Relative power and the risk of conflict*

James D. Fearon
Stanford University

NOT QUITE COMPLETE DRAFT – Comments appreciated

September 3, 2021

Abstract

Is the risk that bargaining will fail and costly conflict occur greater when observable indices of relative power are balanced or out of balance? Or should we expect conflict risk to be insensitive to the balance of power? Using a mechanism-design approach adapted to specific features of anarchical settings (no third-party enforced commitments), I identify the lowest feasible risk of conflict when parties observe binary private signals bearing on their chances of winning in a fight. I find that quite generally, extreme imbalance in observable capabilities should associate with a low or zero risk of conflict. By contrast, for balance or moderate imbalance, the risk of conflict depends on how sensitive military odds are to small changes in relative capabilities. When small changes in relative power have a large effect on military odds, conflict is more likely in the optimal mechanism when power is roughly balanced. When defense is more advantaged in the sense that one needs a larger imbalance in forces to get a good chance of winning, the risk of conflict may be zero for an interval around 50-50 and highest for an intermediate imbalance of (ex ante observable) factors. For lower stakes or higher conflict costs, relative defense dominance can make the risk of conflict insensitive to the observable balance of forces, except at the extremes.

1 Introduction

Ordinarily, if a buyer and seller cannot agree on a price then the result is no trade. They just walk away. By contrast, in many contexts that are often described as "political," bargainers can attempt to seize part or all of the good in question if not satisfied with negotiations. They

*I am very grateful to Gabriel Carroll, Alex Debs, Amanda Kennard, Adam Meirowitz, and Robert Powell for comments and conversations at various stages of this long-suffering project.

can fight, in various fashions. Expectations about what would happen in a fight influence prior negotiations.

Examples: States bargain over the division of territory or other high-stakes issues, either during or in the shadow of war. Domestic political factions bargain over control of government offices either during or in the shadow of civil war. Parties in Congress bargain over the terms of a budget resolution in the shadow of a governmental shutdown. Litigants in a civil dispute bargain over terms of pre-trial settlement in the shadow of going to court.

How do observable indicators of prospects in the fight influence the likelihood of a negotiated agreement? On the one hand, one might expect that costly conflict would be most likely when observable indicators are relatively balanced, because it is then more likely that both sides are optimistic about their chances. On the other hand, perhaps an imbalance of observable indicators tempts the apparently stronger side to aggression. Alternatively, perhaps both effects are present, rendering the risk of costly conflict insensitive to observable measures of relative prospects in a fight.

In this paper I use a mechanism-design approach to ask about the relationship between relative power and the risk of conflict, in situations where the risk arises from the parties having private information about their odds in a fight. For concreteness I use the example of states bargaining over, say, territory. But the problem is more general, arising in a many political contexts.

The states observe private signals concerning their military prospects, and then choose "crisis bargaining strategies" that yield a chance of war and a chance of a negotiated settlement. Negotiated settlements must satisfy an ex post participation constraint, meaning that the states must prefer any realized settlement to their value for fighting *after* having updated based on observing the other side's crisis bargaining strategy.

This mechanism-design approach permits a characterization of the "least war-prone crisis," which has the lowest ex ante risk of war possible given the requirements of incentive compatibility and ex post stability.

The relationship between publicly observable indicators of relative power and the risk of conflict in the least war-prone crisis proves to depend on a feature of the fighting technology known as "the offense-defense balance" in Political Science (Jervis, 1978; Glaser and Kaufmann, 1998), or "decisiveness" in Economics (Hirshleifer, 2000). This is the sensitivity of the odds of winning to small changes away from balanced capabilities.

When small changes in relative power have a large effect on military odds, conflict is more likely in the optimal mechanism when power is roughly balanced. When defense is more advantaged in the sense that one needs a larger imbalance in forces to get a good chance of winning, the risk of conflict may be zero for an interval around 50-50 and highest for an intermediate imbalance of (ex ante observable) factors. Alternatively, when the stakes are higher, relative defense dominance can make the risk of conflict insensitive to the observable balance of forces, except at the extremes. Indeed, a general result is that regardless of the offense-defense balance, for a sufficiently extreme imbalance in observable indicators, the risk of conflict in the least war-prone bargaining process goes to zero.

These results are established using two different models of how the bargainers form estimates of their prospects in a fight. In the more standard case of *private capabilities*, they privately observe some facts about their military capabilities, in addition to publicly observed things like aggregate army size or terrain.

In a second model – *private analysis* – they form estimates of their odds based on fallible analysis of a very complex problem. For example, they observe the results of privately conducted war games, which are like random draws from a true, underlying probability of winning. Here, observable indicators of relative power determine the common prior belief about the underlying probability, and the parties' private signals are correlated (which need not be true for private capabilities).

The classic mechanism-design analysis of buyer-seller bargaining (Myerson and Satterthwaite, 1983) has been developed for conflict settings by Banks (1990), Bester and Wärneryd (2006), and Fey and Ramsay (2007, 2011). More recently, Hörner, Morelli and Squintani (2015) and Meirowitz et al. (2019) use the approach to analyze the strategies and possibilities for international mediation of interstate disputes.

In contrast to these articles, this paper develops the distinction between public and private signals of prospects in a fight, formalizing these and the relationship between them in two different ways. This is necessary in order to study the relationship between observable measures of relative power and the likelihood of conflict. In addition, to get results on the impact of varying relative power, we have to solve the mechanism design problem with asymmetric players – asymmetric in that they have different observable capabilities and, possibly, values for what is at stake. The existing literature in this area either does not solve for the optimal mechanism or considers more tractable symmetric cases, precluding results on observable measures of relative power and the risk of conflict.

Another difference concerns the ex post participation constraint, the requirement that to be implementable, negotiated deals must be preferred to fighting after bargaining strategies have been observed. In the classic mechanism-design models of buyer-seller bargaining, it is assumed that players can commit in advance to abide by any division of the good that the mechanism (or "mediator") recommends. In these settings, such ex ante participation constraints are justified by the idea that contracts can be used, or that we want to know what would be possible if contracts could be used.

The analysis here agrees with Fey and Ramsay (2011), who argued that for interstate conflict settings, states cannot commit to implement agreements that they see as worse than fighting after the fact. They suggested using an ex post condition they called "voluntary agreements" that is similar to the ex post constraint used here.[1] This is plausibly the case not just for interstate bargaining but for many conflict bargaining contexts, including in domestic politics.

The ex post condition can make conflict strictly more likely than if the parties could commit in advance to deals suggested by a mediator. The reason is that if bargaining reveals that one state saw an unfavorable signal about its military prospects, the other state has to get enough to not want to use force given this new information. That means that the temptation to "act tough" is greater with the ex post constraint, which means that the risk of conflict must be greater in order to convince players that acting tough signals that the other really is strong and so concessions are warranted.

This observation suggests a reason for why bargaining in anarchical or quasi-anarchical contexts would be on average more conflict-prone than in settings where ex ante contracts are feasible. Inability to commit (anarchy) proves to worsen the bargaining inefficiency resulting from private information, by increasing the gains from "acting tough" if there is a negotiated settlement.

In a remarkable result, Hörner, Morelli and Squintani (2015) show that an ex post participation constraint does *not* bind if the states observe only the recommended negotiated settlement, and settlement proposals can be randomized in a particular way that is outside the control of the states – for example by a literal mediator. By contrast, the analysis here follows the idea of "posterior implementability" introduced by Green and Laffont (1987), in the assumption that the players observe each other's bargaining strategies. This means that if different types choose different strategies, the players can update, whereas in the mediation models, the states observe only the mediator's proposal, which can conceal information. If we think of a mechanism as representing what is possible in a class of actual bargaining games, then it seems natural (or certainly reasonable) to assume that the players observe the other side's bargaining strategy.[2]

[1] The analysis in Fey and Ramsay (2011) focused almost entirely on implications of incentive compatibility, so that the implications of the ex post condition were not explored.

[2] As Hörner, Morelli and Squintani (2015) note, their mediator "must be able to commit to not seek a peaceful agreement in all circumstances" (1494). A mediator who just wants peace – plausibly the case for most international mediators – will want to secretly deviate from the optimal mechanism by making recommendations that induce peace for sure. The parties cannot observe this and it is hard to see how reputational concerns could

The next section provides motivation by reviewing the empirical and theoretical literature on the balance of power and interstate war. Section 3 states the mechanism-design problem and discusses the two models of private information about chances of winning, private capabilities and private analysis. Section 4 gives general results on how relative power influences whether the first-best outcome of peace-for-sure is possible, and section 5 then identifies and interprets the optimal mechanism in the case of symmetric players. Section 6 gives results for asymmetric players, meaning for imbalanced ex ante measures of relative strength. Section 7 concludes.

2 The balance of power and the likelihood of war

As noted, the question of how observable measures of relative power in a "fight" affect the odds of a deal appears in many more contexts than international politics. By way of motivation, however, in this section I briefly review some of the literature from this domain, where the question has probably had the longest run.

Authors associated with the Realist tradition have often held that war is most likely when power, or military capabilities, are *out of balance*, because the stronger side is then tempted to aggression. For examples, see Claude (1962, 61-63), Kissinger (1979, 55,67), and Mearsheimer (1990).[3]

In direct contrast another set of scholars, sometimes also described as Realists, has taken the opposite position. Simmel (1904), Organski (1959), Blainey (1973), and Waltz (1979) argued that war was most likely when observable measures of power were balanced, as then it is more likely that both sides will be optimistic about their chances. If power is clearly out of balance, they expect the weaker side to realize this and concede without a fight. These arguments assume that the parties to potential conflict are privately trying to estimate their chances, and that they may get it wrong, for one reason or another.

Blainey's treatment is the most developed. He argues more generally that war is caused by *mutual optimism* about chances of winning, and that peace "breaks out" when combatants

work to keep mediators "honest" given the large stakes, the infrequency of these events for any one mediator, and lack of observability of mediator deviations. Another crucial assumption for Hörner et al.'s result is that the mediator can cause the states to go to war by "walking away" from negotiations. In effect, the states can be committed to fight by participating in the mediation. It is not clear, however, why a mediator walking away is not simply a prelude to crisis bargaining in the sense of military moves and direct offers by the parties (tacit or explicit). After the mediator's decision to walk away reveals information about what the mediator must have heard, the relevant continuation payoffs should arguably be associated with an actual crisis bargaining game that follows, not immediate war.

[3]Morgenthau (1967, e.g., 204) often linked balance to peace but was not very consistent on this point.

learn from fighting what the true balance of power between them is. Mutual optimism is more likely, he argues, when observable measures are more equal.[4]

Last, in take-it-or-leave-it (tili) bargaining models of interstate conflict, one state chooses a continuous demand, perhaps an amount of territory, which the other state can then either acquiesce to or go to war to reverse. These models of the strategic situation suggest that *the conflicting forces behind these two intuitions tend to cancel each other out.* The stronger the first state, the bigger the demand it makes ("temptation for aggression"), but the bigger the concession that the state receiving the demand is willing to allow.

For example, the Athenians and Melians agreed that the Athenians were much more powerful, and the Athenians accordingly made extreme demands on the Melians. As it happened these went just beyond what the Melians were willing to accept, rather than chance a fight hoping to force less extreme demands. If the Athenians and Melians had been more equal in observable capabilities, the Athenians would have made less extreme demands, but this would not necessarily have implied a lower risk of war because the Melians would have been more willing to try war for any given set of demands.

If states have linear (risk neutral) preferences over territory, the two effects exactly offset each other and the equilibrium risk of war is independent of the observable balance of power.[5] In these models, the probability of war depends not on the observable balance of power but on the demanding state's value for what is at stake relative to the costs of fighting, which determines how much risk of rejection it is willing to run when it makes its demand.

Fearon's tili model is a model of a fait accompli attempt. The demanding state seizes territory and the other side can either accept or fight. Powell (1996) assumes instead that both sides must agree to change an exogenously fixed status quo. In this type of model, the risk of war depends on whether the distribution of benefits in the status quo is out of alignment with the distribution of military power. When benefits and relative power are in alignment, neither state has a credible threat to go to war and the status quo is stable. But if something happens to create uncertainty about whether one side is still willing to live with the status quo versus

[4]In line with Blainey's treatment, in this paper I use "mutual optimism" to refer to the sum of the player's estimates of their chances of winning a fight after getting private information about their conflict prospects, but before bargaining. That is, mutual optimism refers to "interim" beliefs in the game-theoretic sense, and mutual optimism is a matter of degree rather than a yes-or-no quality. On the question of whether and under what circumstances yes-or-no optimism (beliefs that sum to more than 1) can be maintained by Bayes rational players through the course of bargaining and decisions to go to war, see Fey and Ramsay (2007), Slantchev and Tarar (2011), and Debs (2020). This is perfectly possible in the mechanism-design model analyzed here, as is likely to be the case for any model in which bargainers can take actions that commit themselves to fight with positive probability, as suggested by Slantchev and Tarar (2011).

[5]Fearon (1992, 1995). Wittman (1979) saw and characterized the offsetting forces in a model of intra-war negotiations.

trying conflict, such as a shift in observable capabilities, then the risk of bargaining failure and war goes up. As in the tili model but for a different reason, the status-quo model does not suggest any consistent empirical relationship between the distribution of observable capabilities and the risk of war.

Over the years a substantial body of empirical work has developed that accepts Blainey's and the traditional balance-of-power arguments as given and tries to evaluate their validity against data. The results are mixed. Some studies find support for the hypothesis that power imbalances favor peace. Others support the contrary claim. A third group of studies finds no significant relationship in either direction. Perhaps overall there is tendency in favor of the imbalance-favors-peace side, especially for extreme imbalance.[6]

But if there *is* a relationship between the dyadic balance of power and the likelihood that a dispute will escalate to war, it cannot be a very strong one. With few exceptions, even when the quantitative studies have found a statistically significant relationship between relative power and dispute escalation, relative power variables have accounted for little of the variance. States have often fought when relative capabilities were roughly balanced and often when they were not.[7] In a 1989 review essay, Jack Levy concluded that "[Much] evidence suggests that the dyadic balance of power between two states is a poor predictor of the probability of war between them ..." (Levy, 1989, 242). Subsequent research does not seem to change this assessment.

The lack of a strong relationship might be related to the offsetting dynamics identified in the bargaining models that generate conflict risk from incomplete information about "resolve," or value for the stakes versus costs of fighting (Fearon, 1995; Powell, 1996). Another possibility is that the raw correlations are mixing together heterogeneous effects. Perhaps power parity favors conflict in some circumstances and peace in others. The analysis below suggests that one such factor could be the offense-defense balance.

---

[6]Supporting the balance-of-power hypothesis: Wright and Wright (1983), Ferris (1973), Siverson and Tennefoss (1984). Supporting power preponderance hypothesis: Moul (1985, 1988, 2003), Garnham (1976*a*,*b*)), Weede (1976), Organski and Kugler (1980) (with the caveat that parity promotes war only when a "power transition" is taking place). John Oneal (pers. comm.) observes that when the log of the ratio of the capabilities of the stronger to weaker state is included in typical MIDs regressions, it almost always takes a negative sign, consistent with power-preponderance. (e.g., Oneal and Russett, 1999). Supporting neither: Karsten, Howell and Allen (1984), Maoz (1983). In a lab experiment, Herbst, Konrad and Morath (2017) find that the risk of conflict is unaffected by the distribution of power when a mediator proposes the division, but that power imbalance favored conflict when the players made their own demands in a Nash demand game-type set up. In their experiment probabilities of prevailing in conflict were determined in a second-stage conflict model with endogenous investments in fighting effort, which adds some complexity to interpreting results.

[7]The exception is Moul (1988), who found that major powers in Europe almost never fought in the period 1815 to 1939 when capabilities were out of balance, according to his measure. By restricting his attention to major powers, Moul throws out a class of cases – major-minor disputes – with larger variance on the independent variable. If the relationship is as strong as his analysis suggests, then there should never be wars between minor and major powers, which is not so.

