defined in the baseline game, and A accepts. The $\overline{\theta}$ type does not propose an alliance. And E attacks after seeing $\varphi = \emptyset$ or $\varphi = 1$. By the antagonistic condition it is optimal for E to attack after $\varphi = \emptyset$ and we can choose beliefs off the path following $\varphi = 1$ such that E attacks.

Next, note that A's belief after observing alliance offer $t^*(\underline{\theta})$ puts probability 1 on P being the $\underline{\theta}$ type. Thus, by construction of $t^*(\underline{\theta})$ it is optimal for A to accept. By assumption in the two type model the $\underline{\theta}$ type prefers to form an alliance with A.

Finally, it must be that the $\overline{\theta}$ type does not want to deviate and form an alliance by mimicking the $\underline{\theta}$ type (we can eliminate offers $t \neq t^*(\underline{\theta})$ using off-path beliefs). This holds if

$$\begin{aligned} \omega_P(\overline{\theta}, 0) &\geq \omega_P(\overline{\theta}, 1) - t^*(\underline{\theta}) \\ \Leftrightarrow \omega_A(\underline{\theta}, 0) - \omega_A(\underline{\theta}, 1) &> \omega_P(\overline{\theta}, 1) - \omega_P(\overline{\theta}, 0), \end{aligned}$$

which is true by assumption. □

Lemma B7. *The probability of war in any public alliance equilibrium is 0. If P is antagonistic and $\omega_A(\underline{\theta}, 0) - \omega_A(\underline{\theta}, 1) > \omega_P(\overline{\theta}, 1) - \omega_P(\overline{\theta}, 0)$ then the probability of war in any secret alliance equilibrium is 1.*

Proof. The proof is analogous to that of Lemma B3. □

Offensive Secret Alliances

Our first extension considers secret alliances that are offensive in nature. In particular, we show that our logic of signalling intentions also applies to these types of alliances. Although the baseline model is fairly general in terms of payoffs following war and peace, it does not fit well for offensive alliances because $\pi(\theta)$ is not allowed to depend on whether P and A have formed an alliance. Thus, if E does not attack then forming an alliance is not beneficial. In particular, the logic for the emergence of secret alliances in the baseline model is not equivalent to an offensive sneak attack logic.

The incentive for secrecy that we aim to model here is illustrated by Italy's role in the Seven Weeks' war.[13] In April 1866, a dispute between Prussia and Austria arose over Schleswig-Holstein, a region which had been jointly administered by Prussia and Austria since 1864. Anticipating that this dispute may escalate, Otto von Bismarck initiated a secret treaty with Italy on April 8th, which bound Italy to attack Austria if war broke out. Bismarck's aim with the treaty was to divert Austrian forces to the south, forcing it to fight a two-front war. This drove the incentive for secrecy. Austria, ignorant of Italy's sympathy towards Bismarck's aims, did not prepare for Italian entry into the conflict, and was caught off guard. This surprise contributed to a relatively swift Prussian victory, with the conflict lasting just over a month. This episode illustrates the logic of secret alliances for the purpose of surprise attack; by concealing a potential partner's interest in joining, a state planning aggression can catch its opponent off guard, preventing them from taking action to shore up their defenses against expansion of a war. We now present an extension to allow for such alliances.

The alliance formation stage proceeds as before. After observing a public alliance or no alliance E decides whether to shore up its defenses, $s = 1$, or not, $s = 0$. Next, state P observes whether E shores up its defenses and decides to attack or not.

If P attacks then it wins the conflict with probability $p(a, s)$ and E wins with probability $1-p(a, s)$. Thus, the probability P wins depends on whether E shores up and whether P and A form an alliance. For simplicity, we assume that if E shores up it is certainly victorious, $p(a, 0) = p(a, 1) = 0$. Otherwise, if E does not shore up then $0 < p(0, 0) = \underline{p} < p(1, 0) = \overline{p} < 1$. P's payoff from war is $p(a, s) + (1 - p(a, s))\theta - c_P$ and its payoff from peace is θ. A's payoff from war is $p(a, s) - ac_A$ and its payoff from peace is 0. Finally, E's payoff from war is $p(a, s)\theta + (1 - p(a, s)) - c_E - sk$ and its payoff from peace is $1 - sk$. Let $\Theta = \{0, 1\}$. Thus, when $\theta = 1$ state P and state E have perfectly aligned interests and when $\theta = 0$ state P and state A are aligned preferences opposed to state E. Recall that $\mu_0 = Pr(\theta = 1)$.

