

Mutual Optimism as a Rationalist Explanation of War

Branislav L. Slantchev*

Department of Political Science, University of California – San Diego

Ahmer Tarar

Department of Political Science, Texas A&M University

December 26, 2009

Abstract. Blainey (1988) popularized the argument that crises are more likely to end in war when two nations disagree about their relative bargaining power. This is known as the *mutual optimism* cause of war. Fey and Ramsay (2007) argue that existing models of crisis bargaining cannot properly evaluate this explanation. They propose an alternative class of models with features they claim are better suited for the purpose, and prove that war cannot occur with positive probability in equilibrium in any model of this class. They conclude that the mutual optimism explanation is logically incoherent. We provide a modern rationalist specification of the mutual optimism explanation using our standard models and defend them against the five allegations Fey and Ramsay make against them. We further show that their result has little to do with optimism but instead depends on two indefensible assumptions: (1) actors can unilaterally avoid war by imposing peace terms on opponents who find them worse than war, and (2) the terms of any peace outcome cannot depend on crisis behavior. Their class of models cannot be used to study crisis bargaining in general and mutual optimism in particular.

*We have benefitted from illuminating discussions with Bob Powell, Kris Ramsay, and Mark Fey. We thank Christina Schneider and Megumi Naoi for helpful comments on previous drafts. This research was supported by the National Science Foundation (grant SES-0518222 to Slantchev and grant SES-0518945 to Tarar).

Blainey (1988, 114) offers what has become one of the most celebrated explanations of why war occurs: “war is usually the outcome of a diplomatic crisis which cannot be solved because both sides have conflicting estimates of their bargaining power.” In this view, war happens when at least two states embroiled in a dispute cannot construct a mutually acceptable deal that would satisfy their conflicting demands, and when one of them—often in exasperation—resorts to arms in an attempt to impose its will by force. The crucial part of the explanation requires one to specify the reason, or reasons, for that inability to agree on a peaceful deal.

Blainey’s argument is that the fundamental cause is the two sides’ conflicting estimates about how much they can secure by force. Because a state can always go to war to enforce its demands, what it expects to gain by fighting constitutes the minimum it would require in any peaceful negotiation. These expectations are based on subjective estimates of one’s probability of winning the war (which depends, among other things, on the distribution of power, and the quality and morale of one’s armed forces), the costs required to do so, the duration of fighting, the behavior of allies and antagonistic states alike, the ability of one’s economy and population to sustain the war effort, and so on. If both sides expect to gain a lot by fighting—perhaps because both expect to win with near certainty at an acceptably low cost—then there may exist no negotiated deal that can satisfy their minimum demands. War becomes the inevitable outcome.

This argument is now generally known as the *mutual optimism* explanation of war, and is among the most widely accepted explanations of why war occurs.¹ In a recent article in the *American Journal of Political Science*, Fey and Ramsay (2007, henceforth F&R) argue that this explanation is mistaken on logical grounds, and that war cannot occur between two actors because of mutual optimism about the likelihood of victory. Furthermore, war occurs in many formal models under incomplete information not because of mutual optimism but because actors are ‘locked in’ by the extensive form of the game (to wit, war happens by assumption because analysts have not given actors enough flexibility to avoid it). If correct, their analysis would have serious implications for the conflict literature, formal and non-formal alike. It would show that a well-established explanation for war is theoretically incoherent. It would also reveal that in the large majority of formal models of war under incomplete information, war occurs as a consequence of arbitrary modeling restrictions. In other words, these models would not provide any explanation of war whatsoever!

We have two goals in this article. The first is to provide a theoretical account of the mutual optimism mechanism that is more comprehensive than has been done previously. One of the major problems with F&R’s critique is that it is predicated on an exceptionally narrow view of this mechanism, a view that is not in line with developments in the theoretical literature over the past three decades. The evolution of the mutual optimism explanation is implicit in our existing models, and we now make it explicit. We use a simple ultimatum crisis bargaining model to establish a coherent specification of this mechanism. Since this model is part of the family F&R criticize, we use it to demonstrate the flaws in their reasoning and explain why their seemingly intuitive critique cannot be used to undermine the validity of this explanation for war. Furthermore, we show that their indictment of formal models of

¹Wittman (1979); Lebow (1981, 242–43); Levy (1983, 82–86); Blainey (1988); Jervis (1988, 676); Van Evera (1999, 16); Johnson (2004); Stoessinger (2005, 211).

war under incomplete information is likewise baseless, for both technical and conceptual reasons.

Our second goal in this article is to demonstrate that F&R's model fails to establish the incoherence of the mutual optimism explanation even within the narrow confines of their own conception of it. Even if we grant that F&R's is an acceptable definition of this mechanism, we can still show that their results have nothing to do with mutual optimism but are, in fact, due to two highly implausible assumptions they make. We show that F&R's model would not admit war even in situations where war *should* occur given the payoffs under complete information. Since there is no optimism in this situation, such a result implies that the absence of war under incomplete information might have nothing to do with optimism either. In fact, this is precisely the case: we demonstrate that F&R's "no-war" result is the consequence of their assumptions that (i) allow one actor to impose peace on terms that are worse than the opponent's expected payoff from war, and (ii) do not allow these terms to depend on crisis behavior. In making war a mutual act, F&R have neglected that peace must be a mutual act as well. But if both sides are to agree on the peace terms, one must specify what these terms are and how the actors get to them. They cannot be exogenously specified without reference to crisis behavior.

We conclude that the mutual optimism mechanism, in its modern specification, is more sophisticated than the simple version F&R set up, that it does provide a coherent rationalist explanation for war, and that F&R-style arguments cannot be used to undermine it.

1 The Mutual Optimism Mechanism: A Synthesis

How does mutual optimism cause war? According to F&R, actors involved in a crisis evaluate the military balance and then decide whether to take the available (exogenously fixed) peace settlement or 'bet' on obtaining a better payoff by fighting:

A typical story for how war might result from mutual optimism is as follows. Suppose the leaders of two countries have information about their military forces and tactics that their opponent does not. Moreover, suppose that this information influences each leader's assessment of their country's likelihood of success in combat. If both leaders then believe that their side possesses the "stronger" force, both sides may think they will prevail militarily and thus both leaders may choose to fight rather than pursue a peaceful settlement. (741)

We have been unable to find this version of the mutual optimism (MO) mechanism described anywhere in the literature.

Truth be told, we do not have a clear and widely accepted definition of the MO mechanism even though references to it abound (see footnote 1). To begin with, one must distinguish between its rationalist and non-rationalist specifications. Most of the non-formal literature going back to Blainey (1988) is actually in the non-rationalist vein, as Fearon (1995, 391–93) points out, and it is essentially silent about the process through which mutual optimism leads to war; all it says is that optimism results in inflated expectations about fighting relative to peace, which makes actors unwilling to settle on terms their opponents are willing to offer. These inflated expectations might be due to psychological biases, nationalist fervor, bounded rationality, evolutionary adaptation, or private information about

capabilities or resolve, among others. Whatever the source of optimism, the mutual unwillingness to offer better terms somehow causes negotiations to break down in war.

Observe that mutual optimism is only a starting point: it shows why the actors might be unwilling to offer each other war-avoiding peace terms. It does not tell us how this leads to war. This might not be a problem for a non-rationalist account: if an actor believes himself militarily superior to his opponent and thinks she is incapable of seeing the evident truth of that estimate, nothing she does would alter his conviction. He will even be quite willing to fight in order to teach her a lesson (presumably the outcome of the war would cause a revision of beliefs). If she is likewise optimistic, his willingness to fight will similarly signify nothing: she will be just as eager to fight to teach him a lesson.

Whatever the merits of such an account, one cannot maintain a fully rationalist explanation based on MO without specifying precisely how MO causes war in a world in which actors behave rationally and this fact is common knowledge (Fearon, 1995). The formal literature over the past decade and a half has provided us with a coherent mechanism whose fundamental insight is that *MO causes actors to engage in behavior that ends in war even in environments where settlements exist that would make both better off and where they would be able to locate such settlements if they had better information*. It is not that MO simply makes actors to prefer war to peace. Instead, when optimism makes actors unwilling to agree to the terms their optimistic opponents are willing to offer, it is the attempt to overcome this problem that sometimes results in war. In other words, MO leads to war not as a direct consequence of preferences but as a result of strategic behavior by the optimistic actors.

What are these war-causing behaviors? Costly signaling (actions that deliberately risk war to reveal the credibility of one's war expectations) and the risk-return trade-off (action that runs a higher risk of war in order to obtain better terms of peace) are two specific forms that the mechanism can take. Note that, contrary to F&R's claims, these are not *alternatives* to the MO mechanism, they are the means at its core; they are *how the mechanism operates*. Eliminating them would gut the mechanism and render it meaningless.

F&R's version of the MO mechanism ignores all of this and is too simplistic even compared to the somewhat under-theorized non-rationalist specifications. Since the rationalist version is the most developed and because it is, in fact, what F&R attack, we shall henceforth restrict our attention to it. To our knowledge, there is no comprehensive rationalist specification of the MO mechanism in the literature, perhaps because it is so obvious to many. It is therefore with some embarrassment that we offer one, as a means of elaborating its workings and elucidating why all of F&R's critiques fail to undermine it.