Figure 1: The frequency of MIDs as a functions of GDP ratios

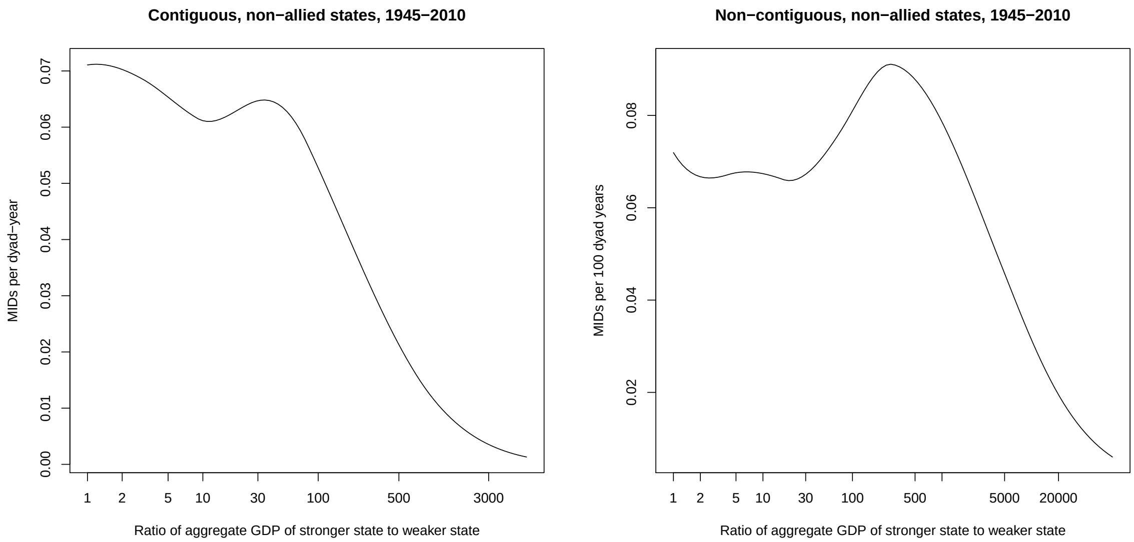


Figure 1 provides a relatively non-parametric look at the empirical relationship between the frequency of serious international disputes between pairs of states, and the ratio of the larger state's aggregate GDP to the smaller state's aggregate GDP for the period 1945-2010. The left panel shows the relationship for contiguous, non-allied dyads while the right panel shows it for non-contiguous, non-allied dyads. In both cases we see that for sufficiently extreme imbalance, the rate of serious disputes goes to zero, and more quickly for the *contiguous* dyads. We also see that in both cases, the rate of conflict is fairly insensitive to this measure of aggregate capabilities until the imbalance gets rather extreme – 30-to-1 for both contiguous and non-contiguous dyads.[8]

The rate of conflict appears maximized at equal GDPs for contiguous states, but is non-monotonic and is maximized for quite unequal states when they are not contiguous (ratio of about 400 to 1).

I consider these patterns in light of the theoretical results further below. To foreshadow, non-contiguity and distance make for relative defense dominance and higher costs of fighting, so that roughly equal-sized, non-contiguous states have little ability to take a fight to the other. Contiguous dyads have greater offensive advantage, and as expected show a mainly decreasing

[8]This is from the MIDs 3.01 data set [need to update], using GDP estimates from an updated version of the estimates used in Fearon and Laitin (2003) (which in this case is mainly from World Bank constant dollar GDP estimates). The dependent variable is 1 if there was at least one militarized interstate dispute initiated by one state in the dyad-year, and zero if not. The smooth is produced by the R package locfit with default smoothing parameters. Dyads are considered in non-allied in a year if they have neither a COW defense alliance nor an entente.

rate of conflict with the degree of power imbalance. Nonetheless, in both cases the frequency of militarized disputes does not vary that much with relative power until the power imbalance is really extreme. The patterns displayed in Figure 1 are consistent with the mechanism-design implications if on average contiguous dyads have moderate defense dominance and higher stakes on average, relative to higher defense dominance and lower stakes for non-contiguous dyads.

3 Conflict bargaining with implicit threats that leave something to chance

This section gives results for a mechanism-design approach to conflict bargaining in which the players observe private signals of the likelihood of military victory and then choose crisis bargaining strategies. Different strategy combinations generate different risks of escalation to war or, if war does not occur, different expected settlement terms.

The main results, presented in the following sections, show how to choose settlement terms and war risks to minimize ex ante expected war risk – this is the "least war-prone crisis" – subject to two sets of constraints. First, each state must do at least as well it could by adopting the crisis bargaining strategy that would be chosen if its leaders had seen a different private signal (incentive compatibility). Second, if war does not occur in the crisis, then each state's negotiated settlement payoff has to be at least as large as its expected payoff from attacking the other *given that* it has now seen the other's crisis bargaining strategy (ex post stability).

This approach "black boxes" how war risk is generated in an interstate or other type of coercive dispute. Implicitly, I am following Thomas Schelling's suggestion in imagining that the parties have ways to put themselves in positions where they are committed to use force – for example, by delegating to troops defending a fixed position. They have ways of "manipulating risk" in a "competition in risk taking" (Schelling, 1960, 1966). We ask what is the most efficient possible coercive bargaining interaction given the problem posed by private information about military prospects, and the mutual optimism that can accompany this.[9]

[9]Schelling was focused on nuclear crisis bargaining, for which the question of how threats to use force could be credible seemed particularly difficult. In Powell's (1990) formalization, nuclear war risk in a nuclear crisis is "autonomous," deriving from pure accidents, irrational panics or fury, or command-and-control failures. Fearon (1994) generates war risk by having crisis escalation increase a leadership's political costs for subsequently backing down.

3.1 Beliefs about military prospects

Formally, the interaction begins with states 1 and 2 privately observing either a favorable or unfavorable signal about their prospects in a war, denoted $a \in \{0, 1\}$ for state 1 and $b \in \{0, 1\}$ for state 2. Values $a = 1$ and $b = 1$ are the favorable signals, and we say that a player with a favorable signal is a "strong type" rather than a "weak type." For type combinations $ab \in \{00, 01, 10, 11\}$, the best estimate of the probability that state 1 would win a war is p_{ab}. That is, p_{ab} is the posterior belief if *both* signals were publicly known. I will use $p_{a\cdot}$ for state 1's interim belief about its win probability conditional on observing signal a, and likewise $p_{\cdot b}$ is state 2's interim belief about state 1's chances after observing signal b.

Nature chooses types. We will consider two ways of thinking about and representing how the states produce their military estimates, termed *private capabilities* and *private analysis*. In both cases, private signals allow the players to update their estimates to $p_{a\cdot}$ and $p_{\cdot b}$ from a common prior estimate that is based on publicly observable factors, such as their economies, militaries, or geography.

The observable factors are summarized by a number $m \in [0, 1]$ which is an index of state 1's relative military power based on any observables, which might include things like total military spending, troops, tanks, or terrain. Estimates depend on both the private signal and observable relative strength, so we can write $p_{ab}(m)$ as state 1's probability of winning when the signals are ab and the distribution of observable factors bearing on military prospects is m.

Private capabilities. The states privately observe some set of facts that bear on their military odds, such as quality of troops, training, or weapons, which amplify or diminish their prospects based on the observable balance, m, alone. For example, with a ratio-form contest-success function with $n > 0$ parameterizing the offense-defense balance, we could have

$$p_{ab}(m) = \frac{m^n}{m^n + (1-m)^n(1+\epsilon)^{n(b-a)}} \tag{1}$$

where $ab \in \{00, 01, 10, 11\}$ is drawn from a common prior distribution $f(ab)$, and $\epsilon > 0$ is a parameter indexing the importance of private information about capabilities relative to observable features. States form an updated estimate based on their private signal and their expectation of what the other saw, given the prior and their signal.[10]

Private analysis. Estimating what will happen in a significant military conflict is a notoriously

[10]With this formulation, when one state is strong and the other weak based on private signals, the strong side gets $n \log(1 + e)$ added to the log odds that it would win a war.

difficult problem. Consider two teams of expert military analysts working independently – general staffs, for example. Even if they had access to exactly the same information they might arrive at different estimates, simply due to the complexity and myriad uncertainties. I propose to model a state's private signal about its conflict odds as the result of its private, fallible analysis of the common situation, which includes publicly observable factors such as aggregate capabilities. For instance, the two general staffs might each observe the results of privately conducted war games.

Formally, it is as if Nature chooses the true probability that state 1 would win at war, $p \in [0, 1]$, according to a commonly known prior density $f(p)$ whose mean $\bar{p}$ reflects publicly observable relative capabilities. Neither state knows the true p. But they do each observe a single 0/1 draw from p and so can update based on the prior f and the draw. This is like observing the result of a war game or, more generally, the results of one's military analysis of the complex problem.

Thus state 1's signal is favorable ($a = 1$) if its draw is 1, and state 2's is favorable if its draw is zero. With some calculations we can show that $p_{1\cdot} = p_{\cdot 0} = \bar{p} + var(p)/\bar{p}$ and $p_{0\cdot} = p_{\cdot 1} = \bar{p} - var(p)/(1 - \bar{p})$. From this it follows that the states' interim beliefs are strictly compatible (sum to one) if one sees a favorable signal and the other an unfavorable signal, whereas there is mutual optimism if both happen to see favorable signals. There is mutual pessimism if both see unfavorable signals.[11]

With private analysis as the source of (possibly) conflicting beliefs, notice that the observable balance of power affects the likelihood of mutual optimism. An observably stronger state is more likely to get a favorable signal about the likely outcome of conflict, while an observably weaker state is more likely to get an unfavorable signal. The ex ante probability that the states will be mutually optimistic is $\bar{p}(1 - \bar{p})$, which is maximized at $\bar{p} = 1/2$ and goes to zero as one side becomes clearly more powerful.

3.2 Game form and mechanism

We represent the territory or issues that the states are in dispute over as a good of size 1, so that in a peaceful bargain state 1 gets $x \in [0, 1]$ and state 2 gets $1 - x$. If they go to war, state 1 gets the whole "pie" with probability p (from above), but they pay costs $c_1 > 0$ and $c_2 > 0$. Because the value of what is at stake has been normalized to 1, the cost terms should be understood as a measure of the states' perception of the *stakes* of the conflict. Smaller c_1

[11]Specifically, $p_{1\cdot} + 1 - p_{\cdot 1} = 1 + var(p)/\bar{p}(1 - \bar{p})$, so that $var(p)/\bar{p}(1 - \bar{p})$ is the degree of mutual optimism. (The variance of f is necessarily less than or equal to $\bar{p}(1 - \bar{p})$.

and c_2 mean that the perceived stakes are higher, and high costs mean that the stakes are low.

In the crisis bargaining game, each state chooses a (pure or mixed) strategy σ_a^1 and σ_b^2 as a function of their type, which are probability distributions on sets of pure strategies, S^1 and S^2. The strategy combination $\sigma_{ab} = (\sigma_a^1, \sigma_b^2)$ produces a probability of war $w(\sigma_{ab})$ and, if war does not occur, expected settlement terms $x(\sigma_{ab})$. Thus expected payoffs given σ_{ab} are

$$
\begin{aligned}
u_1(\sigma_{a\cdot}) &= \mathbb{E}_{b|a}(1 - w(\sigma_{ab}))x(\sigma_{ab}) + w(\sigma_{ab})(p_{ab} - c_1) \text{ and} \\
u_2(\sigma_{\cdot b}) &= \mathbb{E}_{a|b}(1 - w(\sigma_{ab}))(1 - x(\sigma_{ab})) + w(\sigma_{ab})(1 - p_{ab} - c_2).
\end{aligned}
$$

Let π_a^1 be state 1's interim belief that state 2 is a strong type, conditional on state 1 having gotten the signal a. Likewise, let π_b^2 be state 2's interim belief that state 1 is a strong type, conditional on state 2 seeing b. These are the relevant probabilities for the expectation terms above; they are primitives based on the process that generates private information about military odds. For example, in the case of private capabilities, if both states are equally likely to observe a favorable or unfavorable signal, and signals are uncorrelated, then $\pi_a^1 = \pi_b^2 = 1/2$ for all a and b. In the case of private analysis, $\pi_a^1 = 1 - p_{a\cdot}$ and $\pi_b^2 = p_{\cdot b}$.

In equilibrium, each type weakly prefers its strategy to what it would get by mimicking the other type's strategy, so we have the incentive compatibility conditions

$$u_1(\sigma_{a\cdot}) \geq u_1(\sigma_{1-a,\cdot}) \text{ for } a \in \{0,1\} \tag{IC_1}$$

$$u_2(\sigma_{\cdot b}) \geq u_2(\sigma_{\cdot,1-b}) \text{ for } b \in \{0,1\} \tag{IC_2}$$

In the classical mechanism design approach, we would assume that the states' expected payoffs from the crisis bargaining game must be at least as large as their expected value for war after observing their signals, thus $u_1(\sigma_{a\cdot}) \geq p_{a\cdot} - c_1$ and $u_2(\sigma_{\cdot b}) \geq 1 - p_{\cdot b} - c_2$. Further, we would assume that if the equilibrium of the game (or mechanism) meets this interim "participation constraint," then the states would be committed to implement whatever settlement terms $x(\sigma_{ab})$ emerged from the crisis, if war did not occur.

But as Fey and Ramsay (2011) argued in defining a constraint they termed "voluntary agreements," this makes no sense for the armed conflict setting. States or a government and rebel group talking about a peace deal cannot normally commit themselves not to use force after observing the other's negotiating behavior. So in this context it makes more sense to employ an ex post participation constraint: The states must prefer the settlement terms $x(\sigma_{ab})$ to going to war *given what they have learned from observing the other side's actions in crisis* *bargaining.* Formally,

$$x(\sigma_{ab}) \geq \mathbb{E}_{b|\sigma_{ab}}\, p_{ab} - c_1$$
$$1 - x(\sigma_{ab}) \geq \mathbb{E}_{a|\sigma_{ab}}\, 1 - p_{ab} - c_2$$

If different types choose different pure strategies in equilibrium, then the crisis interactions inform them about the other side's private signal. In this case, the ex post participation constraint is the condition that a negotiated settlement $x(\sigma_{ab})$ must be in the "bargaining range" – the set of agreements both prefer to fighting – based on the new, common estimate of relative power, p_{ab}. That is,

$$x(\sigma_{ab}) \in [p_{ab} - c_1, p_{ab} + c_2] \text{ for all } ab. \qquad (IR^{sep})$$

On the other hand, if different types choose the same actions in the crisis, then nothing is learned, and the ex post constraint requires only that the states prefer the negotiated settlement to their interim values for conflict, $p_{a\cdot} - c_1$ and $1 - p_{\cdot b} - c_2$. Note further that choosing the same actions for any signal entails that the negotiated settlement $x(\sigma_{ab})$ must be the same for all situations ab.[12] Since optimistic types have the highest interim values for conflict, it follows that if different types choose the same crisis-bargaining strategies, then there can be positive probability of a negotiated settlement when both sides get favorable signals (mutual optimism) only if there agreements x such that

$$p_{1\cdot} - c_1 \leq x \leq p_{\cdot 1} + c_2, \text{ or} \qquad (IR^{pool})$$
$$p_{1\cdot} - p_{\cdot 1} \leq C \equiv c_1 + c_2.$$

4 Sufficient imbalance of observable military prospects makes for peace

Three straightforward but interesting general results follow concerning the observable balance of power and when the first-best outcome of no conflict is possible.[13]

Proposition 1. *(i) Consider situations in which it is known that there are agreements that (interim) mutually optimistic types prefer to war (IR^{pool} holds). Then the least war-prone crisis has zero risk of conflict. Moreover, in situations where $p_{1\cdot} - p_{\cdot 1} > C$ (i.e., the condition fails), there must be positive risk of conflict in the least war-prone crisis.*

[12] Or really, the distribution on x's must be the same, so that nothing about ab is learned by observing x.

[13] All proofs are in the Appendix.

(ii) Consider the case of private capabilities and any military technology such that $p_{ab}(m)$ approaches 1 or zero as the observable balance of power, m, approaches 1 or zero, respectively. Then for m close enough to 1 or zero, the probability of conflict in the least war-prone crisis is zero.

(iii) Consider the case of private analysis described above. As the prior belief that state 1 would win, based on observable factors, approaches 1 or zero, the probability of conflict in the least war-prone crisis approaches zero.

Part (i) is a familiar result in mechanism-design models of bargaining with correlated values (Bester and Wärneryd, 2006; Fey and Ramsay, 2011; Deneckere and Liang, 2006).[14] Loosely, it says that we can guarantee peace only if there are deals that even the toughest types prefer to armed conflict when the other's behavior tells them nothing new. Substantively, this implies that in such situations, the least war-prone crisis requires that the states are prepared to commit themselves to fight with high probability if the other tries to push for "too much." It is interesting that optimal crisis behavior in these situations has the flavor of "we can't allow them push us to a deal worse than x when we know that they prefer x to fighting."

Parts (ii) and (iii) show that for both private capabilities and private analysis as sources of mutual optimism, highly unbalanced *observable* conflict capabilities make for zero or very low risk of war. The logic is somewhat different in the two cases. With private capabilities, greater imbalance in observable capabilities means that ultimately a state's chances of victory become insensitive to its private information, so that IR^{pool} obtains.