Note that for the aligned type attacking is strictly dominated by not attacking. Consequently,

[13] For more on the Seven Weeks war see Pflanze, Otto (1963) "Bismarck and the Development of Germany, Volume I: 1815-1871" Princeton University Press.

forming an alliance can never improve its payoff. Thus, we look for perfect Bayesian equilibrium in which the $\theta = 1$ type never forms an alliance.[14] Additionally, we focus on pure strategies, although allowing for mixed strategies does not change our conclusions. Also assume $\underline{p} > c_P$, so the $\theta = 0$ type always prefers to attack. And assume $c_A \in (\overline{p} - \underline{p}, 2(\overline{p} - \underline{p}))$, thus, the ally will not intervene absent a positive transfer but if P attacks it is willing to form the alliance. Proposition 5 characterizes behavior in such equilibria.

Proposition 5. *There does not exist an equilibrium in which public alliances occur with positive probability. There exist $\underline{\mu}^O$ such that secret alliances occur in every equilibrium if and only if $\mu > \underline{\mu}^O$. Further, there exists $\overline{\mu}^O > \underline{\mu}^0$ such that, in equilibrium, war occurs with positive probability if $\mu > \overline{\mu}^O$ and never occurs if $\mu \in (\underline{\mu}^O, \overline{\mu}^O]$.*

Proof. The type $\theta = 1$ never forms an alliance. Off-the-path assume μ is sufficiently low that E shores up.

First, there cannot be an equilibrium in which the $\theta = 0$ type always forms a public alliance. If there was such an equilibrium, then state E strictly prefers to shore up after seeing the public alliance and not shore up after seeing nothing. As such the $\theta = 0$ type could profitably deviate to forming a secret alliance and attacking.

Second, assume the $\theta = 0$ type does not form an alliance. It attacks if E does not shore up. Thus, E's utility for shoring up is $1-k$ and its expected utility for not shoring up is $\mu_0 + (1-\mu_0)(1-\underline{p}-c_E)$. Thus, E shores up if $\mu_0 \leq \frac{\underline{p}+c_E-k}{\underline{p}+c_E} \equiv \underline{\mu}^O$. If E shores up there is no incentive for the $\theta = 0$ type to deviate and form an alliance. Thus, this is an equilibrium when $\mu \leq \overline{\mu}^O$.

Third, assume the $\theta = 0$ type forms a secret alliance. If $\mu_0 \leq \frac{\overline{p}+c_E-k}{\overline{p}+c_E} \equiv \overline{\mu}^O$ then it is optimal for E to shore up. The probability of war is 0 so the transfer is free and $\theta = 0$ type has no incentive to deviate.

Fourth, assume the $\theta = 0$ type forms a secret alliance and $\mu_0 \geq \overline{\mu}^O$. Thus, after seeing nothing E does not shore up. Thus, the $\theta = 0$ type will attack. As such, the transfer it makes to A must solve $\overline{p} + t - c_A = \underline{p}$ so $t^* = c_A - (\overline{p} - \underline{p})$. Thus, $\theta = 0$ does not want to deviate to no alliance if $\overline{p} - t^* - c_P \geq \overline{p} - c_P$ which holds iff $2(\overline{p} - \underline{p}) \geq c_A$. □

As in the baseline model, State P worries that forming a public alliance will signal misaligned interests to E. In this case, sending a negative signal mitigates P's ability to launch an effective attack because E can prepare in anticipation. However, despite the similar logic for secret offensive alliances and secret defensive alliances, they emerge under different conditions. In particular, secret offensive alliances are most prevalent when E believes P is likely to be friendly, whereas secret alliances are most prevalent in the baseline model when E believes P is likely to be antagonistic.