1.1 The Standard Model

We begin by describing the standard setup for the inefficiency puzzle (Fearon, 1995). Two risk-neutral states, S (henceforth "she") and D (henceforth "he"), dispute the distribution of an infinitely divisible good whose size is normalized to 1. They can either agree to divide it peacefully or fight over it. War is a winner-take-all costly lottery: D wins with probability $p \in (0, 1)$, S wins with probability $1 - p$, and both suffer costs, $c_S, c_D > 0$. The expected payoffs from war are $p - c_D$ for D and $1 - p - c_S$ for S . Since they sum up to less than the total size of the benefit, there always exists a range of settlements, $B = [p - c_D, p + c_S]$,

that both sides prefer to war. A fully rationalist account must explain why the two sides fail to reach an agreement when the existence of this range is common knowledge.

Consider now a simple variant of the ubiquitous ultimatum crisis bargaining model (henceforth, “standard model”). There is some status quo division of the good $(d, 1 - d)$, where $d \in [0, 1]$ is D ’s share and $1 - d$ is S ’s share. A state is “satisfied” with the status quo if its payoff from living peacefully at that distribution is at least as high as its expected payoff from war; otherwise it is “dissatisfied.” If $d \in B$, both players are satisfied; if $d < p - c_D$, then D is dissatisfied but S is satisfied, and if $d > p + c_S$, then D is satisfied but S is dissatisfied. If both are satisfied, there is no crisis, there will be no revision of the status quo, and no war. These assumptions imply that at most one player can be dissatisfied (Powell, 1999). Suppose, without loss of generality, that D is dissatisfied. Then there is a crisis in which war cannot be avoided unless the status quo is revised in D ’s favor.

The sequence of moves is as follows: S makes a take-it-or-leave-it (TILI) offer $(y, 1 - y)$, where $y \in [0, 1]$ is D ’s proposed share and $1 - y$ is S ’s proposed share. If D accepts this offer, the status quo is revised accordingly and the game ends peacefully with players getting the payoffs from this new distribution. If D rejects this offer, the game ends in war.

With complete information, this model has a unique subgame-perfect equilibrium (SPE), in which D accepts any $y \geq p - c_D$, and S offers exactly $y^* = p - c_D$. The game ends peacefully with the status quo revised in D ’s favor. The fundamental point is that war will not occur under complete information; *states manage to avoid war because they agree to revise the status quo on mutually acceptable terms.*

To examine the mutual optimism explanation, consider an incomplete-information version of the model. To keep things as simple as possible for a crisp illustration of the results, we consider the two-type case with one-sided asymmetric information about military capabilities.² Assume that D can be strong, so his probability of winning the war is $p_h \in (0, 1)$, or weak, so his probability of winning is $p_w \in (0, 1)$ such that $p_w < p_h$. D knows his own type but S is uncertain: she believes that he is strong with probability q and weak with probability $1 - q$. When she makes her offer, S is unsure whether rejection would lead to a war with a strong opponent or a weak one. The solution concept is perfect-Bayesian equilibrium (PBE). The following proposition describes the well-known result. The proof is straightforward and is omitted.

PROPOSITION 1. *In all PBE, D accepts any $y \geq y_h = p_h - c_D$ if strong, any $y \geq y_w = p_w - c_D$ if weak, and rejects any other offer. The offer S makes depends on the critical belief threshold, $k = \frac{p_h - p_w}{p_h - p_w + c_D + c_S} \in (0, 1)$, as follows:*

- (i) *if $q > k$, then S offers y_h , which D always accepts;*
- (ii) *if $q \leq k$, then S offers y_w , which D accepts if weak but rejects if strong.³*

War occurs when $q \leq k$ and D happens to be strong.

This, of course, is the risk-return trade-off result, which is among the most widely accepted rationalist explanations for war (Powell, 1999). The only way to avoid war in this

²Fearon (1995) considers uncertainty about the costs of fighting rather than relative military capabilities. We chose the latter because it is closer in spirit to the mutual optimism idea as originally proposed by Blainey (1988).

³Technically, if $q = k$, S is indifferent between making the large and small offers, and could choose either one or mix. This is a knife-edge condition and is uninteresting.

setting is to offer D at least his expected payoff from war. S can always do this by offering peace terms that the opponent will accept if he is strong: they are so generous that he would certainly accept them if he happens to be weak. Even though S can always secure peace, doing so is not always optimal because such generous peace terms represent unnecessary concessions if D is weak. On the other hand, failing to offer them carries a risk of war if D is strong. S can resolve this dilemma by balancing the risk of having her offer rejected against the gain of obtaining better settlement terms if it is accepted. Because war is costly, this trade-off is only optimal if the risk is not too high; that is, if she believes that there is a good chance her opponent is weak.

F&R claim that this standard model cannot be used to examine the MO argument, but we now show that this is not so: *the standard model actually provides a well-specified mechanism through which MO can cause war*. F&R's argument has five components: (i) war is a unilateral act in the standard model and does not require the agreement of both players to occur; (ii) because one player can start a war without the consent of the other, MO is irrelevant to its occurrence (it is enough that only one side is optimistic); (iii) when war does occur, players are not optimistic in the instant before war begins; (iv) the risk-return trade-off is an alternative explanation for war, which should be considered separately from MO; and (v) players would want to avoid war on the "eve of war" but are prevented from doing so by arbitrary restrictions on the game-tree. The first four claims are related to MO in particular, and the last is a general attack on almost all existing models of war. We now refute each of these in turn, and in the process provide a comprehensive modern definition of the MO mechanism.

1.2 War Is a Mutual Act in the Standard Model

F&R argue that if war is a unilateral act—meaning that an actor can impose that outcome on the opponent—then “the concept of war by mutual optimism loses meaning” (745). We agree: for MO to provide an explanation for war, it should be the case that war occurs when optimistic actors make choices that only *collectively* lead to bargaining failure. For instance, if one side has a very high expected payoff from war but the other was not allowed to make any concessions, war would occur but it would not be caused by MO. The trouble is, war is not unilateral in the standard model.

How do F&R propose to capture the idea of war as a mutual act? They require that each side has an action (“negotiate”) that guarantees the peace outcome regardless of the action of the other player.⁴ It might appear that in the standard model D unilaterally causes war because it is his rejection that starts it. Since S has no way to “stand firm,” her consent appears unnecessary for war to occur.

Observe now that D certainly has a strategy that guarantees peace regardless of what S does: accept any offer. Moreover, S also has a strategy that guarantees peace, at least for anything that D might rationally do: offer some $y > y_h$. It is true that this is weaker than F&R's requirement because it does not guarantee peace for *any* strategy for D (e.g., a strategy of rejecting all offers). However, it does guarantee peace for any strategy D would rationally play: accept offers that are greater than his expected war payoff. These

⁴F&R sometimes refer to this assumption as the requirement that both sides have to choose to “stand firm” for war to occur. As they note on pp. 739 and 745, this boils down to the same thing.

are the only strategies that D would ever play in equilibrium, and since F&R maintain the rationality assumption, our definition is close to theirs.

Note further that in the endogenous peace-terms setting of the standard model (and unlike the F&R setup where these terms are specified exogenously, a problem that we will return to), the satisfied state will never *choose* war. As Proposition 1 shows, it is always better to make an offer that would satisfy at least the weak opponent because doing so buys some positive probability of peace at terms that are better than war. However, because that state also has the option to make the large peace-inducing offer, the choice to make the limited one that carries a risk of war can naturally be interpreted as that state “standing firm.”

This means that war is a mutual act in the standard model: it can occur *only* if S chooses to make a limited offer (which she certainly knows carries a positive risk of being rejected), and if D chooses to reject it. That is, war occurs only when both actors choose to forsake the strategies that guarantee a peaceful outcome. It takes two to make war here.

Finally, as we explain in some detail in the second part of the article where we consider F&R’s model, the definition of mutuality they require is exceedingly demanding for it implies an ability to impose peace terms on an actor whose expected payoff from war is higher.

1.3 War Occurs *If, and Only If*, Mutual Optimism is Present

F&R further argue that “if the correct model of war is one in which any single country can start a war, the presence of mutual optimism is irrelevant and, therefore, not a coherent rationalist explanation of war” (751). Suppose we were to grant F&R’s highly demanding definition of mutuality. Would doing so make MO irrelevant for the results in the standard model?

To answer this question, we need a more precise definition of optimism. When game-theoretic analyses find that war can occur under incomplete information but not under complete information (e.g., in settings like the standard model), they generally call it war due to incomplete information rather than war due to mutual optimism. There is, however, a clear link between the two.

Let us now define optimism in our model. Consider the uninformed state, S , first. She can be said to be optimistic about her military prospects when she is sufficiently confident that she faces a weak opponent (when $q < k$). On the other hand, she can be said to be pessimistic when she is sufficiently confident that she faces a strong one. This is so because in our specification of the standard model, a war against a strong opponent can end in victory with a lower probability than a war against a weak opponent. If this were a model of two-sided incomplete information, D ’s optimism would be defined analogously. Because we assumed that he knows the actual military balance, the definition here boils down to the actual state of that balance. He can be said to be optimistic when he is the strong type (because his probability of winning is high), and pessimistic when he is the weak type (because that probability is low).