With private analysis, by contrast, a very surprising signal can make a state that is strongly disadvantaged by the observable balance of power highly optimistic, so that IR^{pool} would certainly fail. But this situation becomes increasingly unlikely as the prior odds increasingly favor one side, so the ex ante chance of conflict goes to zero.[15]

Parts (ii) and (iii) of Proposition 1 do *not* say that, in general, the risk of armed conflict is decreasing as observable indicators of relative power grow more imbalanced. They say only that this is true for a sufficiently large imbalance. Section 6 below returns to this question.

[14]In our context, if the bargaining does not reveal information about types, then the ex post constraint is the same as the interim constraint that is applied in the standard analysis (that assumes the parties can commit to implement any deal that emerges).

[15]In other words, with a very large disparity in observable capabilities, it will be rare that the weaker side's "war game" produces a positive result.

5 Ex ante balanced observable capabilities: A price of anarchy is less efficient bargaining

The next step is to characterize the least war-prone crisis when IR^{pool} fails, in which case the first-best of zero conflict risk is not possible.

Define a mechanism $M = \langle w_{ab}, x_{ab} \rangle$ for this problem to be a set of eight numbers in the $[0, 1]$ interval, and note that any given crisis bargaining game and strategies σ_{ab} imply a mechanism, using $w_{ab} = w(\sigma_{ab})$ and $x_{ab} = x(\sigma_{ab})$. Let $\bar{p}_i$ be the prior probability that state i observes a favorable signal.[16] If a fully separating mechanism is optimal, then it is the M that minimizes the ex ante expected probability of conflict subject to the incentive and ex post participation constraints:

$$\min_{\langle w_{ab}, x_{ab} \rangle} \bar{p}_1 \bar{p}_2 w_{11} + \bar{p}_1(1 - \bar{p}_2) w_{10} + \bar{p}_2(1 - \bar{p}_1) w_{01} + (1 - \bar{p}_1)(1 - \bar{p}_2) w_{00}$$

$$\text{such that } IC_1, IC_2, \text{and } IR^{sep}.$$

The Appendix presents complete solutions for the private capabilities and private analysis models described above. When the observable balance of power is unequal, the solutions are very involved. Optimal choice of w_{AB} and x_{AB} is a nonlinear programming problem in which several different constraints can bind (or be slack) depending on parameter values.[17]

To convey basic implications and intuition, this section concludes with a proposition characterizing the least war-prone crisis for the symmetric case, in which the prior military odds are even and $c_1 = c_2 = c$. It turns out that with symmetric players, the least war-prone crisis is the same for the private capabilities and private analysis "technologies" (no longer true with asymmetric players).

Proposition 2. *Consider the symmetric case in which the two players have equal observable capabilities, equal views of the stakes at issue, and the same chance of winning in conflict if one sees the favorable and other the unfavorable private signal. That is, using either the private capabilities or private analysis models of military estimates,* $c = c_1 = c_2$, $p_{10} = 1 - p_{01}$, *and*

[16] With private analysis, $\bar{p}_1$ and $\bar{p}_2$ are pinned down as $\bar{p}$ and $1-\bar{p}$, where $\bar{p}$ is the mean of the prior distribution f. With private capabilities, $\bar{p}_i$ could be any number in $[0, 1]$, although perhaps $\bar{p}_i = 1/2$ – favorable and unfavorable signals are equally likely – is the most useful case to consider. (I find it hard to imagine empirical measures for prior beliefs about the likelihood that the other state gets a favorable signal about its capabilities.) Hörner, Morelli and Squintani (2015) use $q = \bar{p}_1 = \bar{p}_2$ in a related private-capabilities set up.

[17] The objective function is linear but w_{11} multiplies x_{11} in the incentive constraints. After the general results that $w_{00} = 0$ and that the ex post participation constraints bind for x_{10} and x_{01}, there are two relevant incentive constraints and five endogenous variables with upper and lower bounds (w_{11}, w_{10}, w_{01}, x_{00} and x_{11}). This leads to multiple cases, often with non-linear boundaries.

$p_{1\cdot} = 1 - p_{0\cdot}$. *The least war-prone crisis in this symmetric case has the following features.*

- *For any parameters, when both sides observe unfavorable signals (mutual pessimism) war does not occur ($w_{00} = 0$). This result holds in the asymmetric case as well, thus, in general.*
- *By symmetry, $x_{00} = x_{11} = 1/2$.*
- *If $c \geq \bar{c} \equiv p_{1\cdot} - 1/2$, then the optimal mechanism has both states choosing the same crisis bargaining strategy regardless of their private signal, and zero risk of war. This condition is equivalent to saying that the stakes are known to be low enough that there are deals both prefer to fighting even if both sides are privately optimistic about their odds in a fight.*[18]
- *If $c < \bar{c}$ then the ex post participation constraints always bind when they choose different crisis strategies, so that $x_{10} = 1 - x_{01} = p_{10} - c$. In words, if the crisis reveals that one side saw favorable information and the other unfavorable, the "strong" player gets just enough to be willing to accept.*[19]
- *Further, if $c < \bar{c}$ then in the optimal crisis the states condition their bargaining strategy on their signal – a favorable signal causes a state to run a higher risk of conflict – and there is a positive risk of war. Let o be the prior odds that a state gets a favorable signal in the case of private capabilities, or, in the case of private analysis, the interim odds with which a state that gets a favorable signal thinks it will win.*

  If $c \in [\underline{c}, \bar{c})$ where $\underline{c} = (p_{10} - 1/2)/(1 + 2o)$, then $w_{10} = w_{01} = 0$ and

  $$w^*_{11} = \frac{o+1}{o}\frac{p_{10} - 1/2 - c}{p_{10} - 1/2 + c}.$$

  If $c < \underline{c}$, then $w_{11} = 1$ and

  $$w_{10} = w_{01} = \frac{p_{10} - 1/2 - c(1 + 2o)}{p_{10} - 1/2 - 2co}$$

To summarize, if costs are high enough or equivalently if stakes are low enough, both sides adopt relatively conciliatory crisis bargaining approaches regardless of their private information about military prospects. They divide the "pie" and do not risk armed conflict. This good

[18] In the general (possibly asymmetric) case, this condition is $c_1 + c_2 \geq p_{1\cdot} - p_{\cdot 1}$, which is to say that it is common knowledge that there are agreements both prefer to conflict (a bargaining range) even if the states are interim1 mutually optimistic.

[19] This result holds in the general case as well, where $x_{10} = p_{10} - c_1$ and $x_{01} = p_{01} + c_2$.

outcome is made possible by the expectation that deviating to the more aggressive strategy will, despite the high costs of fighting, generate a high risk of escalation to a costly fight.

If costs are in an intermediate range, a state whose leaders get a favorable military signal adopts a tougher approach, and war occurs with positive probability if and only if *both* saw favorable signals (mutual optimism). A pessimistic state (weak signal) adopts a more conciliatory approach, running no risk of war and ceding a larger share if the other side presses its demands aggressively. In this range, if both saw poor military assessments, they are both conciliatory and they reach a negotiated settlement with no risk of escalation.

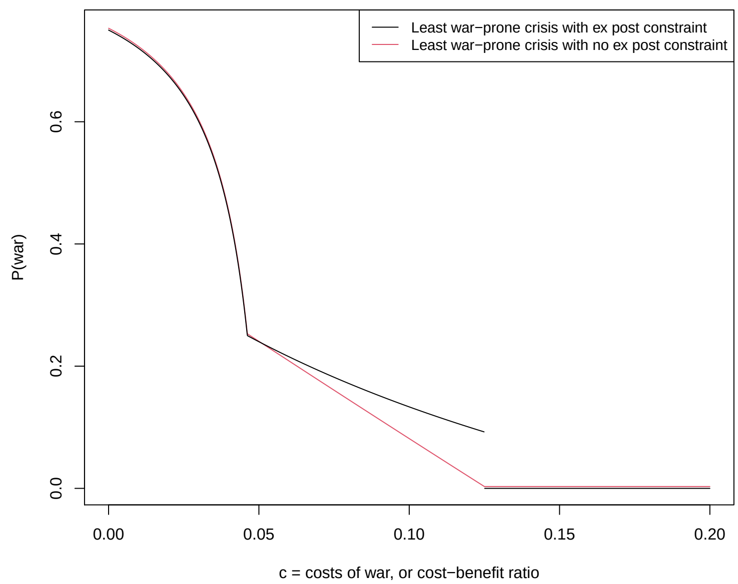
Last, if costs are low enough or stakes are high enough, crisis bargaining is necessarily more dangerous. When both sides get favorable military estimates, they both choose tough strategies and the result is war for sure. A state that gets an unfavorable military estimate chooses a more conciliatory approach, but even so this still produces a positive risk of war if the other side got a favorable signal. In this situation, war can result even though there is "strict mutual pessimism" in the sense that interim estimates sum to less than 1. When both sides saw unfavorable signals, they both choose conciliatory strategies and no war risk is generated.

Further, because in this last case crisis bargaining can result in war even when one side got an unfavorable military estimate, seeing war break out does not imply that one is facing a strong type. So in this case, when the states both got favorable signals about their military prospects, incompatible, mutually optimistic beliefs can remain even conditional on war occurring, since each thinks the other could still be a weak type.

To illustrate, Figure 2 shows the ex ante probability of war as a function of the costs c (equivalently, the cost-stakes ratio) for the symmetric case. Prior beliefs are that state 1's probability of military victory is equally likely to be 3/4 or 1/4. This implies that a state that sees a favorable signal updates to expect a 5/8 chance of winning, and a bad signal leads to the more pessimistic expectation 3/8. If the truth is that one saw "strong" and the other "weak," the best estimate would be that the strong type wins with probability .7 if this information were publicly known. So the ex post constraint requires that a negotiated settlement when one adopts the "tough" strategy and the other the more conciliatory strategy gives at least $.7 - c$ to the tough player. (For convenience, term the tough strategy S and the weak-type strategy W.)

The red line in Figure 2 is the probability of war in the least war-prone crisis *without* the ex post constraint. In that case as in classical mechanism design, the parties are assumed to be able to commit ex ante to implement any negotiated settlement ex post. Notice that for a middle range of costs, the ex post constraint implies they are strictly worse off than if they could tie their hands prior to negotiations. The reason is that they would commit to implement

Figure 2: The probability of war as a function the cost-stakes ratio, symmetric stakes and observable military capabilities



deals in the SW and WS outcomes that give less to the strong type than its after-the-fact value for fighting, once it knows the other side saw an unfavorable signal. Anarchy (here, the ex post constraint) means that the strong type has to get more if it learns the other saw bad information. But this makes bluffing more attractive for a weak type, which in turn means there has to be a higher chance of war in SW and WS or else all types would fight for sure.[20]

The break in the black line at $c = 1/8$ is where it becomes possible to support "pooling" of the tough and weak types, so that the first-best outcome of no conflict and an even split becomes feasible. With $c < 1/8$, tough types prefer taking their chance on war with an estimate of a 5/8ths chance of winning.

6 Imbalance of observable capabilities and the risk of conflict

As noted, with asymmetric capabilities or stakes, identifying the least war-prone crisis becomes much more difficult and the formal characterization is lengthy.[21]

[20]The appendix solves the model with ex ante commitment in the symmetric case ($\bar{p} = 1/2$ and equal costs).

[21]See Appendix. DRAFT NOTE: At this point I have complete characterizations for the private capabilities case in general, and the private analysis case when the prior beliefs about p are Beta distributed. I have not

Figure 3 presents results for the private-analysis case for three different costs/stakes ratios, and three different offense-defense balances. For this model, the states' leaderships have prior beliefs about p, the probability that state 1 would win a war, that are represented by a Beta distribution with mean $\bar{p} \geq 1/2$. In the real world such prior beliefs are based on publicly observable factors bearing on military prospects, such as aggregate GDP, terrain, the size and equipment of their armies, past performance, and so on. The leaderships then observe private 0/1 Bernoulli signals drawn from the true probability p. Thus if state 1 sees "1" it updates to an "interim" estimate of $p_{1\cdot} > \bar{p}$, and likewise if state 2 observes "0" it updates to an interim belief that state 1 would win of $p_{\cdot 1} < \bar{p}$. This would be a case of (interim) *mutual optimism*, since $p_{1\cdot} + 1 - p_{\cdot 1} > 1$.[22]

Qualitative features of the optimal mechanism with asymmetric capabilities or stakes are similar to the symmetric case discussed above. The probability of conflict is (weakly) highest when the leaderships are mutually optimistic, as Blainey (1973) argued. This is because optimism leads a state to chose a more aggressive crisis-bargaining strategy, which creates higher risk of escalation to war. Optimistic types are willing to run this higher risk in order to credibly distinguish themselves from weaker types, in order to get a better deal if the crisis does not result in war.

As before, if both sides happen to get "weak" signals, they both adopt conciliatory approaches in crisis bargaining and make a deal with no risk of escalation. If one side gets the strong signal and the other weak, this makes for a relatively lop-sided deal with no risk of war if the stakes are not very high, but some risk of conflict if the stakes are high enough.

The question of interest, however, is how the risk of escalation varies with the observable balance of forces, which in the private-analysis model is $\bar{p} \geq 1/2$.[23] Figure 3 shows that

- Consistent with Proposition 1, for extreme-enough observable imbalance, the probability of conflict in the optimal mechanism goes to zero, regardless of costs/stakes or the offense-defense balance.[24]
- The more advantaged is offense (in the sense that small changes in the balance of forces

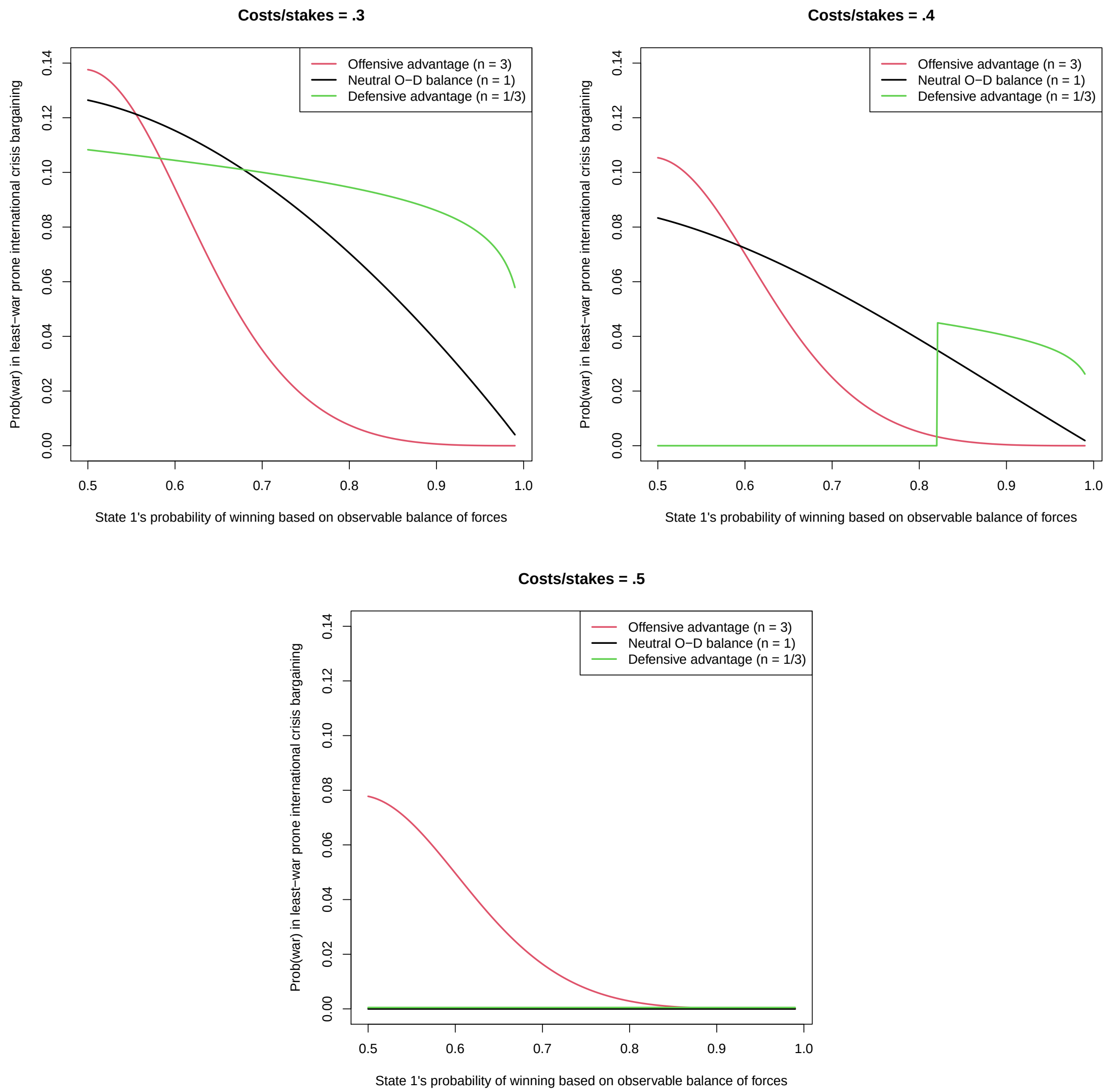
been able to complete the proof in one place (that $x_{10} = p_{10} - c_1$ and $x_{01} = p_{01} + c_2$ in any optimal mechanism), but numerical solutions show that this is true and that the analytical solutions I have are correct – at least for all cases I have checked!