[14] If we relax this restriction then public alliances can occur in equilibrium. However, they are trivial in the sense that there would also exist a payoff equivalent secret alliance equilibrium.

Bargaining

We now allow for bargaining between P and E. Specifically, we modify the previously described alignment special case of our model as described in the text. Assume that after the alliance formation stage E makes an offer $y \in [0,1]$. Next, P accepts or rejects. If P accepts then peace prevails with outcome y in implemented. If P rejects then war occurs. Payoffs from peace and conflict are the same as for the alignment model.

We first establish conditions under which secret alliances do and do not occur in equilibrium.

Proposition 6.

1. *If $\underline{\theta}$ sufficiently close to* 1*, then a secret alliance equilibrium does not exist.*
2. *If $\underline{\theta}$ is sufficiently close* 0*, $\overline{\theta}$ is sufficiently close to* 1*, and the prior belief $Pr(\theta = \overline{\theta})$ is sufficiently high, then a secret alliance equilibrium exists.*

Proof. To prove the first part, assume $\underline{\theta}$ is sufficiently close to 1 such that the $\underline{\theta}$ type accepts $x = 1$ even if it forms an alliance. This implies that the $\overline{\theta}$ type also accepts $x = 1$.

To prove the second part, assume the $\underline{\theta}$ type forms a secret alliance and the $\overline{\theta}$ type does not. Further, let x^* be equal to the offer that would be made to the $\overline{\theta}$ type under complete information. Consider the following assessment. The $\overline{\theta}$ type does not form an alliance and the $\underline{\theta}$ type forms a secret alliance and offers t^*. Following observing nothing, E's updated belief is equal to the prior belief. Following observing a public alliance, E updates to believe that $Pr(\theta = \overline{\theta}) = 1$. E offers $x = 1$ after observing nothing or a public alliance. Each type accepts x if and only if the payoff of accepting is weakly greater than the payoff of rejecting.

Assume $\underline{\theta}$ is sufficiently close to 0 such that the $\underline{\theta}$ type rejects the offer that would be made to. If $Pr(\theta = \overline{\theta})$ is sufficiently close to 1 then it is optimal for E to make the offer to the high type. Thus, for this to be an equilibrium neither type of P must want to deviate. Because E makes the offer $y^* = 1$ following $\varphi = 1$ or $\varphi = \emptyset$ neither type can profitably deviate by changing whether it has an alliance or the type of the alliance. Finally, if $\underline{\theta}$ is sufficiently close to 0 then it prefers to go to war rather than accept this offer, and by assumption is willing to pay the cost t^* to obtain the ally. □

Similar to the baseline model, if P is friendly towards E, in this case all types have an ideal settlement not too far from E, then secret alliances do not form. It is expensive to form an alliance because all types of P are relatively far from A. Thus, even if E proposes its ideal point all types prefer to accept this offer rather than form a secret alliance and fight.

On the other hand, our conditions for secret alliances to arise in equilibrium look less similar to the baseline. This is due to the most significant change between the deterrence and bargaining models which is that there now exist settlements such that E prefers to not fight low types of P, e.g., the $\underline{\theta}$ type. Nevertheless, we see that dispersion in the intentions of P towards E still results in the formation of secret alliances. Consequently, our broader theoretical argument that secret alliances are due to significant uncertainty over intentions continues to hold.

To conclude we establish that secret alliances should be associated with conflict, even when we allow for bargaining.

Proposition 7.

1. *In any secret alliance equilibrium there is a strictly positive probability of war.*
2. *If there exists a public alliance equilibrium with a strictly positive probability of war, then there exists a secret alliance equilibrium with the same probability of war. Moreover, under some conditions, there exists a public alliance equilibrium where the probability of war is* 0*.*

Proof. To prove the first part assume there is an equilibrium in which some type forms a secret alliance with positive probability, but $\rho(\emptyset) = 0$. The payoff from forming the alliance is $\pi_P(\theta) - t^*(\theta) \leq \pi_P(\theta)$. Thus, although $t^*(\theta) =$ when $\rho(\emptyset) = 0$, such a strategy profile being an equilibrium is ruled out by our refinement.