Table 1 shows the conditions for optimism under which war occurs in the standard model. When S is pessimistic, she make the generous offer y_h that D accepts regardless of type. Similarly, when D is pessimistic, he accepts any offer that S makes in equilibrium (because the worst terms she ever offers match his expected war payoff). Hence, with unilateral op-

	<i>D</i> pessimistic (weak, p_w)	<i>D</i> optimistic (strong, p_h)
<i>S</i> pessimistic ($q > k$)	peace (<i>S</i> offers y_h , <i>D</i> accepts)	peace (<i>S</i> offers y_h , <i>D</i> accepts)
<i>S</i> optimistic ($q \leq k$)	peace (<i>S</i> offers y_w , <i>D</i> accepts)	war (<i>S</i> offers y_w , <i>D</i> rejects)

Table 1: Optimism and War in the Standard Model.

timism war cannot occur: *mutual optimism is a necessary condition for war in the standard model*. If, on the other hand, both players are optimistic, *S* makes the low-ball offer y_w and *D* rejects it (because it is worse than his expected war payoff). Hence, with mutual optimism war always occurs: *mutual optimism is a sufficient condition for war in the standard model*. Clearly, MO is quite relevant for the occurrence of war in this model.

F&R argue that to evaluate properly the MO explanation for war, one needs a model in which war occurs only if mutual optimism is present (because otherwise it might be something else causing it, pp. 739, 750). The finding that war occurs in the standard model if, and only if, mutual optimism is present is *stronger* than this requirement because it does not simply satisfy necessity, but also shows that mutual optimism can entirely determine whether war occurs. Hence, the standard model provides a coherent explanation for war based on mutual optimism.⁵

1.4 The Risk-Return Trade-Off Is *Not* an Alternative to Mutual Optimism

We have identified the risk-return trade-off as one means through which mutual optimism can result in war. F&R state that there might be other “viable rationalist explanations” for war, such as “incomplete information mechanisms that are not associated with mutual optimism” (752). The only other such mechanism they cite is the risk-return trade-off, and they are very explicit that they consider it an alternative to mutual optimism (750). (In fact, they reject the risk-return trade-off itself as dependent on arbitrary restrictions in the extensive form, a claim we deal with in the next section.)

It is not entirely clear to us what position F&R wish to take. As far as we have been able to discern, their argument is basically about private information and how it can directly cause war (e.g., when they write that “in this setting, the root cause of war is the inconsistent expectations that arise because of private information” and that they are “working with the definition of mutual optimism as war due to inconsistent beliefs,” 738).

We are somewhat puzzled by this. *Mutual optimism* is just a set of conditions that describe beliefs: both sides having high expectations about war (which might be due to private information about capabilities but which also might be due to a host of other factors, as we noted above). The mutual optimism *explanation* must show how these expectations cause war; it has to specify the reasons these expectations cannot be reconciled without fighting.

⁵This also contradicts F&R’s claim that their results “apply to any game where peace prevails in the absence of mutual optimism” (740). In the standard model, peace does indeed prevail in the absence of mutual optimism and yet their ‘no-war’ theorem does not hold.

In particular, since peace requires that both sides agree to its terms, the mechanism must explain why actors persist in their unwillingness to offer terms that the opponent demands in order to preserve the peace. Our puzzlement stems from F&R's attempt to critique the *explanation* while simultaneously omitting the mechanism through which it operates.

The non-rationalist literature has skirted this requirement by arguing that actors would not reduce their optimism even in the face of abundant evidence that contradicts it simply because they are not rational. This venue is not open to us, but neither it is to F&R who are also interested in a rationalist specification of the MO mechanism. Unfortunately, as we explain in our conclusion, the information transmission mechanism they borrow from the literature on common knowledge does not (and cannot) work in a crisis bargaining setting. Moreover, their modeling approach does not permit them to specify a reasonable mechanism even if they wished to.

In contrast, we have provided one: the risk-return trade-off. As we have been at pains to show, this is not an alternative to the MO explanation, it is one mechanism through which MO can cause war. High expectations about war (because she believes D is weak) cause S to forsake the strategy that guarantees peace and to make a limited offer, which she is fully aware carries a risk of war, to D . High expectations about war cause D to reject this offer even though he is fully aware that doing so will result in war. Thus, when MO is present, the actors engage in specific behaviors and their interaction ends in war.

There are other means through which MO can cause war. For instance, when actors believe their opponent has unreasonably high expectations about war, they might attempt to lower them, which usually entails taking actions that run a risk of war (Schelling, 1966). As is well-known, when both sides are very optimistic, they can end up taking actions that commit them to war (Fearon, 1996). Thus, *costly signaling*, or the attempt to overcome the problem created by MO, is another way through which MO can lead to war. Another example is in contexts where military preparations are very costly: an optimistic actor might choose to under-prepare in the belief that his opponent is weak, and his force levels might prove inadequate to compel an optimistic strong opponent (Slantchev, 2005).

In their desire to purge their model of any such 'alternatives', F&R have ended up gutting the MO mechanism. It is no surprise that they cannot find that mutual optimism can cause war; after all, they have eliminated every connection between the two that we have come up with over the past four decades.

1.5 Players *Can* Be Optimistic “On the Eve of War”

As we noted above, F&R do not actually consider the risk-return trade-off to be a fully coherent explanation for war. The reason can be illustrated very simply with our model. Suppose $q \leq k$ and D is strong, so mutual optimism exists. In equilibrium, S will offer y_w , and D will reject it. Now, in the instance after that rejection but before war begins, S is no longer optimistic. Since D 's strategy is to accept y_w if weak, rejection can only reveal that he is strong. Thus, upon observing that rejection, S will immediately update her beliefs and conclude that D is strong. But we know that in an environment where war is costly but peace is not, she is better off satisfying the strong type instead of fighting: $1 - y_h = 1 - p_h + c_D > 1 - p_h - c_S$. In other words, S would prefer to offer y_h and avoid war but is prevented from doing so by the extensive form of the game. As F&R put it, there

is no mutual optimism “on the eve of war.”

This line of reasoning has two components—the claim that following rejection of her offer, S would want to revise its terms to *ensure peace*, and the claim that she is *artificially constrained by the game-tree*. The latter is really an attack on almost every model of war we have. Because of its importance, we deal with it separately in the section that follows. Here we are content to show that whereas the first part of F&R’s claim does obtain in the two-type model, it fails in models with more than two types (which preserve the MO risk-return trade-off results intact).⁶

Consider a variant of the standard model where D can be either weak, p_w , moderately strong, p_m , or very strong, p_h , with $p_h > p_m > p_w$. S is unsure which type she is facing, but believes that her opponent is strong with probability $q_h \in (0, 1)$, moderate with probability $q_m \in (0, 1)$, and weak with probability $1 - q_m - q_h \in (0, 1)$. The following proposition establishes that mutual optimism will cause war through the risk-return trade-off mechanism in this model as well.

PROPOSITION 2. *In all PBE, D accepts any $y \geq y_h = p_h - c_D$ if very strong, any $y \geq y_m = p_m - c_D$ if moderately strong, any $y \geq y_w = p_w - c_D$ if weak, and rejects any other offer. The offer S makes depends on the critical belief thresholds*

$$k_1 = 1 - q_m \left(1 + \frac{C}{p_m - p_w} \right), \quad k_2 = \frac{p_h - p_m}{p_h - p_m + C}, \quad k_3 = \frac{p_h - p_w - q_m(p_m - p_w + C)}{p_h - p_w + C},$$

where $C = c_D + c_S$, as follows:

- (i) if $q_h > \max\{k_2, k_3\}$, then S offers y_h , which D always accepts;
- (ii) if $q_h < \min\{k_1, k_3\}$, then S offers y_w , which D accepts only if weak;
- (iii) if $k_1 < q_h < k_2$, then S offers y_m , which D accepts only if weak or moderately strong.

War occurs only if S is sufficiently optimistic and D sufficiently strong.

Proof. The strategy for D follows from subgame perfection, and implies that if p_h accepts some offer y , then so will p_m and p_w , and if p_m accepts some offer, then so will p_w . It follows that S will choose among three possible offers: y_h , which all three types accept; y_m , which only the weak and moderate types accept; and y_w , which only the weak type accepts. S will always prefer to offer at least y_w rather than some unacceptable offer $y' < y_w$ that would certainly lead to war: $U_S(y_w) - U_S(y') = (1 - q_h - q_m)C > 0$. Therefore, we only need to consider her preference among the three offers that provide for a chance of peace. As before, we ignore knife-edge conditions.

Algebra shows that S prefers y_w to y_m when $q_h < k_1$, prefers y_m to y_h when $q_h < k_2$, and prefers y_w to y_h when $q_h < k_3$. Conditions (i), (ii), and (iii) follow immediately. \square

Suppose now that $q_h < \min\{k_1, k_3\}$, so S is very optimistic and offers D the worst possible terms that he might ever accept, y_w . Since D would only accept this if he is weak, rejection signals that he is either very strong or moderately strong. The only way for S to ensure peace now would be to offer y_h . It is easy to see that there exist conditions under which S will not want to ensure peace with such a generous offer (if she could make another

⁶We thank an anonymous reviewer whose comments led us to think about this.

TILI offer) but would instead make another limited offer, y_m this time, which the strong opponent would still reject, causing war. As we know from Proposition 1, the condition for this preference is that S 's belief that D is strong (which is the posterior $q_h/(q_m + q_h)$ by Bayes rule) is sufficiently small. Since we already supposed that q_h is small relative to q_m , this condition is easy to satisfy.⁷ In other words, S has retained sufficient optimism even “on the eve of war” and would not make the offer that would guarantee peace. F&R’s claim is an artifact of the two-type case (it also fails if we assume any countable number of types greater than two or a continuum).