[22] Formally, these examples use $\bar{p} = m^n/(m^n + (1-m)^n$ and $p_{ab} = (sm^n + a + 1 - b)/(sm^n + s(1-m)^n + 2)$, where s scales the relative importance of public information versus private information. This is from the updating rule for the Beta distribution $f(p)$ with parameters $\alpha = sm^n, \beta = s(1-m)^n$. Likewise, $p_{a\cdot} = (sm^n + a)/(sm^n + s(1-m)^n + 1)$. For the cases in Figure 3, $s = 1$, $n \in \{1/3, 1, 3\}$, and $C \in \{.2, .3, .4\}$.

[23] Of course, it is without loss of generality that we are making state 1 the observably advantaged state.

[24] The green lines should all go to zero at $\bar{p} \approx 1$ ....

Figure 3: The probability of war as a function of observable imbalance: Private analysis



away from parity have larger effects on the chance of winning), the more we see the highest risk of conflict between ex ante more equal adversaries, and the faster the risk declines towards zero as imbalance grows.

- With greater defense advantage, the risk of conflict is less sensitive to imbalance in observable capabilities, until imbalance becomes extreme. These specific cases do not show it, but if we continue to increase defensive advantage, the green line becomes more flat, making conflict risk almost independent of the observable balance up until extreme inequality.[25] At the same time, although defensive advantage makes conflict risk insensitive to relative power, it reduces conflict risk for any given balance of capabilities.

- Comparing across the three plots we see that, not surprisingly, increasing costs/stakes reduces the lowest feasible conflict risk for any given technology, and that for high enough costs (low enough stakes), the first-best outcome of a deal with zero risk of war becomes possible (third plot, for $n = 1$ and $n = 1/3$).

- Most interestingly, with sufficient defense dominance it can happen that the first-best outcome of zero war risk is feasible for adversaries with sufficiently equal observable capabilities, whereas there must be positive war risk for imbalanced forces that are not too extreme (second plot, $c_1+c_2 = .4$). The reason is that with strong defense dominance, private signals don't make relatively equal adversaries optimistic enough to want to run serious war risks – they are not "tempted to aggression," as in the classical balance-of-power argument. But when one side has a significant advantage, private signals can make the difference $p_{1\cdot} - p_{\cdot 1}$ (mutual optimism) large enough that both sides are willing to run war risk rather than accept the best feasible weak-type deal. Another way to say this is that with defense dominance, there can be a point where chances of victory become relatively sensitive to small shifts when there is a moderate imbalance in observable capabilities, like .85 in the second plot in the Figure.

- Observe that the offense-defense balance does not order the risk of conflict in a simple, monotone way. Between ex ante equal adversaries, greater offensive advantage implies greater risk of war. But between states with a moderate imbalance of observable capabilities, war risk can be significantly greater when there is greater defensive advantage!

  The reason is that with offensive advantage, a significant imbalance in observable capabilities implies a high likelihood that the stronger side would win at war, irrespective of private signals. With more defensive advantage, the weaker side still has a reasonable shot at victory with moderate imbalance, so that optimism can make running significant

[25] This is the same empirical implication as we get from the tili models with private information about costs/stakes, albeit from a different strategic mechanism.

conflict risk reasonable, to prevent exploitation by a weak (bluffing) type of the stronger party.

- Last, observe that for low-enough stakes (third plot), offensive advantage is needed to get any conflict risk at all in the "optimal crisis" situation. What really drives war risk in these models is the relationship between mutual optimism, $p_{1\cdot} - p_{\cdot 1}$, and the total costs/stakes ratio, $C = c_1 + c_2$. If costs/stakes are large, the only way to get conflict is if mutual optimism can be quite large, which requires that private signals have a large impact on a side's estimate of its chances of winning. When the parties also have observable measures available, this is equivalent to saying that offense must be strongly advantaged for private analysis of the military situation to matter enough.

  By contrast, for higher stakes, the observable balance of forces is effectively less consequential for the minimum risk of conflict (except at the extremes).

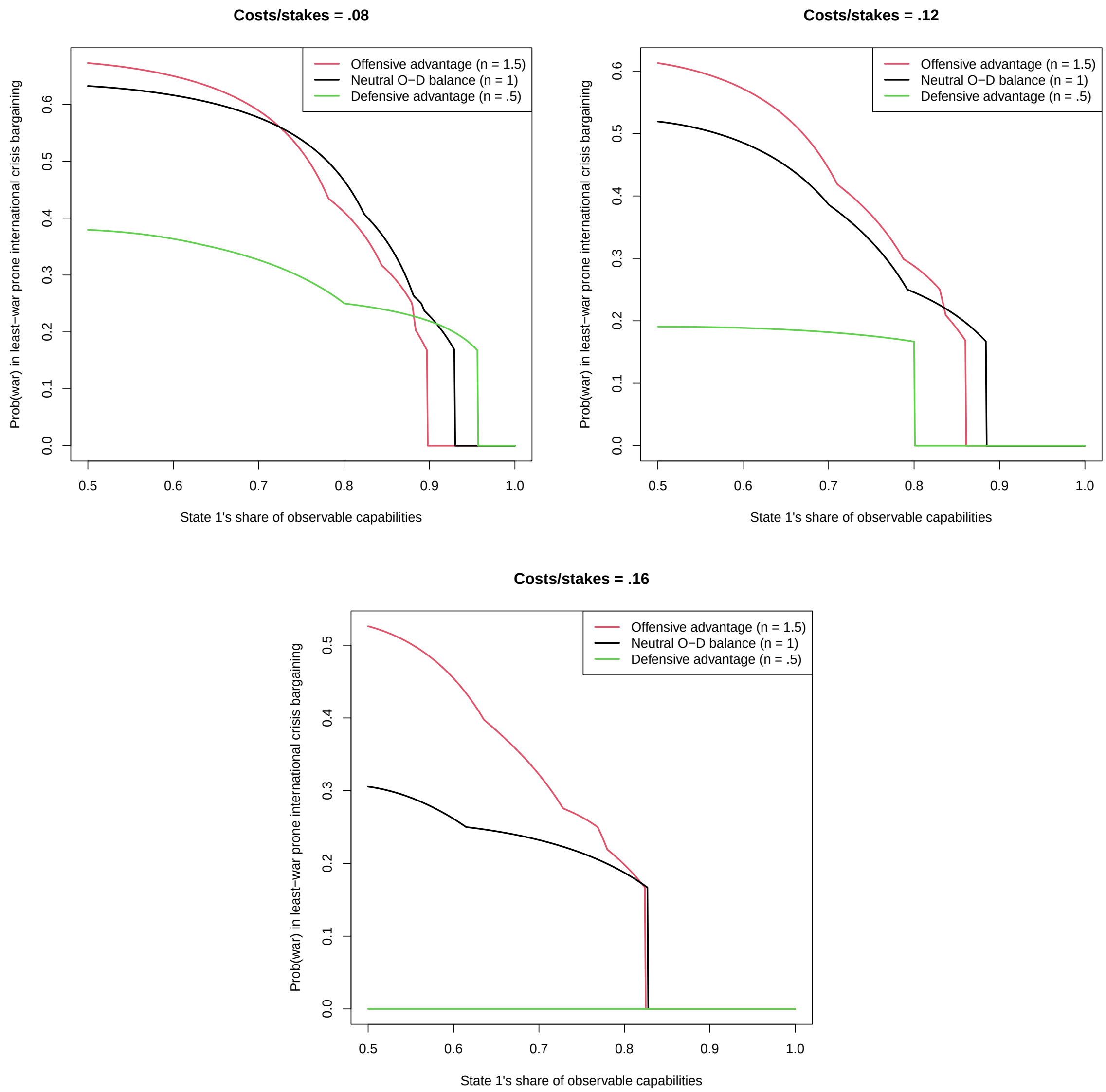
Figure 4 displays parallel cases for private capabilities. Here, the leaderships each observe a private signal of the quality of their forces. It is common knowledge that each side is equally likely to see "strong" or "weak," and their signals are not correlated. These examples use the ratio-form contest-success function in (1), so that the odds that state 1 wins at war if both see "strong" or both see "weak" are $(m/(1-m))^n$, where m is state i's observable capabilities and $n > 0$ is the offense-defense parameter. If state 1 sees "strong" and state 2 sees "weak," state 1's odds are increased by the factor $(1+\epsilon)^n$, where $\epsilon > 0$ reflects the importance of private information relative to observable capabilities (and likewise reduced by this factor if state 2 sees "strong" and 1 sees "weak").

The ratio-form technology with private capabilities has the characteristic that mutual optimism $p_{1\cdot} - p_{\cdot 1}$ is strictly decreasing and goes to zero as state 1's relative power m goes from 1/2 to 1.[26] As a result and as illustrated in the figure, there is always a degree of power imbalance $\bar{m} < 1$ such that the first-best of zero war risk is possible for any $m \geq \bar{m}$. With higher costs (equivalently lower stakes), $\bar{m}$ is smaller so that more moderate power imbalances conduce to peaceful bargaining.

[INCOMPLETE PART: I think it is true that for any military technology such that mutual optimism $p_{1\cdot} - p_{\cdot 1}$ is strictly decreasing and goes to zero as state 1's relative power m goes from 1/2 to 1, the probability of conflict in the optimal mechanism is weakly decreasing. But I have not gotten to trying to prove this yet, and it might only be easily showable for a specific technology, like the ratio form.]

[26] With private analysis, $p_{1\cdot} - p_{\cdot 1}$ is flat when $n = 1$, decreasing with offensive advantage ($n > 1$), and increasing with offensive advantage ($n < 1$). The latter is what is behind the result consistent with traditional balance of power theory in Figure 3.

Figure 4: The probability of war as a function of observable imbalance: Private capabilities



7 Conclusion

In *Leviathan*, Thomas Hobbes saw rough equality of capabilities as a principle cause of conflict in anarchy (chap. 13) and his larger argument can be seen as a ringing endorsement of the proposition that imbalance in capabilities favors peace. The might of Leviathan (the well-ordered commonwealth, or government) should deter any challengers.

Hobbes also argues that small coalitions (gangs) are not enough to ensure security in anarchy because with small numbers, "small additions to one side or the other, make the advantage of strength so great, as is sufficient to carry the Victory; and therefore gives encouragement to Invasion." In other words, with small armed groups, offensive advantage is significant, since small changes in relative capabilities can have a large impact on the chances of Victory (chap. 17). This makes small, equal-sized groups prone to violent conflict.

By contrast, with larger numbers "the odds of the Enemy is not of so conspicuous and visible moment, to determine the [outcome] of Warre, as to move him to attempt" (chap. 17). This is the traditional balance of power argument! With large gangs, such as states, small variations in force sizes do not have as large an effect on the probability of victory, so war is less likely. Hobbes saw interstate relations as an anarchy with a "disposition" to war, but not nearly so prone to violence as "the condition of Meer Nature" between households or clans, because interstate relations are marked by much greater defense dominance.

Hobbes' predictions concerning the relationship between observable capabilities and the risk of violent conflict are similar to the implications of the mechanism-design analysis above, based on asking what is the lowest possible conflict risk when the parties observe private signals of their prospects in a fight. Hobbes' main mechanism was first-strike advantages. But he clearly sees decisions to use force as based on estimates of prospects of success, and he saw these estimates as subject to error that he thought was increasing in equality of the balance of observable capabilities.[27]

The model of "private analysis" introduced here provides a way of formalizing the idea that parties to political bargaining may differ in their fight estimates due to the complexity of the problem, rather than private observations of the quality of their troops, a secret weapon, and so on. Although private information about raw capabilities surely exists and may sometimes be consequential, I believe that much evidence suggests that complexity is typically the more

[27]Distinct from Hobbes (whose analysis on this issue is admittedly just sketched), the mechanism design results do not imply that defense dominance plus equality of power necessarily makes for low risk of conflict. With low enough stakes, we still get maximum risk of war with balanced power, just not much change with greater imbalance.

important source of conflicting estimates and, sometimes, mutual optimism.[28]

References

Banks, Jeffrey S. 1990. "Equilibrium behavior in crisis bargaining games." *American Journal of Political Science* pp. 599–614.

Bester, Helmut and Karl Wärneryd. 2006. "Conflict and the social contract." *Scandinavian Journal of Economics* 108(2):231–249.

Blainey, Geoffrey. 1973. *The Causes of War.* New York: Free Press.

Claude, Inis L. 1962. *Power and international relations.* Random House.

Debs, Alexandre. 2020. "Mutual Optimism and War, and the Strategic Tensions of the July Crisis." *American Journal of Political Science* forthcoming.

Deneckere, Raymond and Meng-Yu Liang. 2006. "Bargaining with interdependent values." *Econometrica* 74(5):1309–1364.

Fearon, James D. 1992. Threats to use force: Costly signals and bargaining in international crises PhD thesis University of California, Berkeley.

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

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

Fearon, James D and David D Laitin. 2003. "Ethnicity, insurgency, and civil war." *American political science review* 97(1):75–90.

Ferris, Wayne H. 1973. *The power capabilities of nation-states: International conflict and war.* Lexington, Mass.; Toronto: Lexington Books.

Fey, Mark and Kristopher W Ramsay. 2007. "Mutual optimism and war." *American Journal of Political Science* 51(4):738–754.

[28]The formal model here maintains the common prior assumption, which effectively means that the parties would update based on the other side's analysis if they had direct access to it. Any number of biases might make players inclined to discount the the other side's analysis, at least in some circumstances. Upward bias in own estimates has obvious implications – greater risk of conflict due to more frequent mutual optimism.

Fey, Mark and Kristopher W Ramsay. 2011. "Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict." *American Journal of Political Science* 55(1):149–169.

Garnham, David. 1976*a*. "Dyadic international war 1816-1965: the role of power parity and geographical proximity." *Western Political Quarterly* 29(2):231–242.

Garnham, David. 1976*b*. "Power parity and lethal international violence, 1969–1973." *Journal of Conflict Resolution* 20(3):379–394.

Glaser, Charles L and Chairn Kaufmann. 1998. "What is the offense-defense balance and how can we measure it?" *International security* 22(4):44–82.

Green, Jerry R and Jean-Jacques Laffont. 1987. "Posterior implementability in a two-person decision problem." *Econometrica* 55(1):69–94.

Herbst, Luisa, Kai A Konrad and Florian Morath. 2017. "Balance of power and the propensity of conflict." *Games and Economic Behavior* 103:168–184.

Hirshleifer, Jack. 2000. "The macrotechnology of conflict." *Journal of Conflict Resolution* 44(6):773–792.

Hörner, Johannes, Massimo Morelli and Francesco Squintani. 2015. "Mediation and peace." *The Review of Economic Studies* 82(4):1483–1501.

Jervis, Robert. 1978. "Cooperation under the security dilemma." *World politics* 30(2):167–214.

Karsten, Peter, Peter D Howell and Artis Frances Allen. 1984. *Military threats: A systematic historical analysis of the determinants of success*. Westport, CT: Praeger.

Kissinger, Henry. 1979. *The White House Years*. New York: Little Brown.

Levy, Jack S. 1989. The Causes of War: A Review of Theories and Evidence. In *Behavior, Society, and Nuclear War*, ed. Philip E. Tetlock et al. Oxford University Press pp. 1151–1169.

Maoz, Zeev. 1983. "Resolve, capabilities, and the outcomes of interstate disputes, 1816-1976." *Journal of Conflict Resolution* 27(2):195–229.

Mearsheimer, John J. 1990. "Back to the future: Instability in Europe after the Cold War." *International security* 15(1):5–56.

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

Morgenthau, Hans Joachim. 1967. *Politics among nations: The struggle for power and peace.* New York: Knopf.