To prove the second part consider a public alliance equilibrium. Let $x^*(Pub)$ be E's offer after seeing $\varphi = Pub$ and let $x^*(\emptyset)$ be E's offer after seeing $\varphi = \emptyset$. For there to be a positive probability of war, P must reject at least sometimes following one or both offers.

To start, note the types cannot be a fully separating in such an equilibrium. If so, E could tailor the offers $x^*(Pub)$ and $x^*(\emptyset)$ to each type and each type must accept with probability 1, otherwise E could profitably deviate to $x^* - \epsilon$ which type θ strictly prefers to accept, for ϵ sufficiently small.

Next, we argue there cannot be such an equilibrium in which a type mixes over forming a public alliance and forming no alliance. To see this, we consider three cases.

First, assume that if this type never rejects then to be indifferent we must have $x^*(Pub) = x^*(\emptyset) = x^*$, but then forming the alliance and accepting x^* rather than forming no alliance and accepting x^* is ruled out by our refinement.

Second, consider the case where such a type sometimes rejects after the offer $x^*(\emptyset)$ and always accepts after $x^*(Pub)$. We must have $\omega_P(\theta, 0) = u_\theta(x^*(Pub))$. However if $\theta = \underline{\theta}$ then $\omega_P(\underline{\theta}, 1) - t^*(\underline{\theta}) > \omega_P(\underline{\theta}, 0) = u_{\underline{\theta}}(x^*(Pub))$, which contradicts that $\underline{\theta}$ always accepts $x^*(Pub)$. Thus, if there is such a type we have $\theta = \overline{\theta}$ and the $\underline{\theta}$ type must choose no alliance or a public alliance with probability 1. In particular, $\underline{\theta}$ must be choosing no alliance, otherwise, E will tailor its offer so that $\overline{\theta}$ always accepts after $\emptyset$. Thus, after seeing Pub E knows $\theta = \overline{\theta}$ and offers $x^*(Pub)$ such that $u_{\overline{\theta}}(x) = \omega_P(\overline{\theta}, 1)$. Under the conjectured strategy the $\overline{\theta}$ type sometimes rejects after $x^*(\emptyset)$ which yields a payoff of $\omega_P(\overline{\theta}, 0)$. However, $\omega_P(\overline{\theta}, 1) - t^*(\overline{\theta}) < \omega_P(\overline{\theta}, 0)$, which contradicts that the $\overline{\theta}$ type is indifferent between accepting $x^*(Pub)$ and rejecting after $x^*(\emptyset)$.

The third case where there is a type that always accepts the offer $x^*(\emptyset)$ and sometimes rejects after $x^*(Pub)$. However, an argument similar to the previous case rules out that this can be an equilibrium.

Finally, consider the case where this mixing type rejects after both offers. Thus, the indifference condition requires. $\omega_P(\theta, 1) - t^*(\theta) = \omega_P(\theta, 0)$. However, $\omega_P(\underline{\theta}, 1) - t^*(\underline{\theta}) > \omega_P(\underline{\theta}, 0)$ and $\omega_P(\overline{\theta}, 1) - t^*(\overline{\theta}) < \omega_P(\overline{\theta}, 0)$.

The above arguments imply that the only possible equilibrium involves both types pooling on forming a public alliance. By the standard argument, E either offers $x^*(Pub)$ to make the $\overline{\theta}$ indifferent, which only the $\overline{\theta}$ type accepts. Or E offers $x^*(Pub)$ which, depending on the parameters either both accept or only the $\underline{\theta}$ type accepts.