One possible response to this would be to agree that S will retain optimism after her initial demand but to argue that she would not be optimistic if she made a second offer and *that* got rejected too. (It might take many more rejections depending on the degree of uncertainty and amount of initial optimism.) In other words, there will always be a “final” offer whose rejection would reduce S 's optimism to the point that she would prefer to make the offer that would guarantee peace. But of course, this just begs the question of how precisely this optimism gets reduced, a question that F&R are at pains to avoid but that is essential to the MO mechanism we have used as an illustration. This leads us to a fundamental point that also addresses F&R’s most sweeping claim.

1.6 War Is *Not* an Artifact of Arbitrary Restrictions on the Game-Tree

The final criticism F&R level at the standard model is that in it war occurs under incomplete information only because the extensive-form of the game does not permit the actors to avoid it once they realize they do not wish to fight and that they would rather reach a peace settlement. Their argument is clearest in the two-type case we have analyzed: after her offer is rejected, S knows she is facing the strong D , and would strictly prefer to make a war-avoiding last-ditch proposal that would satisfy him. However, the game-form does not allow her to make that choice because D 's rejection automatically causes war.⁸ As F&R put it,

Reflecting on this example, we can give an intuitive statement of our main result in the following way. If it is common knowledge that countries are going to fight, and these countries have a “hot line” available, then at least one side will always want to make a call and a proposal that will be accepted and avoid the war. That is, our result

⁷As a numerical example, suppose that $c_D = c_S = 0.1$, $p_h = 3/4$, $p_m = 1/2$, $p_w = 1/4$, and $q_m = 1/3$. Then, $k_1 = 0.4$ and $k_3 = 0.5$, and so S will offer y_w if $q_h < 0.4$. If this offer is rejected, then if S could make another TILI offer, she would offer y_m rather than y_h if $q_h/(q_m + q_h) < 5/9$, which simplifies to $q_h < 5/12$ (> 0.4). Thus, if $q_h < 0.4$, then S initially offers y_w , and if this offer is rejected and she could make another TILI offer, she would still be optimistic enough to offer y_m rather than y_h .

⁸This is the reason F&R do not seem to think that the risk-return trade-off is a viable explanation for war because “there is no way for [S] to react to the private information of [D]” (750). They cite Leventoğlu and Tatar (2008) as allegedly having shown that the risk-return result “is driven to a large extent by (somewhat arbitrary) assumptions regarding the extensive form of the bargaining process” (fn. 17). However, what Leventoğlu and Tatar (2008) actually do is show that private information does not necessarily cause risk-return behavior, as originally suggested by Powell’s (1996) finding of a unique equilibrium with that feature. They show that uniqueness depends on one actor having all the bargaining power, and that if that is not the case, multiple equilibria become possible, and some of them do not exhibit the risk-return trade-off. This implies that mutual optimism may not always “activate” this particular war-causing mechanism. It most certainly does not imply that the mechanism is unreasonable or unrealistic.

applies to a situation where countries can discuss war not only before making it, but after a proposal has been made and rejected. In situations where a firm offer is made that, if rejected, leads to certain war, equilibria with war can exist, but not because of mutual optimism; one side would prefer to settle but is locked into a war by the extensive form (751).

If this argument is correct, it has serious consequences for almost *all* game-theoretic models of crisis bargaining under incomplete information, not just the standard ultimatum model, because these models all have “final” decision-nodes. That is, decision-nodes where the choice to “fight” starts a war irrespective of how this choice would update the other actor’s beliefs; that is, regardless of whether she would want to make an “eve of war” offer. Because these models disallow such presumably war-avoiding offers, they impose war by analyst fiat, not as a consequence of optimal behavior. As a result, they do not explain war at all.⁹

What features do F&R require from a legitimate model of war under incomplete information? They write that “in practice, if one country chooses to stand firm, the other country can stop a war by inducing the bargaining procedure instead” (745). This implies that any actor who would rather bargain than fight should be able to avoid war even after his opponent has chosen to fight. As they put it, “at any given moment before war begins, a state could continue negotiations with the hopes of avoiding a fight” (739). From a substantive standpoint, it is very doubtful that ‘in practice’ an actor can always avoid war when the opponent has decided to fight; and it is perhaps even more doubtful that he can do so ‘at any given moment before war begins’. We leave these problems aside to focus on the theoretical implications that F&R’s claim, if true, would have for crisis bargaining behavior.

F&R do not offer an example of a sequential-move model that satisfies their requirement. We now modify our standard model to allow S to avoid war after D ’s rejection. Consider the following infinite-horizon game: in each period S makes an offer, which D can either accept or reject. If he accepts, the game ends on the peace terms accepted. If he rejects, S can either let the rejection stand or continue negotiations. If she lets the rejection stand, the game ends in war; otherwise, the game continues to the next period. Initially, there is a status quo distribution of benefits, $(d, 1 - d)$, where d is D ’s share. The payoffs are as follows. If an agreement is reached on some division $(y, 1 - y)$ in period t ($t = 0, 1, 2, \dots$), then D ’s payoff is $\sum_{i=0}^{t-1} \delta^i d + \sum_{i=t}^{\infty} \delta^i y$, and S ’s payoff is $\sum_{i=0}^{t-1} \delta^i (1 - d) + \sum_{i=t}^{\infty} \delta^i (1 - y)$, where $\delta \in (0, 1)$ is the common discount factor. (If the players never reach an agreement, the status quo distribution remains in place forever.) If players go to war in some period t , then D ’s payoff is $\sum_{i=0}^{t-1} \delta^i d + \sum_{i=t}^{\infty} (p - c_D)$, and S ’s payoff is $\sum_{i=0}^{t-1} \delta^i (1 - d) + \sum_{i=t}^{\infty} (1 - p - c_S)$. As before, assume that S is satisfied but D is not.

Since we assume one-sided incomplete information, this modification is sufficient to implement F&R’s ‘hotline’ notion that the uninformed actor should have the opportunity to avoid fighting by continuing negotiations even on the “eve of war.” The following proposition fully characterizes the set of equilibrium outcomes.

PROPOSITION 3. *Every subgame-perfect equilibrium of the modified infinite-horizon game*

⁹As we show in the second part of the article, F&R’s own model is vulnerable to the (much stronger) analogue of this argument because it does not allow D to react to S ’s unilateral imposition of peace terms that he finds worse than war.

is peaceful and the status quo is never revised.

Proof. The only way the game can ever end in war is for S to let D 's rejection stand. In any arbitrary period, she can ensure the status quo payoff by any strategy that always continues bargaining. Because she is satisfied, this payoff is strictly higher than her expected war payoff. Hence, S will never let rejection stand. To see that the status quo will never get revised, note that any revision requires that S propose it and D accept it. Because D is already dissatisfied, he will never agree to any redistribution that would leave him worse than the status quo (i.e., any $y < d$), so S cannot obtain better terms. Similarly, S would never propose terms worse than the status quo (i.e., any $y > d$) if these have any chance of being accepted (because she can always keep the status quo payoff). \square

The interaction can end in one of two ways: either in some period S offers the status quo terms and D accepts them, or D always rejects her offers but she never lets the rejection stand (so “negotiations” continue forever). It is straightforward to construct SPE with immediate acceptance of the status quo or interminable negotiations.¹⁰ The structure and degree of incomplete information about D —indeed, whether it is even present—are completely irrelevant. Regardless of how her beliefs might have changed during the interaction, S will never agree to a revision of the status quo in D 's favor. *In every equilibrium, the dissatisfied actor is worse off than his expected payoff from war.*

We can only describe our result here as silly, but it demonstrates the fundamental problem with F&R's requirement and their logic. D is forced into a very unpleasant peace because every time he opts for war (which is better than these peace terms), S never allows fighting to occur but continues the “negotiations” by making a new offer. F&R's logic does not work because (i) it implies that the dissatisfied actor needs the *permission* of the satisfied actor to go to war; and (ii) it implicitly assumes that the “eve-of-war” offer that S will make would avoid war because it would satisfy D .

These two assumptions are incompatible: if S can negate D 's credible threat to fight and can always impose peace that preserves the terms that favor her, she has absolutely no reason whatsoever to offer D better terms. She can have a favorable peace for nothing: the requirement that war cannot begin without her consent frees her from the necessity of inducing her opponent not to fight. We shall call this the *unilateral peace requirement*.

Generally, we think that peace is a mutual act: it requires both sides to agree to abide by its terms. This implies that for peace to obtain, both actors must believe that they can do no better by fighting. In our standard model with complete information peace always obtains but does so only because S agrees to revise the status quo in D 's favor, to the point where D no longer expects to improve upon its terms by fighting. She has to do so because otherwise D will opt out of peace into a war that is even worse for S than the concessions she has to make to avoid it. (As we shall see in the second part, F&R's model short-circuits this logic because it assumes exogenous terms to peace settlements and does not require both players to agree to them.) It follows, then, that the ‘hotline’ argument is specious: in the context F&R propose, there is no reason why the war-avoiding “eve of war” offer should be the revision of terms that would satisfy the actor who is opting to fight.

¹⁰See the technical appendix for two SPE that have these features.

Our argument can go even further. In their criticism of the standard model, F&R argue that “on the eve” of war S is no longer optimistic and would offer terms that would satisfy D in order to avoid war. This claim also needs two assumptions to work: (i) S must be able to revise her beliefs (D 's rejection must be informative); and (ii) S 's “eve of war” offer must be conditional on these beliefs.

These are also incompatible with F&R's unilateral peace requirement. We have already seen that the presence of uncertainty is irrelevant for the equilibrium outcomes. More specifically, this means that even if S 's behavior is in some way conditional on her beliefs, this way is inconsequential; she might as well ignore what she thinks about D and just offer the status quo on the “eve of war” in every period.