Moul, William. 2003. "Power parity, preponderance, and war between great powers, 1816-1989." *Journal of Conflict Resolution* 47(4):468–489.

Moul, William B. 1985. "Balances of Power and European Great Power War, 1815–1939: A Suggestion and Some Evidence." *Canadian Journal of Political Science/Revue canadienne de science politique* 18(3):481–528.

Moul, William Brian. 1988. "Balances of power and the escalation to war of serious disputes among the European great powers, 1815-1939: Some evidence." *American Journal of Political Science* pp. 241–275.

Myerson, Roger B and Mark A Satterthwaite. 1983. "Efficient mechanisms for bilateral trading." *Journal of economic theory* 29(2):265–281.

Oneal, John R and Bruce Russett. 1999. "The Kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885–1992." *World politics* 52(1):1–37.

Organski, A.F.K. 1959. *World Politics.* New York: Knopf.

Organski, A.F.K. and Jacek Kugler. 1980. *The War Ledger.* Chicago: University of Chicago Press.

Powell, Robert. 1996. "Stability and the Distribution of Power." *World Politics* 48(2):239–267.

Schelling, Thomas. 1966. *Arms and Influence.* New Haven: Yale University Press. Chapters 1-3.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* New Haven: Yale University Press.

Shimer, Robert and Iván Werning. 2019. "Efficiency and information transmission in bilateral trading." *Review of Economic Dynamics* 33:154–176.

Simmel, Georg. 1904. "The sociology of conflict. I." *American journal of sociology* 9(4):490–525.

Siverson, Randolph M and Michael R Tennefoss. 1984. "Power, alliance, and the escalation of international conflict, 1815-1965." *American Political Science Review* 78(4):1057–1069.

Slantchev, Branislav L and Ahmer Tarar. 2011. "Mutual optimism as a rationalist explanation of war." *American Journal of Political Science* 55(1):135–148.

Waltz, Kenneth N. 1979. *Theory of International Politics.* Boston: McGraw Hill.

Weede, Erich. 1976. "Overwhelming preponderance as a pacifying condition among contiguous Asian dyads, 1950–1969." *Journal of Conflict Resolution* 20(3):395–411.

Wittman, Donald. 1979. "How a war ends: A rational model approach." *Journal of Conflict Resolution* 23(4):743–763.

Wright, Quincy and Louise Leonard Wright. 1983. *A study of war.* University of Chicago Press.

8 Appendix [not complete]

This section provides complete solutions for the optimal (least war-prone) mechanism that satisfies the incentive compatibility and ex post participation constraints for the "private capabilities" and "private analysis" settings described in the text. I begin with some general results, for any "technology" of generating military prospects and private signals about them, and then develop full solutions for private capabilities and private analysis technologies.

8.1 The mechanism design problem

A mechanism is a list of eight numbers, $M = \langle (w_{ab}), (x_{ab}) \rangle$, where w_{ab} is the probability of conflict and x_{ab} is the share of the resource that state 1 gets when state 1 sees signal $a \in \{0, 1\}$ and state 2 sees $b \in \{0, 1\}$. All of these must be in $[0, 1]$.

p_{ab} is the best estimate of the chance that state 1 wins conditional on knowing the private signals ab. $p_{a\cdot}$ is state 1's interim belief about its chance of winning after observing signal a, and likewise $p_{\cdot b}$ for state 2. Let π^1_a be state 1's interim belief that state 2 is the strong type conditional on observing signal a, and likewise π^2_b is state 2's belief that state 1 is a strong type given signal b. Let $\bar{p}_i$ be the ex ante probability that state i gets the "strong" signal. These, along with the states' known costs for fighting c_1 and c_2, constitute the primitives of the *military technology*, which take specific form below when we analysis private capabilities and analysis settings.

The players' interim expected payoffs are

$$\begin{aligned}
u^1_a &= \mathbb{E}_{b|a}(1 - w_{ab})x_{ab} + w_{ab}(p_{ab} - c_1) \\
&= \pi^1_a \left((1 - w_{a1})x_{a1} + w_{a1}(p_{a1} - c_1)\right) + (1 - \pi^1_a) \left((1 - w_{a0})x_{a0} + w_{a0}(p_{a0} - c_1)\right) \\
u^2_b &= \mathbb{E}_{a|b}(1 - w_{ab})(1 - x_{ab}) + w_{ab}(1 - p_{ab} - c_2) \\
&= 1 - [\mathbb{E}_{a|b}(1 - w_{ab})x_{ab} + w_{ab}(p_{ab} + c_2)] \\
&= 1 - \left[\pi^2_b \left((1 - w_{1b})x_{1b} + w_{1b}(p_{1b} + c_2)\right) + (1 - \pi^2_b) \left((1 - w_{0b})x_{0b} + w_{0b}(p_{0b} + c_2)\right)\right]
\end{aligned}$$

The mechanism design problem is to minimize the ex ante expected chance of conflict subject to the incentive compatibility and ex post participation constraints.

$$\min_M \bar{p}_1\bar{p}_2 w_{11} + \bar{p}_1(1 - \bar{p}_2)w_{10} + (1 - \bar{p}_1)\bar{p}_2 w_{01} + (1 - \bar{p}_1)(1 - \bar{p}_2)w_{00}$$

such that, for $a, b \in \{0, 1\}$,

$$u^1_a \geq \hat{u}^1_a \equiv \mathbb{E}_{b|a}(1 - w_{1-a,b})x_{1-a,b} + w_{1-a,b}(p_{ab} - c_1) \qquad (IC^1_a)$$
$$u^2_b \geq \hat{u}^2_b \equiv 1 - \mathbb{E}_{a|b}\left[(1 - w_{a,1-b})x_{a,1-b} + w_{a,1-b}(p_{ab} + c_2)\right] \qquad (IC^2_b)$$

and the ex post participation constraints. These are most easily and clearly stated for what we can call the separating and pooling cases. In any separating case, different types choose different pure strategies in the crisis bargaining game, so that types are revealed. This implies the ex post constraints

$$x_{ab} \in [p_{ab} - c_1, p_{ab} + c_2] \text{ for all } ab. \tag{IR^{sep}}$$

In the full pooling case, different types choose the same pure or mixed crisis bargaining strategy, so that their posterior beliefs about the other's type are the same as their interim beliefs. This implies that the ex post constraint is the same as the usual interim participation constraint, and that $x = x_{ab}$ for all ab:

$$\begin{aligned} p_{1\cdot} - c_1 \leq x \leq p_{\cdot 1} + c_2, &\text{ or} \\ p_{1\cdot} - p_{\cdot 1} \leq C \equiv c_1 + c_2. & \end{aligned} \tag{IR^{pool}}$$

I consider the possibility of partially revealing strategies in the crisis bargaining game in section XX below.

8.2 Proof of Proposition 1

Proposition 1. *(i) Consider situations in which it is known that there are agreements that (interim) mutually optimistic types prefer to war (IR^{pool} holds). Then the least war-prone crisis has zero risk of conflict. Moreover, in situations where $p_{1\cdot} - p_{\cdot 1} > C$, there must be positive risk of conflict in the least war prone crisis.*

(ii) Consider the case of private capabilities and any military technology such that $p_{ab}(m)$ approaches 1 or zero as the observable balance of power, m, approaches 1 or zero, respectively. Then for m close enough to 1 or zero, the probability of conflict in the least war-prone crisis is zero.

(iii) Consider the case of private analysis described above. As the prior belief that state 1 would win, based on observable factors, approaches 1 or zero, the probability of conflict in the least war-prone crisis approaches zero.

Proof. (i) It is easy to check that if $C \geq p_{1\cdot} - p_{\cdot 1}$, then $w_{ab} = 0$ for all ab and $x \in [p_{1\cdot} - c_1, p_{\cdot 1} + c_2]$ satisfy both the IC and IR^{pool} constraints. To show necessity, suppose $w_{ab} = 0$ for all ab, which makes the incentive constraints

$$\pi_0^1 x_{01} + (1 - \pi_0^1) x_{00} \geq \pi_0^1 x_{11} + (1 - \pi_0^1) x_{10} \tag{IC_{1w}}$$
$$\pi_1^1 x_{11} + (1 - \pi_1^1) x_{10} \geq \pi_1^1 x_{01} + (1 - \pi_1^1) x_{00} \tag{IC_{1s}}$$
$$\pi_0^2 x_{11} + (1 - \pi_0^2) x_{01} \geq \pi_0^2 x_{10} + (1 - \pi_0^2) x_{00} \tag{IC_{2w}}$$
$$\pi_0^2 x_{10} + (1 - \pi_0^2) x_{00} \geq \pi_0^2 x_{11} + (1 - \pi_0^2) x_{01} \tag{IC_{2s}}$$

Let $o_a^1 = \pi_a^1/(1 - \pi_a^1)$ and likewise $o_b^2 = \pi_b^2/(1 - \pi_b^2)$, and note that $o_1^i \leq o_0^i$. (These are the odds that player i thinks $j \neq i$ is strong conditional on seeing a favorable or unfavorable signal themself.) With some rewriting the conditions for players 1 and 2 become, respectively

$$o_1^1(x_{11}-x_{01}) \geq x_{00}-x_{10} \geq o_0^1(x_{11}-x_{01})$$
$$o_0^2(x_{11}-x_{10}) \geq x_{00}-x_{01} \geq o_1^2(x_{11}-x_{10})$$

Thus $o_0^1 \geq o_1^1$ implies $x_{11} \leq x_{01}$ and $x_{00} \leq x_{10}$, and $o_0^2 \geq o_1^2$ implies $x_{00} \geq x_{01}$ and $x_{11} \geq x_{10}$. Putting these together, $x_{10} \geq x_{00} \geq x_{01} \geq x_{11} \geq x_{10}$, so they all must take the same value.

Next, we already know that it cannot be that both types choose the same strategy, since then there is no x such that the strong types don't prefer to fight ex post. If different types choose different strategies, then it must be that there are strategy combinations such that strong types increase their belief that they face a weak type, so that their posterior value for war is at least equal to their ex ante value of $p_{1\cdot}-c_1$ for state 1 and $1-p_{\cdot 1}-c_2$ for state 2. This implies that for both to accept x, it must be that $p_{1\cdot}-c_1 \leq x \leq p_{\cdot 1}+c_2$.

(ii) $\lim_{m\to 1} p_{ab}(m)=1$ implies $\lim_{m\to 1} p_{a\cdot}(m)=1$ and $\lim_{m\to 1} p_{\cdot b}(m)=1$. Thus $\lim_{m\to 1} p_{a\cdot}(m)-p_{\cdot b}(m)=0$ and the condition for the first best is satisfied for any $C>0$ for m close enough to 1. Similarly for $\lim_{m\to 0}$.

(iii) Let $\bar{p}$ be the mean of the prior distribution $f(p)$. Wlog we consider $\bar{p} \geq 1/2$, so player 1 is ex ante observably stronger. It is easy to show that for any prior beliefs f, $p_{10} \geq p_{1\cdot}$ and $p_{1\cdot}=\bar{p}+\sigma^2/\bar{p}$. Note that $\pi_a^1 = 1-p_{a\cdot}$ is player 1's interim belief that player 2 is strong, and $\pi_b^2=p_{\cdot b}$ is player 2's interim belief that 1 is strong.

Consider a mechanism in which player 1 chooses crisis bargaining strategy S regardless of type, while player 2 chooses $W(\neq S)$ or S according to type. Let $w_{11}=1$ and $w_{10}=0$. 2_w wants to play W if

$$1-x_{10} \geq p_{\cdot 0}(1-p_{10}-c_2)+(1-p_{\cdot 0})(1-p_{00}-c_2)$$
$$x_{10} \leq p_{\cdot 0}p_{10}+(1-p_{\cdot 0})p_{00}+c_2$$
$$x_{10} \leq p_{1\cdot}p_{10}+(1-p_{1\cdot})p_{00}+c_2$$

2's choice reveals its type, so ex post 1_s must prefer x_{10} to fighting at $p_{10}-c_1$ if it observes that 2 played W, leading to the condition that

$$p_{10}-c_1 \leq x_{10} \leq p_{1\cdot}p_{10}+(1-p_{1\cdot})p_{00}+c_2$$
$$C \geq (p_{10}-p_{00})(1-p_{1\cdot}).$$

It is easy to check that we can choose $x_{10}=p_{10}-c_1$ and x_{00}, x_{01}, $w_{00}=0$ and w_{01} so that neither type of player 1 wants to deviate to W.

Since $\lim_{\bar{p}\to 1} p_{1\cdot}=1$, we can always implement this mechanism for $C>0$ and large enough $\bar{p}$. The ex ante probability of war is

$$\bar{p}(1-\bar{p})w_{11}+\bar{p}^2 w_{10}+(1-\bar{p})^2 w_{01}=\bar{p}(1-\bar{p})$$

which goes to zero as $\bar{p}$ goes to 1. Since the optimal mechanism by definition cannot do worse than this one, in the optimal mechanism (which can be this one) the claim holds. □

8.3 General results for the optimal mechanism

Fully written-out IC constraints are:

$$u_1^1 = \pi_1^1 \left[(1-w_{11})x_{11} + w_{11}(p_{11}-c_1)\right] + (1-\pi_1^1)\left[(1-w_{10})x_{10} + w_{10}(p_{10}-c_1)\right] \geq$$
$$\hat{u}_1^1 = \pi_1^1 \left[(1-w_{01})x_{01} + w_{01}(p_{11}-c_1)\right] + (1-\pi_1^1)\left[(1-w_{00})x_{00} + w_{00}(p_{10}-c_1)\right] \qquad (IC_{1s})$$
$$u_0^1 = \pi_0^1 \left[(1-w_{01})x_{01} + w_{01}(p_{01}-c_1)\right] + (1-\pi_0^1)\left[(1-w_{00})x_{00} + w_{00}(p_{00}-c_1)\right] \geq$$
$$\hat{u}_0^1 = \pi_0^1 \left[(1-w_{11})x_{11} + w_{11}(p_{01}-c_1)\right] + (1-\pi_0^1)\left[(1-w_{10})x_{10} + w_{10}(p_{00}-c_1)\right] \qquad (IC_{1w})$$

$$u_1^2 = 1-\pi_1^2 \left[(1-w_{11})x_{11} + w_{11}(p_{11}+c_2)\right] - (1-\pi_1^2)\left[(1-w_{01})x_{01} + w_{01}(p_{01}+c_2)\right] \geq$$
$$\hat{u}_1^2 = 1-\pi_1^2 \left[(1-w_{10})x_{10} + w_{10}(p_{11}+c_2)\right] - (1-\pi_1^2)\left[(1-w_{00})x_{00} + w_{00}(p_{01}+c_2)\right] \qquad (IC_{2s})$$
$$u_0^2 = 1-\pi_0^2 \left[(1-w_{10})x_{10} + w_{10}(p_{10}+c_2)\right] - (1-\pi_0^2)\left[(1-w_{00})x_{00} + w_{00}(p_{00}+c_2)\right] \geq$$
$$\hat{u}_0^2 = 1-\pi_0^2 \left[(1-w_{11})x_{11} + w_{11}(p_{10}+c_2)\right] - (1-\pi_0^2)\left[(1-w_{01})x_{01} + w_{01}(p_{00}+c_2)\right] \qquad (IC_{2w})$$

In any separating mechanism, ex post IR implies that all bracketed terms for player 1 are at least great as the expected war payoff, and all bracketed terms for player 2 are no greater than the war term.

We assume in what follows that $p_{1\cdot} - p_{\cdot 1} > C = c_1 + c_2$, so that the first best outcome of zero risk of war is not possible in the optimal mechanism.

Using $\mu_{ab} = 1 - w_{ab}$ as the probability of peace, and $o_a^1 = \pi_a^1/(1-\pi_a^1)$ and $o_b^2 = \pi_b^2/(1-\pi_b^2)$ for the odds that the other state is a strong type conditional on one's own signal, we can more helpfully write these in terms of surpluses over war values as

$$o_1^1\mu_{11}(x_{11}-p_{11}+c_1) + \mu_{10}(x_{10}-p_{10}+c_1) \geq o_1^1\mu_{01}(x_{01}-p_{11}+c_1) + \mu_{00}(x_{00}-p_{10}+c_1) \qquad (IC_{1s})$$
$$o_0^1\mu_{01}(x_{01}-p_{01}+c_1) + \mu_{00}(x_{00}-p_{00}+c_1) \geq o_0^1\mu_{11}(x_{11}-p_{01}+c_1) + \mu_{10}(x_{10}-p_{00}+c_1) \qquad (IC_{1w})$$

$$o_1^2\mu_{11}(p_{11}+c_2-x_{11}) + \mu_{01}(p_{01}+c_2-x_{01}) \geq o_1^2\mu_{10}(p_{11}+c_2-x_{10}) + \mu_{00}(p_{01}+c_2-x_{00}) \qquad (IC_{2s})$$
$$o_0^2\mu_{10}(p_{10}+c_2-x_{10}) + \mu_{00}(p_{00}+c_2-x_{00}) \geq o_0^2\mu_{11}(p_{10}+c_2-x_{11}) + \mu_{01}(p_{00}+c_2-x_{01}) \qquad (IC_{2w})$$

Lemma. $\mu_{00} = 1$ in any optimal mechanism. ($w_{00} = 0$.)