First, assume there is a public alliance equilibrium with a positive probability of war where E offers $x^*(Pub)$ to make the $\overline{\theta}$ indifferent between accepting or rejecting. Thus, $u_E(\overline{x}(\underline{\theta})) \leq \mu_0 u_E(\overline{x}(\overline{\theta}, 1)) + (1 - \mu_0)\omega_E(\underline{\theta})$, when the $\overline{\theta}$ type is willing to accept the offer $\overline{x}(\underline{\theta}, 1)$. In this case, the $\overline{\theta}$ type accepts, thus, $\omega_P(\overline{\theta}, 1) - t^*(\theta) \geq$. On the other hand, the $\underline{\theta}$ type rejects, thus, $\omega_P(\underline{\theta}, 1) > u(x^*)$. Now consider the strategy profile where the $\underline{\theta}$ type forms a secret alliance and the $\overline{\theta}$ type does not. In this case, the payoff of the $\overline{\theta}$ type is $\omega_P(\overline{\theta}, 0) > \omega_P(\overline{\theta}, 1) - t^*(\overline{\theta})$, which holds. The $\underline{\theta}$ type must prefer to form an alliance and reject over forming no alliance and taking $\overline{x}(\overline{\theta}, 0)$. This holds by $\omega_P(\underline{\theta}, 1) - t^*(\underline{\theta}) > \omega_P(\underline{\theta}, 0) = u(\overline{\underline{\theta}, 0}) > u(\overline{\theta}, 1)$. Additionally, the first inequality also implies $\underline{\theta}$ prefers to form the alliance over not and rejecting. Finally, for this to be an equilibrium E must prefer to make the $\overline{x}(\overline{\theta}, 0)$ offer over $\overline{x}(\underline{\theta}, 1)$. This requires $u_E(\overline{x}(\underline{\theta})) \leq \mu_0 u_E(\overline{x}(\overline{\theta}, 0)) + (1 - \mu_0)\omega_E(\underline{\theta})$. Note that $u_E(\overline{x}(\overline{\theta}, 0)) > u_E(\overline{x}(\overline{\theta}, 1))$. From the assumption of the public alliance equilibrium we have $u_E(\overline{x}(\underline{\theta})) \leq \mu_0 u_E(\overline{x}(\overline{\theta}, 1)) + (1 - \mu_0)\omega_E(\underline{\theta})$. Thus, $u_E(\overline{x}(\underline{\theta})) \leq \mu_0 u_E(\overline{x}(\overline{\theta}, 1)) + (1 - \mu_0)\omega_E(\underline{\theta}) < \mu_0 u_E(\overline{x}(\overline{\theta}, 0)) + (1 - \mu_0)\omega_E(\underline{\theta})$, as required. A similar argument shows the case where the $\overline{\theta}$ type rejects the offer $\overline{x}(\underline{\theta}, 0)$ offer.

Second, consider an equilibrium where E offers $x^*(Pub) = \overline{x}(\underline{\theta}, 1)$ to make the $\underline{\theta}$ type indifferent, which the $\underline{\theta}$ type accepts. We must have that $u_{\overline{\theta}}(\overline{x}(\underline{\theta}, 1) < \omega_P(\overline{\theta}, 1)$, otherwise the $\overline{\theta}$ type accepts and the probability of war is 0. Thus, the $\overline{\theta}$ types equilibrium payoff is $\omega_P(\overline{\theta}, 1) - t^*(\overline{\theta})$. However, consider a deviation by the $\overline{\theta}$ type to forming no alliance and rejecting the off-path offer. This yields a payoff $\omega_P(\overline{\theta}, 0) > \omega_P(\overline{\theta}, 1) - t^*(\overline{\theta})$, thus, there cannot be such an equilibrium. □

As before, the occurrence of secret alliances is associated with conflict relative to public alliances. If a type of P anticipates always accepting E's offer then it has no incentive to ever form a secret alliance. This is because P could deviate to no alliance without E observing, and then still accept the offer. Thus, if a type of P ever forms a secret alliance in equilibrium then it must reject E's offer (at least sometimes).

In contrast, there can be public alliance equilibrium in which peace always prevails. As described earlier, bargaining allows E to strike a deal even with the least friendly type of P. Therefore, if the $\underline{\theta}$ type forms a public alliance and the $\overline{\theta}$ type does not, the public alliance allows E to better tailor its offer, the $\underline{\theta}$ type to obtain a better offer, and for the states avoid conflict. Additionally, even when war can occur with public alliances, under the same parameters we can sustain a secret alliance equilibrium with the same probability of war. Thus, even with bargaining we expect a relationship between conflict and secret alliances.