Note now that the “eve of war” concept is itself meaningless in a world where war will never occur. What F&R do not seem to realize is that this also renders their ‘hotline’ useless because the revision of beliefs, the only reason for S to use the ‘hotline’, cannot occur in such a world. This is easily shown in the standard model with two-types where S is no longer optimistic after D rejects her risk-return offer. She is no longer optimistic because D 's rejection was *informative*: it revealed credibly that D is strong, which caused S to revise her beliefs accordingly. The revelation was credible because D would never have rejected her offer if he were weak. This is so because he knows that rejection would lead to war, and his expected war payoff when he is weak is no better than the terms of the risk-return offer. Only a strong D has an expected war payoff that exceeds these terms, and only he would ever reject them. S knows this, which is why she believes what D is telling her implicitly with his rejection: “I am strong.”

Consider now F&R's line of reasoning. Suppose that S were to use the ‘hotline’ upon rejection to offer D war-avoiding terms that, as F&R implicitly assume, D would accept even if he is strong. Doing so would indeed avoid war. F&R take this as evidence that the original model is problematic but instead it reveals the flaw in their argument. Since S will always ensure peace by making this generous offer, rejection no longer means war: it means peace on terms sufficiently attractive to make even the strong D stick with it. Such terms are extraordinarily attractive for the weak D , and he can have them at no risk or cost by simply rejecting S 's initial limited offer and accepting the “eve of war” offer she immediately proffers afterward.

This destroys the signaling property of rejection: since S is aware of these incentives, she would know that whenever she avoids war at the last moment with a generous offer, D would always reject a more limited one even if he is weak. Since *rejection ceases to be informative*, S will not revise her belief when she observes it. Since she would learn nothing from D 's action, her initial optimism will remain intact and she will have no reason at all to make the generous offer. In other words, because S always avoids war, there is no “eve of war”; without such an “eve of war,” D cannot persuade S to reduce her optimism; without such a reduction of optimism, S will not want to use the ‘hotline’. The circle closes and the ‘hotline’ argument falls apart.

The problem of credible information transmission is, of course, a very general phenomenon that has been studied for a long time. When mutual optimism is present and actors cannot find a mutually acceptable peace agreement, they can avoid war only by *overcoming* that optimism. The core insight of the signaling literature in international relations is that one of the principal mechanisms for doing that is for a strong actor to take actions

that generate an irreducible risk of war. He must run a real risk in order to demonstrate his willingness to do so, which in turn reveals that he is not bluffing. In our example, this would imply that D 's rejection can only be informative if it were to cause war with positive probability that S cannot entirely control. That is, *while the avoidance of war is under the collective control of the players, it should not be under individual control.*

As we have argued, the MO explanation requires one to specify the mechanism through which mutual optimism ends in war. The risk-return trade-off is one possibility but the more important one is *credible signaling*: war can occur when actors attempt to reduce each other's optimism through signaling which entails positive risks of war in order to be credible. Although they do not have to, these attempts might well leave actors pessimistic "on the eve of war." Even though mutual optimism might not be present when war breaks out (because the signaling effort was successful), mutual optimism is still the cause of war because it induced the behavior that generated it. For example, escalating audience costs might inadvertently lock actors into war (Fearon, 1996), and the "threat that leaves something to chance" entails deliberate actions that generate risks that neither actor can control (Powell, 1990).

To see how fully F&R rule out signaling as part of any MO explanation, consider their statement that "our assumptions serve to ensure that a profitable deviation from a war equilibrium is present when there is no mutual optimism" (750). But the core insight of credible signaling is that by the time actors overcome mutual optimism, such a foolproof deviation that is guaranteed to lead to peace may no longer be available. We are certainly *not* saying that MO explanations require that when war begins actors should no longer be mutually optimistic. What we *are* saying is that MO explanations do not *require* that they must be, which is precisely what F&R do.¹¹ We have provided examples of mechanisms that can cause war when mutual optimism is present at the outset of the crisis. The fact that they might sometimes make actors fight with regret is irrelevant.

The fundamental problem with F&R's argument is that it actively refuses to consider the process through which actors acquire information in a crisis when mutual optimism prevents them from reaching an agreement. However, the modern rationalist specification of the MO explanation is *precisely* about this process: the forms it can take and the outcomes it can produce.

2 Why War Does Not Occur in the Fey-Ramsay Model

What if the reader does not buy our specification of a rationalist MO explanation with its information transmission mechanisms? Would that mean that F&R's indictment of it stands? We now show that even if one were to adopt F&R's restrictive view of the MO explanation, their model still fails to demonstrate that it is incoherent.

F&R's approach is to analyze a general class of models that are supposedly more appropriate than standard crisis bargaining models, and show that in this class of models there exists no (Bayesian Nash) equilibrium in which war occurs (we shall henceforth refer to this, their main result in Theorem 1, as the "no-war" result.) While their approach permits

¹¹For instance, on p. 751 they clearly state, "... our definition of mutual optimism that, at the instant before the war starts, both sides believe they will prevail."

the analysis of an entire class of models, it is quite abstract and they never offer an example of an actual model that belongs to this class. We construct just such a model and then use it to show how a combination of assumptions that F&R make creates a class of models that cannot be used to evaluate the coherence of the MO explanation in particular and crisis bargaining in general.

2.1 The Basic F&R Model

In F&R’s model, two states, S and D , embroiled in a crisis simultaneously choose from a set of actions. Given F&R’s assumption that the negotiation payoffs are unique in each state of the world, we can reduce the set of actions to two: “stand firm” (F) and “negotiate” (N). Since neither war nor negotiation payoffs can depend on how that outcome is reached, nothing is added by considering more complicated action sets. To capture the notion of war as a mutual act, F&R assume that it can only occur if both actors choose to stand firm, $\langle F, F \rangle$. Otherwise, the outcome is a negotiated settlement where payoffs are identical for the strategy profiles $\langle F, N \rangle$, $\langle N, F \rangle$, and $\langle N, N \rangle$.

Whereas payoffs are not allowed to depend on the crisis behavior of the actors, they can depend on the true state of the world denoted by ω , which can be one from a countably finite set $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$. Actors may obtain private information about the true state of the world before they make their choices. The war payoff in some state ω is specified as in the common costly-lottery winner-take-all model we have been using, $W_i(\omega) = p_i(\omega) - c_i(\omega)$, where $p_i(\omega) \in (0, 1)$ is actor i ’s probability of winning if the true state of the world is ω , and $c_i(\omega) > 0$ are his war costs in that state. The usual assumptions apply: ties are not allowed, $p_S(\omega) + p_D(\omega) = 1$, and war is inefficient: $W_S(\omega) + W_D(\omega) < 1$ for all $\omega \in \Omega$. Negotiated settlements, denoted by $r_i(\omega)$, on the other hand, are efficient: $r_S(\omega) + r_D(\omega) = 1$. (This is just for convenience: the results hold as long as negotiations are less costly than war.) The negotiation payoffs are specified exogenously for each state of the world and cannot vary with the strategies used during the crisis.

		D	
		N	F
S	N	$r_S(\omega), r_D(\omega)$	$r_S(\omega), r_D(\omega)$
	F	$r_S(\omega), r_D(\omega)$	$W_S(\omega), W_D(\omega)$

Figure 1: F&R’s basic model.

Figure 1 shows F&R’s basic model. Private information about ω obeys standard rationality postulates (e.g., a player cannot exclude the true state of the world from the set of states he believes possible; see Osborne and Rubinstein (1994, Ch. 5) for formal definitions). F&R’s “no-war” theorem establishes that $\langle F, F \rangle$ can never be a Bayesian-Nash equilibrium. Because this result applies to all games in this class, they conclude that “our result that war cannot occur in equilibrium implies that mutual optimism is not a valid rationalist explanation for war... Our result shows that one prominent explanation, war by mutual optimism, is not a coherent and internally consistent theory of war within the rationalist framework” (752).

How does this inference work? Let us grant for a moment that this model is ideally suited

to examine the MO explanation, as they claim.¹² F&R insist, correctly, that if MO is to make any sense as an explanation, *war should not occur if actors have complete information, only when they have private information*. F&R reason that if war does not occur in equilibrium in the latter case under *any* information partitions satisfying the standard game-theoretic postulates, then MO (or private information actually) is not a valid rationalist explanation for war. They further conclude that models, in which war does occur only under incomplete information, admit such a result for artificial structural reasons. We have already dispensed with the last claim, so here we examine the validity of F&R’s line of reasoning.

2.2 War Does Not Occur with Complete Information... Even When It Should

Consider F&R’s model with complete information, i.e., there is only one state of the world, ω_1 . Suppose now that in this state of the world $p_i(\omega_1) = \frac{1}{2}$ and $c_i(\omega_1) = \frac{1}{4}$, so $W_i(\omega_1) = \frac{1}{4}$. Bargaining is efficient and $r_i(\omega_1) = \frac{1}{2}$, so each actor strictly prefers the exogenous negotiated settlement to war. The model is shown in Figure 2(a).

		<i>D</i>	
		<i>N</i>	<i>F</i>
<i>S</i>	<i>N</i>	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}$
	<i>F</i>	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{4}, \frac{1}{4}$

(a) State of the world is ω_1 .