Proof. Suppose $\mu_{00} < 1$. Increasing it gains slack on both weak type IC constraints, with strict gain for at least one, given $x_{00} \in [p_{00}-c_1, p_{00}+c_2]$ by ex post IR. If $x_{00} \in [p_{01}+c_2, p_{10}-c_1]$, then increasing μ_{00} relaxes both strong-type incentive constraints, so it could not be that $\mu_{00} < 1$ in an optimal mechanism. If $x_{00} < p_{01}+c_2$, then IC_{1s} is relaxed by increasing μ_{00}, while IC_{2s} is unaffected since, by ex post IR, if $2s$ deviated to the weak-type strategy when $x_{00} < p_{01} + c_2$, it prefers fighting against $1w$, so μ_{00} has no effect on its payoff. The parallel argument applies for $x_{00} > p_{10} - c_1$. □

Lemma. In any optimal FR mechanism in which there is a positive probability of conflict, $x_{10} = p_{10} - c_1$ and $x_{01} = p_{01} + c_2$.

Proof. [NOT THERE YET] By ex post IR, in an FR mechanism, $x_{10} \geq p_{10} - c_1$. If $x_{10} > p_{10} - c_1$, then IC_{1s} does not bind, since the other terms in parens are all at least zero in an ex post IR mechanism. For $x_{10} > p_{10} - c_1$, reducing x_{10} relaxes IC_{1w} (by making the temptation to mimic $1s$ smaller); has no effect on IC_{2s}, since for $x_{10} \geq p_{10} - c_1$ $2s$ would fight if it deviated and faced $1w$; and relaxes IC_{2w} (since then $2w$ cedes less if it faces $1s$). Thus if a mechanism satisfies the constraints for $x_{10} > p_{10} - c_1$, it also satisfies them for $x_{10} = p_{10} - c_1$.

Lemma. If $w_{11} > 0$ in the optimal mechanism, then at least one of the weak types IC constraints binds.

Proof. Suppose to the contrary that $w_{11} > 0$ and both IC_{1w} and IC_{2w} are slack. Then we can reduce w_{11} a small amount without affecting IC, since (for the strong types) ex post IR implies $x_{11} \in [p_{11} - c_1, p_{11} + c_2]$. (Also, changing w_{11} has no effect on the ex post constraints.) □

Using binding ex post IR constraints, IC_{1w} and IC_{2w} are

$$\begin{aligned}
0 \leq\ & \pi_0^1 w_{11}(x_{11} - p_{01} + c_1) + (1 - \pi_0^1) w_{10}(p_{10} - p_{00}) - \pi_0^1 w_{01} C \\
& - \pi_0^1 (x_{11} - p_{01} - c_2) - (1 - \pi_0^1)(p_{10} - x_{00} - c_1) \\
0 \leq\ & \pi_0^2 w_{11}(p_{10} - x_{11} + c_2) - \pi_0^2 w_{10} C + (1 - \pi_0^2) w_{01}(p_{00} - p_{01}) \\
& - \pi_0^2 (p_{10} - x_{11} - c_1) - (1 - \pi_0^2)(x_{00} - p_{01} - c_2)
\end{aligned}$$

It is convenient to define $o_1 = \pi_0^1/(1 - \pi_0^1)$ and $o_2 = \pi_0^2/(1 - \pi_0^2)$ and rewrite these weak type incentive constraints as

$$\begin{aligned}
0 \leq\ & w_{11}(x_{11} - p_{01} + c_1) + w_{10}(p_{10} - p_{00})/o_1 - w_{01} C \\
& - x_{11} + x_{00}/o_1 + p_{01} - p_{10}/o_1 + c_2 + c_1/o_1
\end{aligned} \tag{2}$$

$$\begin{aligned}
0 \leq\ & w_{11}(p_{10} - x_{11} + c_2) - w_{10} C + w_{01}(p_{00} - p_{01})/o_2 \\
& + x_{11} - x_{00}/o_2 - p_{10} + p_{01}/o_2 + c_1 + c_2/o_2
\end{aligned} \tag{3}$$

The ex post IR constraints can be shown to imply that all the terms in parentheses here are positive.

Lemmas. In an optimal mechanism in which there is positive probability of war:

1. At least one of $ic1$ and $ic2$ must bind (by which I mean, equals zero). Further, if $ic1 > 0$ then $w_{10} = 0$ and if $ic2 > 0$ then $w_{01} = 0$.

   Otherwise, if $w_{11} > 0$ then we could reduce it a little and still satisfy $ic1$ and $ic2$.

Likewise, if $w_{10} > 0$ or $w_{01} > 0$ we can reduce it while maintaining slack on both sides. If $w_{10} > 0$ we could decrease it and still having slack on $ic1$ and $ic2$ while decreasing the probability of war, and similarly for w_{01}.

2. If $ic1 > 0$ and $ic2 = 0$ then $w_{10} = 0$, $x_{11} = ub$, and $x_{00} = lb$.

   Increasing x_{11} tightens $ic1$ and relaxes $ic2$ by the same amount, so that if $ic1$ were slack and $x_{11} < ub$, we could increase x_{11} making for both slack, which would violate (1). Further, since reducing x_{00} tightens $ic1$ while relaxing $ic2$ we must have $x_{00} = lb$ for the same reason.

3. If $ic1 = 0$ and $ic2 > 0$ then $w_{01} = 0$, $x_{11} = lb$, and $x_{00} = ub$.

   If $w_{01} > 0$ we could decrease it and still have slack on $ic2$ while decreasing the probability of war. If $x_{11} > lb$, we could reduce it, creating slack on both sides, which would violate (1). Likewise, if $x_{00} < ub$, then we could increase it creating slack on both sides.

4. Let $u = p_{10} - p_{00}$ and $d = p_{00} - p_{01}$. Let $C = c_1 + c_2$. If $C^2 > ud/o_1o_2$ then either $w_{10} = 0$ or $w_{01} = 0$ or both.

   Let $dx \in \mathbb{R}$ be a small change in the variable x. We have

$$
\begin{aligned}
dic1 &= dw_{10}u/o_1 - dw_{01}C \\
dic2 &= dw_{10}(-C) + dw_{01}d/o_2
\end{aligned}
$$

   With rearranging, we can show a small reductions in dw_{10} and dw_{01} imply $dic1 > 0$ and $dic2 > 0$ when the condition in the lemma obtains, implying that we could gain slack on both constraints while reducing the probability of war. So $w_{10} = 0$ or $w_{01} = 0$ or both.

5. From (2) and (3), if $w_{10} > 0$ and $w_{01} > 0$ then both $ic1$ and $ic2$ bind.

6. If both $ic1$ and $ic2$ bind, then (a) if $o_1 < o_2$, then either $x_{00} = ub$ or $x_{11} = lb$; (b) if $o_1 > o_2$ then $x_{00} = lb$ or $x_{11} = ub$.

   Suppose to the contrary of both (a) and (b) that x_{00} and x_{11} are in their relevant (lb, ub). Increasing x_{11} by the small amount dx_{11} tightens $ic1$ by $1 - w_{11}$ and relaxes $ic2$ by the same amount. Increasing x_{00} by the small amount dx_{00} relaxes $ic1$ by dx_{00}/o_1 and tightens $ic2$ by dx_{00}/o_2. Thus if $o_1 < o_2$, a small increase dx_{00} gains more slack on $ic1$ than it loses on $ic2$, so by compensating with a small enough reduction in x_{11} we could produce slack on both constraints, which contradicts Lemma 1. It follows that if $o_1 < o_2$, then either $x_{00} = ub$ or $x_{11} = ub$, so no further reduction or increase is possible.

   The same kind or argument works for (b) in the other directions.

7. If $w_{10} > 0$ ($w_{01} > 0$) then $ic1$ binds ($ic2$ binds). If not, could reduce w_{10} (w_{01}) and create slack on both sides. If $w_{10} > 0$, then $x_{00} = ub$, since otherwise we could increase x_{00}, shifting $ic2$ and $ic1$ upwards by $1/o_1.$ and $o_0.$, creating a gap, and then reducing w_{10}.

8. The prior belief $f(p)$ having a Beta distribution is sufficient for it to be the case that in an optimal mechanism, either $w_{11} < 1$ and $w_{10} = w_{01} = 0$, or $w_{11} = 1$ and either $w_{10} > 0$ or $w_{01} > 0$ (or both).

If $w_{11} < 1$, suppose we increase w_{11} by a small amount, ϵ_{11}, This increases $ic1$ by $\epsilon_{11} o_{0\cdot} u$ and decreases $ic2$ by $\epsilon_{11} C$. This would allow reducing w_{10} by $\epsilon_{10} : \epsilon_{11} o_{0\cdot} u = \epsilon_{10}(x_{11} - p_{01} + c_1)$, or reducing w_{01} by $\epsilon_{01} : \epsilon_{11} C = \epsilon_{01} d/0_{1\cdot}$.

In the first case, the ex ante probability of war is reduced if $\bar{o}\epsilon_{11} < \bar{o}^2 \epsilon_{10}$, or

$$\frac{o_{0\cdot} u}{x_{11} - p_{01} + c_1} < \bar{o}.$$

The largest possible value for the LHS occurs when $x_{11} = lb = p_{11} - c_1$, implying that $o_{0\cdot} u/d < \bar{o}$ is sufficient. This holds for any Beta f, under which $u = d$. ($o_{0\cdot} < \bar{o}$ always.) So we could reduce the probability of war further and so have a contradiction.

In the second case (w_{01}), increasing w_{11} by ϵ_{11} allows reducing w_{01} by ϵ_{01} such that $\epsilon_{01} d/o_{1\cdot} = \epsilon_{11}(p_{10} - x_{11} + c_2)$, and this will reduce the probability of war if $\epsilon_{01} > \bar{o}\epsilon_{11}$, which is

$$\frac{p_{10} - x_{11} + c_2}{d/o_{1\cdot}} > \bar{o}.$$

The LHS has its minimum value at $x_{11} = p_{11} + c_2$, implying the sufficient condition $o_{1\cdot} u/d > \bar{o}$, which is always true with the Beta f prior.

9 Symmetric players (equal observable balance and stakes)

With symmetric players, $c = c_1 = c_2$, $p_{00} = 1/2$, $p_{10} = 1 - p_{01}$, $w_{10} = w_{01}$, $x_{11} = x_{00} = 1/2$.

9.1 Private capabilities

$o = o_1 = o_2$.

The incentive constraints are now the same for both states, and can be rewritten as

$$w_{11}(p_{10} - 1/2 + c) + w_{10}(p_{10}/o - 1/2o - 2c) \geq (p_{10} - 1/2 - c)(1 + 1/o)$$

With $\bar{p} = o/(o+1)$ the prior chance that a state is the strong type, the problem is minimize $\bar{p}^2 w_{11} + 2\bar{p}(1-\bar{p}) w_{10}$ subject to this constraint.

Clearly, $c \geq p_{10} - 1/2$ allows the first-best of $w_{11} = w_{10} = w_{01} = 0$. This is the same as the general condition in Proposition 1 for this case.

For $c < p_{10} - 1/2$, solving for w_{10} in the constraint (since it must bind) and substituting into the objective function, we find that for any o, it is best to increase w_{11} before w_{10}. This result leads to

$$w_{11}^* = \frac{o+1}{o} \frac{p_{10} - 1/2 - c}{p_{10} - 1/2 + c} \text{ and } w_{10}^* = 0 \text{ if } C \in \left[\frac{p_{10} - 1/2}{1 + 2o}, p_{10} - 1/2\right]$$

and

$$w_{11}^* = 1 \text{ and } w_{10}^* = \frac{p_{10} - 1/2 - c(1+2o)}{p_{10} - 1/2 - 2co} \text{ if } c < \frac{p_{10} - 1/2}{1+2o}.$$

9.2 Private analysis

With an equal observable balance of power, $\bar{p} = 1/2$, and $p_{0\cdot} = 1 - p_{1\cdot}$. Recalling that $o_1 = (1 - p_{0\cdot})/p_{0\cdot}$ and $o_2 = p_{1\cdot}/(1 - p_{1\cdot})$, this implies that $o_1 = o_2$, exactly as in the private capabilities case. The analysis is then the same as above, except that $\bar{p}$ is constrained to be 1/2, which is a subcase for that proof.

However, whereas with private capabilities, $o = \bar{p}/(1 - \bar{p}$ is the prior odds that a state gets the favorable signal about its capabilities, with private analysis, $o = p_{1\cdot}/(1 - p_{1\cdot})$ is state 1's interim belief if it sees a favorable signal, which is the same as state 2's belief about state 1's odds if it sees the unfavorable signal.

10 Private capabilities

This section works through the complete solution for conflict and signal technologies with $o = o_1 = o_2$ and $d > u$, which is the case for the ratio-form when $m > 1/2$. (Recall $u \equiv p_{10} - p_{00}$ and $d \equiv p_{00} - p_{01}$.) The incentive constraints for full separation become

$$\begin{aligned} 0 \le\ & w_{11}(x_{11} - p_{01} + c_1) + w_{10}u/o - w_{01}C \\ & - x_{11} + x_{00}/o + p_{01} - p_{10}/o + c_2 + c_1/o \end{aligned} \tag{4}$$

$$\begin{aligned} 0 \le\ & w_{11}(p_{10} - x_{11} + c_2) - w_{10}C + w_{01}d/o_2 \\ & + x_{11} - x_{00}/o - p_{10} + p_{01}/o + c_1 + c_2/o \end{aligned} \tag{5}$$

10.1 $w_{11} = w_{10} = w_{01} = 0$

[...]

10.2 $w_{11} > 0, w_{10} = w_{01} = 0$

[...]

10.2.1 $ic1 > ic2 = 0$

By Lemma X, $x_{11} = ub$ and $x_{00} = lb$. Substituting into $ic2$ and solving for w_{11} leads to

$$w_{11}^* = 1 - \frac{C(1+o) - d}{uo},$$

which for $w_{11}^* \in [0,1]$ requires $C \in [d/(1+o), (d+uo)/(1+o)]$.

At w_{11}^* $ic1$ must be positive, which holds if

$$C^2(1+o) + Co(d-u) - (d^2 - u^2) < 0.$$

Let the roots of this quadratic be $\underline{C}^{1+}$ and $\bar{C}^{1+}$ (they always exist given $d > u$). The smaller root is negative, so this condition becomes $C \in \left[0, \bar{C}^{1+}\right]$.[29]

With algebra it can be shown that $\bar{C}^{1+} < (d+uo)/(1+o)$, and it is greater than $d/(1+o)$ if $o > u/(d-u)$. So this case occurs for

$$C \in \left[\frac{d}{1+o}, \bar{C}^{1+}\right] \text{ if } o > \frac{u}{d-u}$$

and not otherwise.

10.2.2 $ic2 > ic1 = 0$

By Lemma 3, $x_{11} = lb$ and $x_{00} = ub$. Substituting into $ic1$ and solving for w_{11} leads to

$$w_{11}^* = 1 - \frac{C(1+o) - u}{do}$$

which for $w_{11}^* \in [0,1]$ requires $C \in [u/(1+o), (u+do)/(1+o)]$.

At w_{11}^* $ic2$ must be positive, which holds if

$$C^2(1+o) - Co(d-u) + d^2 - u^2 < 0$$

Let $\underline{C}^{2+}$ and $\bar{C}^{2+}$ be the roots of this quadratic,[30] which are real and both positive when

$$o > \hat{o} \equiv \frac{2\left(d + u + \sqrt{2d(d+u)}\right)}{d-u}.$$

If this condition fails, then this case cannot exist because there is no C such that $ic2 > ic1 = 0$.

[29] These are $C^{1+} = \left(-o(d-u) \pm \sqrt{o^2(d-u)^2 + 4(1+o)(d^2-u^2)}\right)/2(1+o)$.