State A Proposes Alliance

In our baseline model state P holds the bargaining power with A in deciding the type of alliance and the transfer. Here, we show that our conclusions about the emergence of secret alliances in equilibrium are not sensitive to this assumption. Assume now that A proposes the transfer $t \geq 0$ and type of alliance $x \in \{Pub, Sec\}$, after which P can accept or reject. Everything else remains the same as in the baseline model. P's expected utility for accepting alliance (x, t) is

$$\rho(\varphi(x,1))\omega_P(\theta,1) + (1 - \rho(\varphi(x,1)))\pi_P(\theta) - t$$

and rejecting yields

$$\rho(\emptyset)\omega_P(\theta,0) + (1 - \rho(\emptyset))\pi_P(\theta).$$

Thus, in any equilibrium, if A proposes an alliance of type $x \in \{Pub, Sec\}$ and P accepts then the transfer is given by

$$t^*(\theta) = \rho(\varphi(x,1))\Big(\omega_P(\theta,1) - \pi_P(\theta)\Big) - \rho(\emptyset)\Big(\omega_P(\theta,0) - \pi_P(\theta)\Big)$$

For A to be willing to propose an alliance that is accepted by P thus requires

$$\begin{aligned}&\rho(\varphi(x,1))\omega_A(\theta,1) + (1 - \rho(\varphi(x,1))\pi_A(\theta) + t^*(\theta) \geq \rho(\varphi(x,1))\omega_A(0,\theta) + (1 - \rho(\varphi(x,1)))\pi_A(\theta)\\ &\Leftrightarrow \omega_P(\theta,1) - \omega_P(\theta,0) \geq \omega_A(\theta,0) - \omega_A(\theta,1). \qquad (17)\end{aligned}$$

Flipping the inequality in (17), this is exactly the condition that defines Θ^*. Thus, for any $\theta \in \Theta^*$ A and P never form an alliance in equilibrium, which is the same as in the baseline model.

We now argue that the results of the baseline model hold in this extension. First, notice that the difference in A's expected utility from no alliance and proposing an accepted alliance (net of the transfer) is the same as the difference in the baseline model for P from no alliance and proposing an accepted alliance (net of the transfer). Thus, the inequalities that must hold for A and P to form an alliance of type x in equilibrium are exactly the same as those in the baseline model. Second, if we fix a set of θ choosing no alliance, secret alliance, and public alliance, then E's expected utility from attacking following $\varphi = Pub$ or $\varphi = \emptyset$ is the same regardless of which player proposed the alliance. Consequently, the proofs used in the baseline model to sustain when certain types of equilibria exist or do not exist can be applied to the extension unchanged, outside of relabeling the incentive constraints for P as the incentive constraint for A.

C Data on Secret Alliances

In the discussion of the paper, we assess empirical patterns of secret alliances using data from ATOP (Leeds et al., 2002). Here we discuss why we believe the data is informative, despite the secret nature of the alliances we study.

First, while we would not expect to observe more recent secret alliances in this data, the most dramatic shift in the prevalence of secret alliances goes back over 70 years. Thus, based on the secret alliances we do observe, we would expect such alliances to have become public at this point.

Second, the overall trends should still be valid assuming there is not a time-varying change in the propensity of secret alliances to become public. Furthermore, any time-varying change would have to be significant to overturn the large broader trends we observe.

Finally, we observe even sensitive secret alliances formed in the postwar era become public. For example, historians have extensively documented highly sensitive secret pacts such as the Protocol of Sevres, which was a secret agreement between Israel and France that secured French support in the ensuing Suez Crisis. In the years since the crisis, historians have uncovered confirmatory evidence of not only the agreement, but its specific text and the circumstances of its negotiation (Shlaim, 1997). This suggests that the ATOP data does capture non-trivial cases of secret alliances.