		<i>D</i>	
		<i>N</i>	<i>F</i>
<i>S</i>	<i>N</i>	$\frac{4}{5}, \frac{1}{5}$	$\frac{4}{5}, \frac{1}{5}$
	<i>F</i>	$\frac{4}{5}, \frac{1}{5}$	$\frac{1}{4}, \frac{1}{4}$

(b) State of the world is ω_2

Figure 2: Parameterized Specifications of F&R’s Model.

It is readily verified that only $\langle N, N \rangle$, $\langle N, F \rangle$, and $\langle F, N \rangle$ are Nash equilibria. In particular, $\langle F, F \rangle$ is not an equilibrium, so the “no-war” result holds. The F&R assumptions do not appear problematic.

Consider now an alternative state of the world, ω_2 , where everything is the same as in ω_1 except that $r_S(\omega_2) = \frac{4}{5}$ and $r_D(\omega_2) = \frac{1}{5}$. This model is shown in Figure 2(b). Here, *S* still prefers the negotiated settlement to war but *D* strictly prefers war. It is easily verified that only $\langle N, N \rangle$ and $\langle N, F \rangle$ are Nash equilibria. Again, $\langle F, F \rangle$ is not an equilibrium, and the “no-war” result holds. However, we now have a problem because given the payoffs, *war should occur* in this scenario!

To see this, observe that in all Nash equilibria *D* accepts a peace settlement that gives him a payoff that is strictly worse than his payoff from war. In other words, *D* would rather fight than accept such a meager settlement but cannot do so. Why not? Because if he chooses *F*, *S*’s best response is to choose *N* and avoid the war. F&R refer to this ability innocuously as “inducing the bargaining procedure” but what it really means is an ability to impose peace on a dissatisfied opponent. Therefore, F&R’s model *artificially precludes war through its structural assumptions*, the very sin they accuse the standard model of committing.

¹²For example, on p.739, they write that they are analyzing “a class of games that capture the key features of the mutual optimism argument.”, and on p.750 that they are “formalizing the mutual optimism hypothesis and using assumptions designed to test this hypothesis” and “Our game-theoretic setting is thus designed to create conditions in which there is a clear link between mutual optimism and war.”

Furthermore, in F&R's model, the sin is deadly. In the standard endogenous-offer model, war will *not* occur with these war payoffs under complete information because S would offer D enough to make war unprofitable (in our simple TILI model, she would offer $\frac{1}{4}$, which he will accept). Crucially, war is avoided under complete information precisely because there is no mutual optimism: players know the true state of the world and are able to negotiate the settlement accordingly. Of course, F&R's model also has no mutual optimism (after all, we are dealing with complete information) but the reason war does not occur *has nothing to do with its absence*. In contrast to the standard model where S has to make adjustments to the status quo to get D to agree not to fight, in F&R's model she simply chooses not to fight and preserves the beneficial status quo. This is their *unilateral peace assumption*.

F&R make war a mutual act by assuming that both actors must stand firm. This *might* be a reasonable approach if they allowed the negotiation payoffs to depend on *how* the outcome is reached. For example, we agree that for war to occur, both sides must choose to fight. If one country attacks another, the target could refuse to fight, but the important point is that it can only do so at the price of satisfying the attacker's demands; i.e., by surrendering. The standard endogenous-offers model captures this mutuality of war even though surrendering is not modeled explicitly. Suppose S wished to surrender when D rejects her offer and attacks. In our winner-take-all context, this is equivalent to agreeing that D should have the entire good. Since the payoff to holding no share of it is zero, S would only surrender if war is even worse (e.g., her cost of fighting is high relative to her probability of winning). If she is going to surrender, however, D 's expected payoff from rejection is no longer $p - c_D$ but 1, which implies that he can credibly demand the *entire* good, which S would indeed offer. When we specify the war payoffs for the outcome that follows D 's rejection in the standard model, we are implicitly assuming that each side would rather fight than surrender the entire good (i.e., $p_i - c_i > 0$ for each i).

F&R's model *could* capture this, but that would require that the negotiation payoffs depend on *how* the settlement is reached. It would matter which actor chooses to "stand firm" and which chooses to "negotiate". In the strategy profile $\langle N, F \rangle$, it is S that surrenders and D attacks, and in $\langle F, N \rangle$ it is the reverse. The payoffs associated with these two peace outcomes *should* be different, with actors getting worse payoffs when they surrender relative to when their opponent surrenders to them. The strategy profile $\langle N, N \rangle$ would then represent a more equitable negotiated outcome in which each actor would presumably obtain more than what he would have gotten by surrendering. However, F&R explicitly rule this out. This is their *behavior-independent peace payoffs* assumption. (It might be that F&R's simultaneous-move model is not the best way to capture the interaction when crisis behavior is payoff-relevant—it does not allow actors to reconsider their decisions and react to the opponent's behavior—but this is not essential to our point.)

F&R state (p.750) that "war can be validated as a consequence of mutual optimism only in a model in which an equilibrium with war exists even though *either* side can deviate to negotiate some settlement." The problem is that when one side deviates into negotiations, the "settlement" can make the other side worse off than war. The dissatisfied actor has no say in whether peace can be had on such terms. Whereas war might be a mutual act in the limited sense that both sides must agree to fight for it to occur, peace is *certainly* a mutual act in that both must agree *not* to fight for it to occur. We would argue that mutuality of

peace is an essential requirement for any reasonable model of crisis bargaining or mutual optimism.

The fundamental flaw in F&R's model is caused by the combination of their *unilateral peace* and *behavior-independent peace payoffs* assumptions. Taken together, they allow actors to impose settlement terms on an opponent who would rather fight than accept them. We cannot even call this outcome "peace" because the dissatisfied actor would immediately begin a new crisis. The essential problem of peace is that one's optimism about war makes it very difficult to meet the demands of a dissatisfied opponent when such demands are themselves inflated by optimism as well. This is why *mutual* optimism is such a hard problem to overcome in a crisis.

F&R seem to think that satisfaction and dissatisfaction are outside the purview of the MO explanation—they offer "dissatisfied revisionist states" as an alternative explanation for war (739). We would not even know how to *define* these terms without reference to beliefs about expected war payoffs: after all, an actor is dissatisfied with a peaceful distribution of the benefit when he expects to gain more by fighting, and is satisfied with it when he expects even less from war. Optimism inflates war expectations and shrinks the range of settlements that an actor considers acceptable: when the peace terms fall out of this range, the actor becomes dissatisfied with them. *Satisfaction and dissatisfaction are just labels for the relationship between the peace and war payoffs, they are always relative. If one is going to explain the absence of war, then one must explain how the terms of peace satisfy both actors.*

2.3 The Structural Causes of Peace with Incomplete Information

The complete information analysis reveals the potential problems that the twin structural assumptions of unilateral peace and payoff-irrelevant behavior might pose. We now show that they actually provide the core of F&R's main "no-war" result under incomplete information as well.

Consider a sequential-move variant of F&R's basic model. As they note on p. 746, their result "applies equally to decisions made simultaneously and sequentially," and such a model makes it easier to illustrate our points. Assume that there are only two states of the world, ω_1 and ω_2 , and there is incomplete information about the realization. D knows the true state of the world, but S does not—she believes that it is ω_1 with probability $1 - q$, and ω_2 with probability q . S can negotiate or stand firm. If she negotiates, the game ends with the peace settlement. If she stands firm, D can negotiate or stand firm in turn. If he negotiates, the game ends with the peace settlement. If he stands firm, the game ends in war. Payoffs depend on the state of the world but not on crisis behavior. Figure 3 shows the extensive-form of this game.

This model satisfies all assumptions of F&R's class of models. In particular, (i) war is inefficient, (ii) the negotiation outcomes are efficient, (iii) war only occurs if both sides stand firm, (iv) each player has a strategy that guarantees a negotiated settlement regardless of the other side's strategy, (v) the negotiation payoffs are unique in each state of the world, (vi) there is a common prior, and (vii) each actor's possibility correspondence is partitional. Consequently, the "no-war" theorem holds for this particular model as well, as is established in the following result.

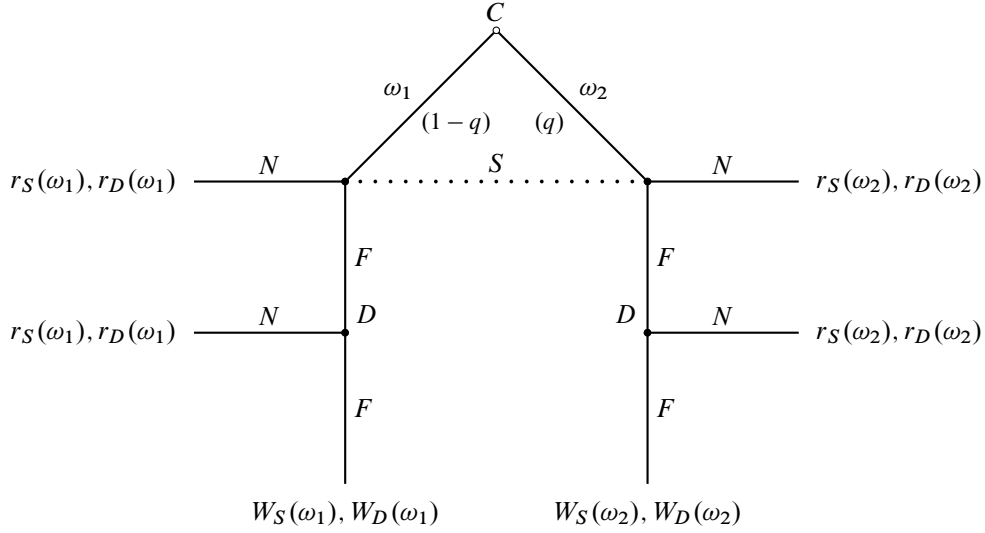


Figure 3: A game that conforms to all F&R assumptions.