[30] These are $C^{2+} = \left(o(d-u) \pm \sqrt{o^2(d-u)^2 - 4(1+o)(d^2-u^2)}\right)/2(1+o)$.

With algebra (actually, Mathematica) it can be shown that $[\underline{C}^{2+}, \bar{C}^{2+}] \subset [u/(1+o), (u+do)/(1+o)]$, so for this case to obtain we need $o > \hat{o}$ and $C \in [\underline{C}^{2+}, \bar{C}^{2+}]$. (I note in passing that $\hat{o} > 1$, and so this case does not exist for perhaps the most plausible assumption of $o = 1$.)

Moreover, when $o > \hat{o}$, $\bar{C}^{1+} < \underline{C}^{2+}$, which means that the conditions for $ic2 > ic1 = 0$ and $ic1 > ic2 = 0$ cannot hold at the same time.

10.2.3 $ic1 = ic2 = 0$

With $o = o_1 = o_2$, $x_{11} - x_{00}/o$ adds to $ic2$ and subtracts from $ic1$. Substituting x_{00}/o from $ic2$ into $ic1$ thus eliminates the x terms, and solving yields

$$w_{11}^* = (1 + 1/o)\,\frac{u + d - C}{u + d + C}.$$

This is in $[0, 1]$ when $C \in [(u + d)/(1 + 2o), u + d]$.

Writing $ic1$ in terms of w_{11}^*,

$$0 = w_{11}^*(x_{11} - p_{01} + c_1) - x_{11} + x_{00}/o + p_{01} - p_{10}/o + c_2 + c_1/o.$$

We need to identify conditions such that there are values of x_{00} and x_{11} within their bounds such that this equality is satisfied. Since the RHS is increasing in x_{00} and decreasing in x_{11}, we need RHS > 0 when $x_{00} = ub$ and $x_{11} = lb$, and RHS < 0 when $x_{00} = lb$ and $x_{11} = ub$. (It doesn't matter if we use $ic1$ or $ic2$ since they are the same at w_{11}^*.) With the help of Mathematica, we find that

$$\begin{array}{lll} C \in [(u+d)/(1+2o), u+d] & \text{if} & o < u/(d-u) \\ C \in [\bar{C}^{1+}, u+d] & \text{if} & o \in [u/(d-u), \hat{o}] \\ C \in [\bar{C}^{1+}, \underline{C}^{2+}] \text{ or } C \in [\bar{C}^{2+}, u+d]\} & \text{if} & o > \hat{o}. \end{array}$$

10.3 $w_{11} = 1$ and $w_{10} > 0, w_{01} = 0$ or $w_{10} = 0, w_{01} > 0$

If $w_{11} = 1$ and $ic1 > 0$ and $ic2 = 0$, then from Lemmas, $w_{10} = 0$ and $x_{00} = lb$. Solving $ic2 = 0$ for w_{01} yields

$$w_{01}^* = 1 - C(1 + o_2)/d,$$

which implies that we need $C < d/(1 + o_2)$. With the condition for $ic1 > 0$, we have

$$C \in \left[\sqrt{\frac{du}{o_1(1 + o_2)}}, \frac{d}{1 + o_2}\right].$$

For private capabilities with $o = o_1 = o_2$, this interval exists only if $o > u/(d - u)$.

If $w_{11} = 1$ and $ic1 = 0$ and $ic2 > 0$, then from Lemmas, $w_{01} = 0$ and $x_{00} = ub$. Solving $ic1 = 0$

for w_{10} yields

$$w_{10}^* = 1 - C(1+o_1)/u,$$

which implies that we need $C < u/(1+o_1)$. With the condition for $ic2 > 0$, we have

$$C \in \left[\sqrt{\frac{du}{o_2(1+o_1)}}, \frac{u}{1+o_1}\right].$$

In the private capabilities case with $o = o_1 = o_2$ and $d > u$, this interval does not exist, so this case does not exist, and we have

$$w_{10}^* = 0 \text{ and } w_{01}^* = 1 - C(1+o)/d \text{ for } C \in \left[\sqrt{\frac{du}{o(1+o)}}, \frac{d}{1+o}\right] \text{ if } o > \frac{u}{d-u}.$$

There is also a case where $ic1 = ic2 = 0$ and $w_{10} = 0$ and $w_{01} > 0$. Solving under these assumptions for w_{01} gives

$$w_{01}^* = \frac{C(1+2o) - u - d}{Co - d},$$

which, given $d > u$, is in $[0,1]$ when $C \in [u/(1+o), (d+u)/(1+2o)]$. Solving for x_{00} at this w_{01}^* and finding conditions for $x_{00} \in [lb, ub]$, we find that this case exists for

$$C \in \left[\sqrt{\left(\frac{d}{2o^2}\right)^2 + \frac{du}{o^2}} - \frac{d}{2o^2}, \frac{d+u}{1+2o}\right] \text{ if } o < u/(d-u)$$
$$C \in \left[\sqrt{\left(\frac{d}{2o^2}\right)^2 + \frac{du}{o^2}} - \frac{d}{2o^2}, \sqrt{\frac{du}{o(1+o)}}\right] \text{ if } o > u/(d-u)$$

10.4 $w_{11} = 1, w_{10} > 0, w_{01} > 0$

From Lemma X, both ic constraints must bind.

$$0 = ic1 = w_{10}u/o - w_{01}C + x_{00}/o - p_{10}/o + C + c_1/o$$
$$0 = ic2 = w_{01}d/o - w_{10}C - x_{00}/o + p_{01}/o + C + c_2/o.$$

Eliminating x_{00} and solving for the w's gives

$$w_{10} = 1 - \frac{dC}{du - C^2o^2} \text{ and } w_{01} = 1 - \frac{C^2o}{du - C^2o^2}.$$

Conditions for both of these to be in $[0, 1]$ reduce to

$$C \in \left[0, \sqrt{\left(\frac{d}{2o^2}\right)^2 + \frac{du}{o^2}} - \frac{d}{2o^2}\right],$$

and $x_{00} = p_{00} + c_2$.

10.4.1 $w_{11} = 1$ and w_{10} or $w_{01} = 1$

Not possible. By Lemmas, both ic's bind if $w_{10} > 0$ and $w_{01} > 0$, which leads to case above with exact solutions for these two w's for C less than a threshold.

11 Private analysis

$\bar{p} \geq 1/2$, $u = d$ and $o_2 > o_1$ (i.e., state 1's probability of being the strong type is greater than 2's probability conditional both seeing unfavorable signals.)

11.1 $w_{11} = 1, w_{10} > 0, w_{01} > 0$

From Lemma X, both ic constraints must bind.

$$\begin{aligned}
0 = ic1 &= w_{10}u/o_1 - w_{01}C + x_{00}/o_1 - p_{10}/o_1 + C + c_1/o_1 \\
0 = ic2 &= w_{01}d/o_2 - w_{10}C - x_{00}/o_2 + p_{01}/o_2 + C + c_2/o_2.
\end{aligned}$$

From Lemma X, $o_2 > o_1$ implies $x_{00} = ub$. Substituting for $x_{00} = ub$ and solving for the w's yields

$$w_{10} = 1 - \frac{uC}{u^2 - C^2 o_1 o_2} \text{ and } w_{01} = 1 - \frac{C^2 o_2}{u^2 - C^2 o_1 o_2}.$$

For both to be in $[0, 1]$ we need[31]

$$C \leq u \min \left\{ \sqrt{\frac{1}{o_2(1 + o_1)}}, \sqrt{\left(\frac{1}{2o_1 o_2}\right)^2 + \frac{1}{o_1 o_2}} - \frac{1}{2o_1 o_2} \right\}.$$

The first term is $\gtrless$ than the second as $o_1 + 1 \gtrless o_2$.

Define u times the first term in the min as $\hat{C}_1$ and u times the second as $\hat{C}_2$. Summing up,

[31] Note that $u^2 > C^2 o_1 o_2$ is implied by this condition.

Case 1a: $w_{11} = 1, w_{10} > 0, w_{01} > 0$

$$\begin{aligned}
w &= \left(1, 1 - uC/(u^2 - C^2 o_1 o_2), 1 - C^2 o_2/(u^2 - C^2 o_1 o_2)\right) \\
x &= (ub, \cdot) \text{ for} \\
C &< \hat{C}_2 \text{ if } o_2 < o_1 + 1 \\
C &< \hat{C}_1 \text{ if } o_2 > o_1 + 1.
\end{aligned}$$

11.1.1 $w_{11} = 1$ and $w_{10} > 0, w_{01} = 0$ or $w_{10} = 0, w_{01} > 0$

If $w_{11} = 1$ and $ic1 > 0$ and $ic2 = 0$, then from Lemmas, $w_{10} = 0$ and $x_{00} = lb$. Solving $ic2 = 0$ for w_{01} yields

$$w_{01}^* = 1 - Co_2/u,$$

which implies that we need $C < u/o_2$. With the condition for $ic1 > 0$, we get

$$C \in \left[\hat{C}_2, u/o_2\right],$$

which exists (as an interval) only when $o_2 < o_1 + 1$.

If $w_{11} = 1$ and $ic1 = 0$ and $ic2 > 0$, then from Lemmas, $w_{01} = 0$ and $x_{00} = ub$. Solving $ic1 = 0$ for w_{10} yields

$$w_{10}^* = 1 - C(1 + o_1)/u,$$

which implies that we need $C < u/(1 + o_1)$. With the condition for $ic2 > 0$, we have

$$C \in u\left[\sqrt{\frac{1}{o_2(1 + o_1)}}, \frac{1}{1 + o_1}\right],$$

which exists as an interval when $o_2 > o_1 + 1$.

Summing up, we have

Case 1b: $w_{11} = 1, w_{10} = 0, w_{01} > 0$

$$\begin{aligned}
w &= (1, 0, 1 - Co_2/u) \\
x &= (ub, \cdot) \text{ if } o_2 < o_1 + 1 \text{ and} \\
C &\in [\hat{C}_2, u/o_2].
\end{aligned}$$

Case 1c: $w_{11} = 1, w_{10} > 0, w_{01} = 0$

$$\begin{aligned}
w &= (1, 1 - C(1 + o_1)/u, 0) \\
x &= (ub, \cdot) \text{ if } o_2 > o_1 + 1 \text{ and} \\
C &\in [\hat{C}_1, u/(1 + o_1)].
\end{aligned}$$

11.2 $w_{11} > 0, w_{10} = w_{01} = 0$

11.2.1 $ic1 > ic2 = 0$

By Lemma X, $x_{11} = ub$ and $x_{00} = lb$. Substituting into $ic2$ and solving for w_{11} leads to

$$w_{11}^* = 1 - \frac{u - C(1+o_2)}{uo_2},$$

at which value $ic1$ must be positive, which holds if

$$0 > C^2(1 + 1/o_2) + u^2(1/o_1 - 1/o_2)$$

which is not possible for any $C > 0$. So this case cannot happen.

11.2.2 $ic2 > ic1 = 0$

By Lemma 3, $x_{11} = lb$ and $x_{00} = ub$. Substituting into $ic1$ and solving for w_{11} leads to

$$w_{11}^* = 1 - \frac{C(1+o_1) - u}{uo_1}$$

at which value $ic2$ must be positive, which holds if

$$\begin{aligned} C^2(1 + 1/o_1) + u^2(1/o_2 - 1/o_1) < 0 \\ C < C^{yn} \equiv u\sqrt{\frac{o_2 - o_1}{o_2(o_1+1)}}. \end{aligned}$$

$w_{11}^* \in [0, 1]$ requires $C \in [u/(1 + o_1), u]$, so this case exists for C in the intersection of these two intervals, if this is not empty, which is the case when $o_2 > 1 + o_1$ and for

$$C \in \left[\frac{u}{1+o_1}, C^{yn}\right].$$

Summing up:

Case 2: $w_{11} < 1, w_{10} = 0, w_{01} = 0$

$$\begin{aligned} w &= \left(1 - \frac{C(1+o_1) - u}{uo_1}, 0, 0\right) \\ x &= (ub, lb) \\ C &\in (u/(1 + o_1), C^{yn}) \text{ and } o_2 > o_1 + 1. \end{aligned}$$

11.2.3 $ic1 = ic2 = 0, x_{00} = p_{00} + c_2$

If $o_1 < o_2$ then from Lemma 6 (both ic's bind), either $x_{00} = ub$ or $x_{11} = ub$, so that we have in either case two equations and two unknowns (w_{11} and the x). We solve for w_{11} and then use a constraint to derive conditions on w_{11} such that the relevant $x \in [lb, ub]$. This section solves for this case when $x_{00} = ub$.

Solving for w_{11} and x_{11} leads to

$$w_{11}^* = \frac{u(2 + 1/o_1 + 1/o_2) - C(1 + 1/o_1)}{2u + C},$$

which is in $[0, 1]$ when $C \in [\underline{C}, \bar{C}]$, where we define

$$\underline{C} = u\frac{o_1 + o_2}{o_2(1 + 2o_1)} \text{ and } \bar{C} = u\frac{o_1 + o_2 + 2o_1o_2}{o_2(1 + o_1)}.$$

$x_{11} > p_{11} - c_1$ when $C > C^{yn}$ and $x_{11} < p_{11} + c_2$ when

$$C^2o_1o_2 - Cu(o_1 + o_2) + u^2(o_2 - o_1) > 0.$$

This has roots (when they exist)

$$\frac{u}{2o_1o_2}\left(o_1 + o_2 \pm \sqrt{(o_1 + o_2)^2 - 4o_1o_2(o_2 - o_1)}\right).$$

Call these C^l and C^h. They exist, and both are positive, when $(o_1 + o_2)^2 > 4o_1o_2(o_2 - o_1)$, which after some work we find to be the case when either $o_1 < 1/4$ or

$$o_2 < \hat{o}_2 \equiv \frac{o_1}{4o_1 - 1}\left[2o_1 + 1 + \sqrt{o_1(2 + o_1)}\right].$$

When this conditions fails (i.e., $o_1 > 1/4$ and $o_2 > \hat{o}_2$ so there are no real roots to the quadratic), $x_{11} < p_{11} + c_2$ for any $C > 0$. Thus when $o_2 > \hat{o}_2$ and $o_1 > 1/4$, the conditions for this case are that $C \in [\max\{C^{yn}, \underline{C}\}, \bar{C}]$. From algebra, when $o_2 > \hat{o}_2$, the first term in the max is greater than the second, so this becomes $C \in [C^{yn}, \bar{C}]$.

If $o_2 < \hat{o}_2$ or $o_1 < 1/4$, then the roots are real and both positive, and the condition for $x_{11} < p_{11} + c_2$ is that either $C < C^l$ or $C > C^h$. So this case requires either

$$C \in \left[\max\{C^{yn}, \underline{C}\}, \min\{C^l, \bar{C}\}\right] \text{ or } C \in \left[\max\{C^{yn}, \underline{C}, C^h\}, \bar{C}\right].$$

Observe that if $C^l > \bar{C}$, then $C^h > \bar{C}$ also, so the second interval does not exist and the first interval becomes $C \in [\max\{C^{yn}, \underline{C}\}, \bar{C}]$. From algebra we find that $C^l > \bar{C}$ when $o_1 < \sqrt{2} - 1$.

Given this, more work shows that $C^{yn} \gtrless \underline{C}$ as $o_2 \gtrless o_1 + 1$, so we have

$$
\begin{aligned}
&C \in \left[C^{yn}, \bar{C}\right] \text{ if } o_1 < \sqrt{2} - 1 \text{ and } o_2 > o_1 + 1 \\
&C \in \left[\underline{C}, \bar{C}\right] \text{ if } o_1 < \sqrt{2} - 1 \text{ and } o_2 < o_1 + 1.
\end{aligned}
$$

For $o_1 > \sqrt{2} - 1$, the first interval is $\left[\max\{C^{yn}, \underline{C}\}, C^l\right]$, and the lower bound is again determined by $o_2 \gtrless o_1 + 1$, leading to

$$
\begin{aligned}
&C \in \left[C^{yn}, C^l\right] \text{ if } o_1 > \sqrt{2} - 1 \text{ and } o_2 > o_1 + 1 \\
&C \in \left[\underline{C}, C^l\right] \text{ if } o_1 > \sqrt{2} - 1 \text{ and } o_2 < o_1 + 1.
\end{aligned}
$$

Now consider the higher interval, which exists only when $o_1 > \sqrt{2} - 1$. From lots of algebra or Mathematica, one finds that $(o_2 < \hat{o}_2 | o_1 < 1/4)$ implies that C^h is greater than than both C^{yn} and $\underline{C}$, and also $C^h < \bar{C}$. Thus this case can obtain for

$$
C \in \left[C^h, \bar{C}\right] \text{ if } o_1 > \sqrt{2} - 1 \text{ and } o_2 < \hat{o}_2.
$$

Summing up, we have

Case 3a: $w_{11} < 1, w_{10} = 0, w_{01} = 0$

$$
\begin{aligned}
&w = \left(\frac{u(2 + 1/o_1 + 1/o_2) - C(1 + 1/o_1)}{2u + C}, 0, 0\right) \\
&x = (ub, x^*_{11}) \text{ if } (o_2 < \hat{o}_2 | o_1 < 1/4) \text{ and} \\
&C \in \left[C^{yn}, \bar{C}\right] \text{ if } o_1 < \sqrt{2} - 1 \text{ and } o_2 > o_1 + 1 \\
&C \in \left[\underline{C}, \bar{C}\right] \text{ if } o_1 < \sqrt{2} - 1 \text{ and } o_2 < o_1 + 1 \\
&C \in \left[C^{yn}, C^l\right] \cup \left[C^h, \bar{C}\right] \text{ if } o_1 > \sqrt{2} - 1 \text{ and } o_2 > o_1 + 1 \\
&C \in \left[\underline{C}, C^l\right] \cup \left[C^h, \bar{C}\right] \text{ if } o_1 > \sqrt{2} - 1 \text{ and } o_2 < o_1 + 1 \\
&\quad \text{or if } o_2 > \hat{o}_2 \text{ and } o_1 > 1/4 \text{ and } C \in [C^{yn}, \bar{C}].
\end{aligned}
$$

11.2.4 $ic1 = ic2 = 0, x_{11} = p_{11} + c_2$

For $x_{11} = ub = p_{11} + c_2$, eliminating x_{00} leads to

$$
w^*_{11} = \frac{2u + uo_1 + uo_2 - C(1 + o_2)}{uo_2 + uo_1 + Co_1},
$$

which is in $[0,1]$ when $C \in [\underline{C}^\circ, \bar{C}^\circ]$, defining

$$\underline{C}^\circ = \frac{2u}{1+o_1+o_2} \text{ and } \bar{C}^\circ = \frac{2u+uo_1+uo_2}{1+o_2}.$$

$x_{00} > p_{00} - c_1$ when $C > C^{yn}$. $x_{00} < p_{00} + c_2$ when $C \in [C^l, C^h]$ when these roots exist, and when they do not, this condition cannot be satisfied. Therefore this case does not exist for $o_2 > \hat{o}_2$ and $o_1 > 1/4$.

So, when $(o_2 < \hat{o}_2 | o_1 < 1/4)$, the case requires $C \in [\max\{\underline{C}^\circ, C^{yn}, C^l\}, \min\{C^h, \bar{C}^\circ\}]$.

More algebra: $(o_2 < \hat{o}_2 | o_1 < 1/4)$ implies that $C^{yn} < C^l$; $C^l \gtrless \underline{C}^\circ$ as $o_2 \gtrless o_1 + 1$; and $C^h \gtrless \bar{C}^\circ$ as $o_2 \gtrless o_1 + 1$. So when $(o_2 < \hat{o}_2 | o_1 < 1/4)$ this case requires

$$\begin{aligned} &C \in [C^l, \bar{C}^\circ] \text{ if } o_2 > o_1 + 1 \\ &C \in [\underline{C}^\circ, C^h] \text{ if } o_2 < o_1 + 1. \end{aligned}$$

Summing up, we have

Case 3b: $w_{11} < 1, w_{10} = 0, w_{01} = 0$

$$\begin{aligned} w &= \left(\frac{2u + uo_1 + uo_2 - C(1+o_2)}{uo_2 + uo_1 + Co_1}, 0, 0 \right) \\ x &= (x^*_{00}, ub) \text{ if } (o_2 < \hat{o}_2 | o_1 < 1/4) \text{ and} \\ C &\in [C^l, \bar{C}^\circ] \text{ if } o_2 > o_1 + 1 \\ C &\in [\underline{C}^\circ, C^h] \text{ if } o_2 < o_1 + 1. \end{aligned}$$

11.2.5 $w_{11} = w_{10} = w_{01} = 0$

This is the first-best case. The IC constraints become

$$\begin{aligned} 0 &\leq -x_{11} + x_{00}/o_1 + p_{01} - p_{10}/o_1 + c_2 + c_1/o_1 \\ 0 &\leq x_{11} - x_{00}/o_2 - p_{10} + p_{01}/o_2 + c_1 + c_2/o_2, \text{ or} \end{aligned}$$

$o_2 > o_1$ implies that unless either $x_{00} = ub$ or $x_{11} = ub$, we could increase both slightly and gain slack on both constraints. So to find the lowest costs such that both constraints hold, either $x_{00} = ub$ or $x_{11} = ub$. Substituting in turn, and asking about conditions under which the other x can be within bounds and still have the constraints satisfied, I find that the first-best obtains when

$$C \geq \left\{ u(1 + 1/o_2), u\frac{o_1 + o_2 + 2o_1o_2}{o_2(1+o_1)} \right\}.$$

The first term is greater than the second with $o_1 > 1$ and the second is greater than the first when $o_1 < 1$.

Summing up,

Case 4a: $w_{11} = 0, w_{10} = 0, w_{01} = 0$

$$\begin{aligned} w &= (0,0,0) \\ x &= () \text{ if} \\ C &\geq \bar{C} \text{ if } o_1 < 1 \\ C &\geq u(1 + 1/o_2) \text{ if } o_1 > 1. \end{aligned}$$

The above describes the optimal mechanism given that the states choose different crisis bargaining strategies. To complete the analysis we need to ask if there is a mechanism in which state 1, state 2, or both choose the same strategies that satisfies the constraints and produces a lower ex ante probability of conflict.[32]

Case 4a: Peace for sure when both types of both states choose the same strategy.

If both types choose the same strategy – or equivalently if the direct mechanism prescribes the same x and w regardless of what is reported – ex ante payoffs are $(1-w)x + w(p_{a\cdot} - c_1)$ for state 1 and $(1-w)(1-x) + w(1-p_{\cdot b} - c_2)$ for state 2. Since the mechanism (or crisis behavior) reveals no information about type, the ex post constraints are simply

$$\begin{aligned} x &\geq p_{a\cdot} - c_1 \\ 1 - x &\geq 1 - p_{\cdot b} - c_2, \end{aligned}$$

which implies that these are satisfied iff $x \in [p_{1\cdot} - c_1, p_{\cdot 1} + c_2]$.

In the direct mechanism interpretation, IC is satisfied since payoffs do not depend on type. In the crisis bargaining game interpretation, deviation to a different strategy is suboptimal if, say, it is expected to generate $w = 1$.

When $C \geq p_{1\cdot} - p_{\cdot 1}$ the optimal mechanism has $w_{11} = w_{10} = w_{01} = 0$. It is perfectly possible for parameters to be such that this condition is met but $C < \max\{u(1 + 1/o_2), u(2 + o_{0\cdot} + 1/o_{1\cdot})/(1 + o_{0\cdot})\}$, so that there is a positive probability of conflict in the best mechanism that has different types choosing different strategies.

Case 4b: State 1 "pools," choosing the same strategy regardless of type, and state 2 chooses different strategies depending on type.

[32]These distributions on outcomes can be implemented with a direct revelation mechanism – the states truthfully report their types to a "mediator" – provided that we make the assumptions listed in footnote 14. As Shimer and Werning (2019) note, "modeling ex post participation constraints requires us to be explicit about what a buyer learns from the mechanism." In our case, first, the mediator/mechanism can hide information, meaning that she does not have to implement outcomes that respect the ex post participation constraints given her knowledge of the players' types. Second, the players do not observe each other's truthful reports; otherwise the ex post constraints would bind. And third, in the mediator interpretation, the mediator must be able to compel the players to fight rather than continue bargaining outside the mechanism (the equivalent of this condition is actually required in all standard mechanism design analyses of bargaining).

There is clearly no gain to having $w_{10} > 0$ as this lowers state 1's payoffs while hurting IC for state 2. So $w_{10} = 0$.

Start by assuming a case in which $w_{11} = 1$. Ex post IR_{1s} requires that $x_{10} \geq p_{10} - c_1$, and IC_{2s} requires that $1 - x_{10} \leq 1 - p_{0\cdot} - c_2$ or $x_{10} \geq p_{0\cdot} + c_2$. IC_{2w} requires $1 - x_{10} \geq 1 - p_{1\cdot} - c_2$, or $x_{10} \leq p_{1\cdot} + c_2$. So we want to choose the largest x_{10} that satisfies IC_{2w}, leading to the requirements that

$$\begin{aligned} p_{1\cdot} + c_2 &\geq \max\{p_{10} - c_1, p_{0\cdot} + c_2\} \text{ or} \\ C &\geq p_{10} - p_{1\cdot}. \end{aligned}$$

Next, we ask when we can do better by having $w_{11} < 1$.

Ex post IR requires that if the negotiated settlement for SW occurs, $x_{10} \in [p_{10} - c_1, p_{1\cdot} + c_2]$ and if the negotiated settlement for SS occurs, $x_{11} \in [p_{11} - c_1, p_{0\cdot} + c_2]$.[33] These yield necessary conditions (for the intervals to exist)

$$C > \max\{p_{10} - p_{1\cdot}, p_{11} - p_{0\cdot}\}.$$

State 2's incentive constraints are

$$ic2w \equiv (1-w_{11})(1-x_{11}) + w_{11}(1-p_{\cdot 0}-c_2) \leq 1-x_{10} \leq ic2s \equiv (1-w_{11})(1-x_{11}) + w_{11}(1-p_{\cdot 1}-c_2)$$

or (using $p_{\cdot 0} = p_{1\cdot}$ and $p_{\cdot 1} = p_{0\cdot}$)

$$x_{11}(1 - w_{11}) + w_{11}(p_{0\cdot} + c_2) \leq x_{10} \leq x_{11}(1 - w_{11}) + w_{11}(p_{1\cdot} + c_2).$$

By IR_{2s}, $x_{11} \leq p_{0\cdot} + c_2 < p_{1\cdot} + c_2$, so both LHS and RHS decrease as w_{11} decreases, which implies that we do best to make x_{10} as small as possible. It binds for $1s$ at $x_{10} = p_{10} - c_1$, leading to

$$x_{11}(1 - w_{11}) + w_{11}(p_{0\cdot} + c_2) \leq p_{10} - c_1 \leq x_{11}(1 - w_{11}) + w_{11}(p_{1\cdot} + c_2).$$

Suppose that $x_{11} < p_{0\cdot} + c_2$ and $w_{11} > 0$. Then we can slightly increase x_{11} while keeping IR_{2s} satisfied, while reducing w_{11} and restoring or maintaining IC on both sides. So IR_{2s} binds and $x_{11} = p_{0\cdot} + c_2$. This leads to the result that this mechanism is feasible with

$$w_{11}^* = \frac{p_{10} - p_{0\cdot} - C}{p_{1\cdot} - p_{0\cdot}}$$

which is between zero and one when $C \in (p_{10} - p_{1\cdot}, p_{10} - p_{0\cdot})$. Combining with the condition above, this case requires

$$C \in (\max\{p_{10} - p_{1\cdot}, p_{11} - p_{0\cdot}\}, p_{10} - p_{0\cdot}).$$

[33] Using $p_{\cdot 0} = p_{1\cdot}$ and $p_{\cdot 1} = p_{0\cdot}$, and the fact that these bind on $2s$ before $2w$.

For $f \sim$ Beta, it is easy to show that the second term in the max is always strictly greater; thus $C \in (p_{11} - p_{0\cdot}, p_{10} - p_{0\cdot})$.

Since $p_{10} > p_{1\cdot}$, case 4a is feasible for the upper part of this range, and it has a lower probability war (zero). So, putting it together with the $w_{11} = 1$ condition above, case 4b is $w^*_{10} = 0$ and

$$\begin{aligned} w^*_{11} &= \frac{p_{10} - p_{0\cdot} - C}{p_{1\cdot} - p_{0\cdot}} \text{ when } C \in (p_{11} - p_{0\cdot}, p_{1\cdot} - p_{0\cdot}) \text{ and} \\ w^*_{11} &= 1 \text{ when } C \in [p_{10} - p_{1\cdot}, p_{11} - p_{0\cdot}] \end{aligned}$$

$x_{10} = p_{1\cdot} + c_2$ when $w^*_{11} = 1$, and $x^*_{10} = p_{10} - c_1$ and $x_{11} = p_{0\cdot} + c_2$ when $w^*_{11} < 1$. (x_{11} is immaterial when SS means war for sure.)

There is no simple analytic expression for when the ex ante probability of war in Case 4b is lower than for the best alternative "separating" mechanism (cases 1-3).

Case 4c: State 2 "pools" on W and state 1 plays according to type. This case can be supported as an equilibrium for some parameters, but is never optimal. Case 4b exists in same range and has lower ex ante probability of war, as shown next.

By a similar logic in 4b, $w_{00} = 0$. IR requires that $x_{00} \in [p_{0\cdot} - c_1, p_{01} + c_2]$ and $x_{10} \in [p_{1\cdot} - c_1, p_{11} + c_2]$, implying conditions

$$C > \max\{p_{0\cdot} - p_{01}, p_{1\cdot} - p_{11}\}.$$

State 1's incentive constraints are

$$ic1w \equiv (1 - w_{10})x_{10} + w_{10}(p_{0\cdot} - c_1) \leq x_{00} \leq ic1s \equiv (1 - w_{10})x_{10} + w_{10}(p_{1\cdot} - c_1)$$

Reducing w_{10} shifts both sides of the IC condition upwards, so optimally we want $x_{00} = p_{01}+c_1$. Then

$$(1 - w_{10})x_{10} + w_{10}(p_{0\cdot} - c_1) \leq p_{01} + c_2 \leq (1 - w_{10})x_{10} + w_{10}(p_{1\cdot} - c_1)$$

and a similar argument leads to the conclusion that $x_{10} = p_{11} - c_1$, and

This leads to the result that this mechanism is feasible with

$$w^*_{10} = \frac{p_{1\cdot} - p_{01} - C}{p_{1\cdot} - p_{0\cdot}}$$

which is between zero and one when $C \in (p_{0\cdot} - p_{01}, p_{1\cdot} - p_{01})$. Combining with the condition above, this case requires

$$C \in (\max\{p_{0\cdot} - p_{01}, p_{1\cdot} - p_{11}\}, p_{1\cdot} - p_{01}).$$

For $f \sim$ Beta, it is easy to show that the second term in the max is always strictly greater. Also, the upper bound is greater than $p_{1\cdot} - p_{0\cdot}$, and case 4a is feasible for C greater than this and has zero chance of war. So case 4c has a chance of being optimal only if $C \in (p_{1\cdot} - p_{11}, p_{1\cdot} - p_{0\cdot})$.

The upper bound is the same as for Case 4b and is the lower bound for fully peaceful Case 4a. For $f \sim$ Beta and $\bar{p} > 1/2$, algebra shows that the lower bound is greater than $p_{11} - p_{0\cdot}$, the lower bound for the range of C in which Case 4b has $w_{11} \in (0, 1)$. So Case 4c is supportable only when Case 4b exists also and has $w_{11} \in (0, 1)$ as given above. The ex ante chances of war in these two cases compare as

$$(1 - \bar{p})\frac{p_{10} - p_{0\cdot} - C}{p_{1\cdot} - p_{0\cdot}} \gtrless \bar{p}\frac{p_{1\cdot} - p_{01} - C}{p_{1\cdot} - p_{0\cdot}},$$

where the LHS is Case 4b and the RHS Case 4c. From algebra, the numerator on the LHS is strictly less than the numerator on the RHS, which implies, with $\bar{p} > 1/2$, that war is less likely ex ante in Case 4b.

Proof of Proposition 2a

Case 4b obtains whenever $C \in [p_{10} - p_{1\cdot}, p_{1\cdot} - p_{0\cdot}]$. For any prior beliefs f, $p_{10} \geq p_{1\cdot}$ and $p_{1\cdot} = \bar{p} + \sigma^2/\bar{p}$. Since the latter approaches 1 as $\bar{p}$ approaches 1, the difference $p_{10} - p_{1\cdot}$ approaches zero as $\bar{p}$ goes to 1. Thus, for large enough $\bar{p}$, Case 4b can obtain for any fixed $C > 0$.

The ex ante probability of war is Case 4b is $(1 - \bar{p})w_{11}$, which approaches zero as $\bar{p}$ approaches 1. For costs C greater than the upper bound of the range for Case 4b, Case 4a obtains, in which the probability of war is zero. $\square$