PROPOSITION 4. *The game in Figure 3 has no perfect Bayesian equilibrium (PBE) in which war occurs with positive probability.*

The proof is straightforward and proceeds along the lines of the one case we discuss here.¹³ Assume that D strictly prefers the negotiated settlement to war in ω_1 but not in ω_2 . That is, $W_D(\omega_1) < r_D(\omega_1)$ and $W_D(\omega_2) > r_D(\omega_2)$, so D is satisfied in ω_1 and dissatisfied in ω_2 . With these parameter values, D 's equilibrium strategy must be to negotiate in ω_1 and stand firm in ω_2 .

Consider now the optimal strategy for the uninformed player. Because S does not know the true state of the world, from her perspective D is potentially dissatisfied so her expected payoff from standing firm is $U_S(F) = (1 - q)r_S(\omega_1) + qW_S(\omega_2)$. Her expected payoff from negotiations is $U_S(N) = (1 - q)r_S(\omega_1) + qr_S(\omega_2)$. She would strictly prefer to negotiate if, and only if, $U_S(N) > U_S(F)$, which simplifies to $r_S(\omega_2) > W_S(\omega_2)$.

Note that this condition does not depend on her beliefs: it is completely irrelevant whether she is optimistic or not. Her choice is *entirely* determined by her preferences in the state of the world where D is known to be dissatisfied. As we know, our standard assumption of war being costlier than peace allows only one actor to be dissatisfied in any given state of the world. If D is dissatisfied in ω_2 , it must be that S is satisfied in that state of the world. This means that $r_S(\omega_2) > W_S(\omega_2)$, which in turn implies that $U_S(N) > U_S(F)$. Therefore, she will choose N , ending the game immediately with the negotiated settlement.

Because S 's decision only depends on her preferences in ω_2 but not ω_1 , her belief about the true state of the world—and hence any possible optimism—is irrelevant. Unlike the standard model, S does not care about D 's private information. To see why this happens, consider first ω_1 , where D is known to be satisfied. If this is the true state of the world, then standing firm would cause him to negotiate, and S 's payoff would be $r_S(\omega_1)$. If she simply chose to negotiate immediately, her payoff would again be $r_S(\omega_1)$, courtesy of the

¹³The complete proof for all parameter values can be found in the technical appendix.

behavior-independent payoff assumption. This holds for any possible state of the world where D is satisfied: if the true state of the world happens to be among them, peace would occur whether S chooses to stand firm or negotiate, and her payoff would be the same either way.

This now means that from her perspective, the only difference between her expected payoff from standing firm and negotiating would arise from the states of the world, like ω_2 , where D is dissatisfied—if she stands firm and the true state is among them, he would stand firm too and war would occur. To see why her beliefs about the likelihood of different states in which D is dissatisfied are also irrelevant, observe that in any such state she has to be satisfied:

$$\begin{aligned} W_D(\omega_2) \geq r_D(\omega_2) &\Rightarrow W_D(\omega_2) \geq 1 - r_S(\omega_2) && \text{(because } r_D(\omega_2) = 1 - r_S(\omega_2)\text{)} \\ &\Rightarrow r_S(\omega_2) > W_S(\omega_2) && \text{(because } 1 - W_D(\omega_2) > W_S(\omega_2)\text{)}. \end{aligned}$$

Is it easy to see that this holds for *any* state in which D is dissatisfied, so in any such state S 's payoff from war is strictly worse than her payoff from peace. Therefore, for any positive probability that D is dissatisfied, her expected payoff from standing firm is *strictly lower* than her expected payoff from negotiating. Consequently, in any equilibrium she strictly prefers to negotiate, and can obtain a settlement, courtesy of the *unilateral peace assumption*.

Both assumptions are necessary and sufficient for the result. Sufficiency follows immediately from the argument above. To see their necessity, suppose players could not impose peace unilaterally but their peace payoffs were still independent of behavior. War would then occur in any state of the world in which D is dissatisfied: S cannot alter the exogenous terms to make him satisfied and cannot simply impose the terms to avoid fighting. Thus, the “no-war” result cannot obtain without the *unilateral peace assumption*.

To see that this result also requires payoff-irrelevant behavior, suppose players could impose peace unilaterally but their peace payoffs depended on how that peace was obtained. We now show that an *arbitrarily small* difference in these payoffs in just *one* state of the world causes the “no-war” result to break down.

Consider the sequential-move game in Figure 4. It is the same as the game in Figure 3 except that in ω_1 , S gains a slight advantage, $\varepsilon > 0$, if she stands firm and D induces the negotiations compared to the settlement in which she negotiates first. We have in mind an advantage so small that D would still strictly prefer to negotiate in ω_1 : $0 < \varepsilon < r_D(\omega_1) - W_D(\omega_1)$. Suppose, as before, that D strictly prefers war in ω_2 which, of course, implies that S strictly prefers peace there. Compared to the game in Figure 3, we now allow negotiation payoffs to differ in just one state of the world and the difference is so small that both players are still satisfied there. War would not occur with complete information: $\langle F, N \rangle$ is the unique Nash equilibrium under ω_1 , and $\langle N, F \rangle$ and $\langle N, N \rangle$ are Nash equilibria under ω_2 . With incomplete information, however, the result changes dramatically, as the following proposition shows.

PROPOSITION 5. *If $q < k = \frac{\varepsilon}{r_S(\omega_2) - W_S(\omega_2) + \varepsilon} \in (0, 1)$, then the game in Figure 4 has a unique perfect Bayesian equilibrium: S stands firm, and D stands firm if dissatisfied (i.e., in ω_2) and negotiates if satisfied (i.e., in ω_1). War occurs with probability $q > 0$.*

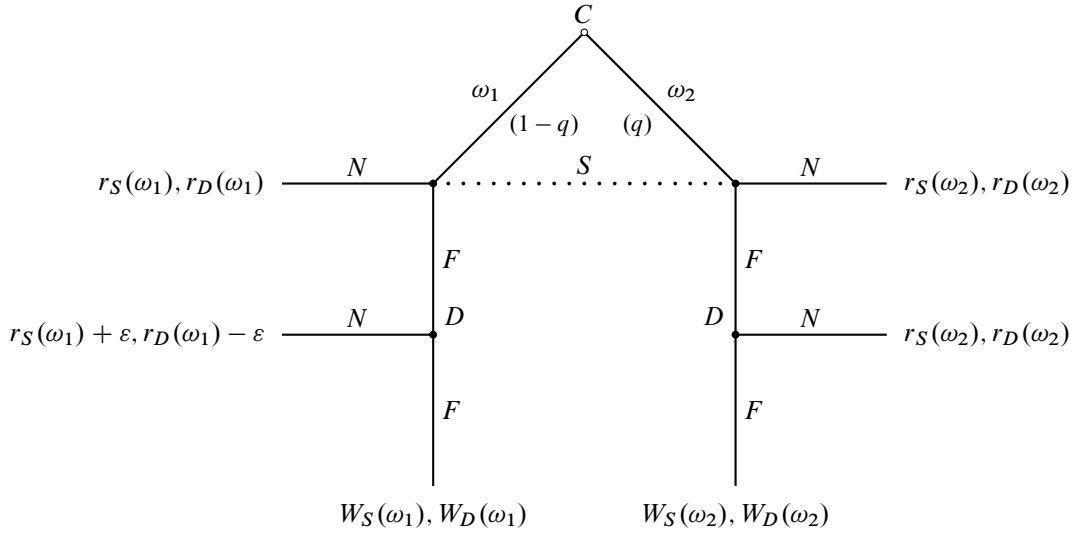


Figure 4: A game that conforms to all F&R assumptions except behavior-independent peace payoffs.

Proof. Given the parameter assumptions, in any PBE D must negotiate in ω_1 and stand firm in ω_2 . Therefore, $U_S(F) = (1 - q)[r_S(\omega_1) + \epsilon] + qW_S(\omega_2)$, and $U_S(N) = (1 - q)r_S(\omega_1) + qr_S(\omega_2)$. She would strictly prefer to stand firm if, and only if, $U_S(F) > U_S(N)$, which simplifies to $q < k$. \square

It is not just that war occurs with positive probability in some equilibrium, it does so in the *only* PBE that exists when players are mutually optimistic. The necessity of mutual optimism follows because war occurs only when both players stand firm: S does so if her belief that D is satisfied is sufficiently high, and D does so only if he is actually dissatisfied (so knows that his war payoff is higher than the available peace terms). If the true state of the world happens to be ω_2 , war would occur even though each player could impose peace unilaterally.

Why does such a minor relaxation of the behavior-independent peace assumption cause a breakdown of the “no-war” result? After all, the negotiation payoffs differ only in a state of the world where D is satisfied and would never stand firm; how can it then lead to war? It can lead to war because now S is no longer indifferent *how* the peace terms are obtained. She strictly prefers D to induce the negotiations, which he would do whenever he is satisfied. If S is very optimistic, she believes that the risk from standing firm is so small that the minor gain from forcing D to negotiate trumps the probability that he would stand firm as well. It is precisely because the gain accrues in a state of the world where war does not occur that she is willing to run that risk. S 's optimism leads her to discount the probability that war will occur when she stands firm. Of course, when war does occur, S has reasons to regret her decision (recall that she is satisfied in ω_2). However, this does not alter the fact that her decision is optimal *ex ante* given her optimistic beliefs. The reader would naturally wonder whether it would be possible to avoid war if she could offer D better terms. This, however, would take us out of F&R's world back to the standard

endogenous-offer models.

We conclude that the absence of war in F&R's model has nothing to do with private information, but is a direct consequence of their two structural assumptions. We believe these assumptions to be indefensible from a substantive perspective. However, even if we were to grant them, F&R's model would remain silent about the validity and coherence of the MO explanation. Since nothing in their result actually depends on beliefs, the "no-war" theorem cannot be used to evaluate an explanation for which beliefs are crucial.

3 Conclusion

When we first learned of F&R's result, we were quite startled. Their mathematical logic is impeccable, and the implications seemed to reach far beyond being just a challenge to the MO explanation. We were at a loss when trying to defend the occurrence of war in our existing models against a general result under such apparently reasonable assumptions. Our discussions with Mark and Kris have been enormously beneficial, and we came to realize that we disagree with them about what mutual optimism is. In doing so, we were forced to define very precisely our understanding of the rationalist MO explanation. As we have shown in this article, the dispute is not a matter of semantics, it is a fundamental disagreement about the essence of crisis bargaining. Our position can be summarized succinctly: *the modern rationalist version of the MO explanation specifies the mechanism that leads from mutual optimism to fighting: war occurs as a result of strategic behavior intended to overcome the inability to reach an agreement caused by this optimism.* In our construction of this argument, we identified one feature that we believe *any* reasonable model of crisis behavior should possess, and another that is desirable of most.

First, in any model of crisis bargaining, *peace should be a mutual act; the avoidance of war cannot be under the individual control of any actor.* The standard models all assume that peace is only under collective control. While it is true that war is also consensual, it is only so in the very limited sense that both sides must agree to fight for it to occur. The standard model implicitly assumes mutuality of war when it assumes that unconditional surrender to the attacker's demands—what it would take to get him to halt his attack—is worse than fighting for the target. The unilateral peace assumption that F&R make is substantively implausible: if we lived in such a world, any state could preserve a favorable status quo by refusing to fight revisionist opponents dissatisfied with its terms. For instance, the U.S. could have won the second phase of the war in Iraq by the simple expediency of declining to fight the insurgents. (Hussein could have won the first phase by not fighting the invasion. . . which he tried to no avail.) Thus, any model used to study crisis behavior must specify every peace outcome in such a way that the expected payoffs for both players are no worse than their expected war payoffs when that outcome occurs. This is just a variant of the venerable assumption of anarchy in international relations theory. In the endogenous-offer model this arises naturally from strategic behavior that ensures that at the time of agreement both sides find the terms acceptable.

Second, in most model of crisis bargaining, *crisis behavior should be payoff-relevant.* The standard endogenous-offers model assumes that behavior can affect the terms of the settlement (through the offers), and the signaling models assume that it can affect the peace payoffs (through audience costs, mobilization costs, or through which actor capitulates first)

and even the war payoffs (through military preparations). The only exception we can think of that would still be reasonable are “burning bridges” models where an actor can irreversibly commit to fighting when attacked. We have our scruples about those as well, but at any rate F&R’s model is not among them.

F&R seem to think that crisis behavior should affect expected payoffs only through changes in beliefs. They write that the “fundamental reason that mutual optimism cannot lead to war is that if both sides are willing to fight, each should infer that they have either underestimated the strength of the opponent or overestimated their own strength. In either case, these inferences lead to a peaceful settlement of the dispute” (738). As we have seen, *that* is not the reason they obtain their “no-war” result, but the claim reveals the fundamental flaw in their approach to crisis bargaining in general. The problem is that these inferences about willingness to fight cannot come for free as they do in their model. There is a price tag attached to any learning that can happen in a crisis, and it usually involves a higher risk of war.

But how can F&R obtain such frictionless learning in their model? By taking the economics literature on efficient exchange (see their references to the “no trade theorem” on p. 740), forcing crisis bargaining to conform to its assumptions, and applying its results. The *Betting on a Die* game that F&R use to provide intuition for the “no-war” theorem illustrates well the assumptions of economic exchange. Each player privately rolls a die and observes the number, and both decide if they want to bet. If both agree to bet, they pay a small cost and the player with the smaller number pays 1 dollar to the winning opponent. If at least one refuses to bet or if the numbers are the same, nobody pays anything (and no costs are incurred without a bet). As is well known, this game does not have an equilibrium in which betting occurs. In economics, where agents have private information about the expected value of a risky asset and have to agree to terms of trade, this is known as the “no trade” theorem. (More generally, it cannot be common knowledge that both players are willing to bet or trade a risky asset.) F&R describe (p.740) their innovation as allowing the “status quo” outcomes (no betting, no trade), not just the “war” outcome (bet, trade), to depend on the state of the world too.

This, however, is not enough to transform the betting/trade environment into one appropriate for studying crisis bargaining or the MO explanation for war. Betting and trade are both voluntary and non-coercive: if some player does not like what her private information tells her, she can simply decline to participate, which would give her the status quo payoff. If the opponent does not like the status quo relative to what he expects to gain from betting or trading, he cannot coerce her to take the bet or trade or compensate him in any way. In such a context, learning can happen through simple statements about one’s willingness to bet or by revealing the terms of trade one is willing to accept by writing them in a contract. Learning is frictionless because nobody has any reason to use the action—betting or writing—to signify anything else than what it does. If the bet is concluded when players shake hands or the trade occurs when they sign the terms of the contract, extending one’s hand and signing one’s name are unambiguously related to the private information players have: there is no incentive to pretend to be stronger than you really are. If anything, in these scenarios players want to pretend exactly the opposite: they want their opponents to believe that they do not want to bet or trade. Unfortunately, that fiction cannot be sustained at the instant when the actual bet or trade must occur: their willingness to engage in it reveals

their true beliefs.

The gap between this environment and crisis bargaining is staggering. If one player extends his hand, the other can politely turn him down, and he cannot whack her on the head until she agrees to terms he believes he would have gotten had she agreed to bet. It does not matter how beneficial these expected terms might be relative to the status quo payoff he obtains when she walks away. If whacking is allowed, extending one's hand is no longer unambiguously related to his private information because it is now a threat: "give me what I ask for or else I will whack you." What now matters is whether he means it because it would determine the terms she would have to offer to prevent him from whacking her. His declaration of willingness can no longer be taken at face value because he now has incentives to pretend to want to whack her if she fails to meet his terms. This necessitates actions designed to dispel such illusions, and so on... it is a whole other ball game.

What F&R have done is create a class of models they call "crisis bargaining models" but that have been shorn of the very features that define this phenomenon. Avoiding betting or trade are unilateral acts and can be had without reference to the payoffs these outcomes yield to the other player. Peace is not. By making their *unilateral peace* and *behavior-independent peace payoffs* assumptions, F&R have imposed the characteristics of economic exchange on crisis bargaining. It is no surprise that they find an analogous result. The problem is that it can tell us nothing about the phenomenon we are studying.

References

- Blainey, Geoffrey. 1988. *The Causes of War*. 3rd edn. New York: The Free Press.
- Fearon, James D. 1995. Rationalist Explanations for War. *International Organization*, **49**(3), 379–414.
- Fearon, James D. 1996. *Bargaining Over Objects that Influence Future Bargaining Power*. Paper presented at the annual meeting of the American Political Science Association, Washington, D.C., August 28-31.
- Fey, Mark, and Ramsay, Kristopher W. 2007. Mutual Optimism and War. *American Journal of Political Science*, **51**(4), 738–754.
- Jervis, Robert. 1988. War and Misperception. *Journal of Interdisciplinary History*, **18**(4), 675–700.
- Johnson, Dominic D. P. 2004. *Overconfidence and War: The Havoc and Glory of Positive Illusions*. Cambridge: Harvard University Press.
- Lebow, Richard Ned. 1981. *Between Peace and War: The Nature of International Crisis*. Baltimore: The Johns Hopkins University Press.
- Leventoğlu, Bahar, and Tarar, Ahmer. 2008. Does Private Information Lead to Delay or War in Crisis Bargaining? *International Studies Quarterly*, **52**(3), 533–553.
- Levy, Jack S. 1983. Misperception and the Causes of War: Theoretical Linkages and Analytical Problems. *World Politics*, **36**(1), 76–99.

- Osborne, Martin J., and Rubinstein, Ariel. 1994. *A Course in Game Theory*. Cambridge: The M.I.T. Press.
- Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. Cambridge: Cambridge University Press.
- Powell, Robert. 1996. Bargaining in the Shadow of Power. *Games and Economic Behavior*, **15**(2), 255–89.
- Powell, Robert. 1999. *In the Shadow of Power*. Princeton: Princeton University Press.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven: Yale University Press.
- Slantchev, Branislav L. 2005. Military Coercion in Interstate Crises. *American Political Science Review*, **99**(4), 533–547.
- Stoessinger, John G. 2005. *Why Nations Go To War*. 9th edn. Belmont: Wadsworth.
- Van Evera, Stephen. 1999. *Causes of War: Power and the Roots of Conflict*. Ithaca: Cornell University Press.
- Wittman, Donald. 1979. How a War Ends: A Rational Model Approach. *The Journal of Conflict Resolution*, **23**(4), 743–63.