Uncertainty in Crisis Bargaining with Multiple Policy Options: A Game-Free Analysis*

Brenton Kenkel† Peter Schram‡

March 25, 2021

Abstract

We examine the relationship between private capabilities and war in a new class of crisis bargaining games. Whereas traditional models consider interactions that end in either war or a peaceful bargain, we assume actors can also engage in low-level, costly policy options that shape final political outcomes in their favor (e.g., sanctions, arming, cyberattacks). Using the tools of Bayesian mechanism design, we analyze the relationship between private information about one's war payoff and the equilibrium outcomes of these flexible-response crisis bargaining games. In contrast with traditional models, we find that private willingness to fight may decrease probability of war or a state's expected utility from settlement. A key factor is whether private signals of wartime strength are positively or negatively correlated with a state's effectiveness at low-level policy responses.

*We thank Andrew Coe, Brad Smith, William Spaniel, and participants at the Vanderbilt Conflict Workshop for helpful comments.

†Assistant Professor, Department of Political Science, Vanderbilt University. `brenton.kenkel@vanderbilt.edu`

‡Assistant Professor, Department of Political Science, Vanderbilt University. `peter.schram@vanderbilt.edu`

"Our traditional approach is either we're at peace or at conflict. And I think that's insufficient to deal with the actors that actually seek to advance their interests while avoiding our strengths."
—General Dunford (2016), Chairman of the U.S. Joint Chiefs of Staff.

The relationship between private information and war is central to contemporary international relations theory (Fearon, 1995; Powell, 1999; Schultz, 1999; Wagner, 2000; Slantchev, 2003b; Meirowitz *et al.*, 2008; Chassang and Padró i Miquel, 2009; Acharya *et al.*, 2015; Gurantz and Hirsch, 2017; Spaniel and Bils, 2018). Theories of how private information affects war are typically developed using crisis bargaining games,[1] in which states engage in a series of negotiations over a divisible asset and each state may go to war as an outside option. These models assume that one or more states receive a private signal about their war payoffs, unknown to their adversaries, that represents their hidden capabilities or willingness to go to war. Across the diverse set of models of crisis bargaining, two monotonicity results consistently emerge. First, when an actor receives a better private signal about its war payoff—i.e., the actor possesses a greater private willingness to go to war or a more robust hidden capability for use in war—negotiations are more likely to end in war. Second, when an actor receives a better private signal about its war payoff, that actor will attain a greater final expected utility. These relationships between private willingness to fight and the outcomes of bargaining are not specific to any particular bargaining game, but instead emerge in the equilibria of *all* crisis bargaining games (Banks, 1990; Fey and Ramsay, 2011).

The crisis bargaining framework has shaped the theoretical study of the causes of war, but traditional models rest on an implausible assumption: that actors either reach a peaceful and efficient bargain, or else engage in full-scale war (or some similar costly engagement). Recent research has relaxed this dichotomy, recognizing that crisis actors in practice face a vast array of policy options between completely efficient peace and totally destructive war. These policy options include implementing sanctions or tariffs (Coe, 2014; McCormack and Pascoe, 2017; Spaniel and Malone, 2019; Joseph, 2020); offering third-party support to rebels (Schultz, 2010); pursuing a wide range of low-level operations sometimes classified as gray zone conflict (Mazarr, 2015b; Gannon *et al.*, 2020), hybrid conflict (Lanoszka, 2016) or hassling (Schram, Forthcoming, 2020); engaging in cyberwarfare (Gartzke and Lindsay, 2015; Baliga *et al.*, 2020); entering into brinkmanship (Powell, 1989, 2015); and arming (Schultz, 2010; Debs and Monteiro, 2014; Gurantz and Hirsch, 2017; Coe and Vaynman, 2020).

Once we allow for flexible policy options in between total peace and total war, the concept of

[1] We use the broad definition of crisis bargaining games from Fey and Ramsay (2011).

a private willingness to go to war becomes more complicated. In standard crisis bargaining models, private information is unidimensional, only affecting the payoffs from total war. In practice, however, what makes a state better at war might also affect its success using these intermediate policy options. For example, if a state has a wide range of privately known cyber-exploits, then the state knows that it could perform well in a conventional war that uses expansive cyberattacks or in a precise cyberattack against a target's infrastructure. As a result, when a state possesses strong private cyber-capabilities, both war and low-level conflict could be better options for the state. Alternatively, many of the policy levers listed above produce different second-order effects. For example, if a leader is privately concerned about losing popular domestic support, then the leader may be more willing to fight a war to create a rally-around-the-flag effect and less willing to implement tariffs or a covert low-level attack. As a result, having a strong private desire to garner domestic support could make war a better option while making low-level conflict or economic warfare worse options. Whether private war payoffs are associated with greater or lower payoffs from alternative policy options is an empirical question—one whose answer varies across cases and contexts. What is important for our purposes is that while these linkages undoubtedly exist, their effects on the outcomes of crisis bargaining have not been systematically examined.

To summarize, consistent with the wide range of observed international crisis behaviors, contemporary international relations research has begun embracing that actors face a wide range of potentially-related policy choices within a crisis. But previous attempts to draw general conclusions about crisis bargaining models have implicitly ruled out this diverse array of policy choices (Banks, 1990; Fey and Ramsay, 2011). We bring these literatures together, using the game-free analysis tools pioneered for traditional crisis bargaining to study how private information affects outcomes in bargaining games with a wider variety of policy options. We ultimately find that the general results for ordinary crisis bargaining models do not generally hold in this more complex, realistic strategic environment. The effect of private signal on equilibrium outcomes is dependent on the direction and magnitude of the relationship between private wartime strength and a state's effectiveness at low-level, flexible policy responses.

This paper re-examines the relationship between private information and war within what we call *flexible-response crisis bargaining games*. These games share the key features of crisis bargaining games—one actor has a private type that influences war payoffs, and actors negotiate or can escalate to war—but also allow for multiple forms of costly conflict short of all-out war. Most importantly for our purposes, we also allow the private type to influence the payoffs of these other forms of conflict as well. We find that previously established results for crisis bargaining models, like the relationship between greater private war payoffs and a greater final likelihood of war, break down in the flexible-response models. Using a game-free analysis along the lines of previous mechanism design research (Banks, 1990; Fey and Ramsay, 2011), we show that these differences with conventional findings are not the result of any particular game form or functional form. Instead, these differences can arise in *any* flexible-response crisis bargaining game, depending on particular assumptions about the strategies available to players short of war.

Flexible-response crisis bargaining games consider two competing players, a challenger and a defender, in a crisis. These games begins with the challenger selecting some level of transgressions (possibly none) where, following Gurantz and Hirsch's (2017) use of the term, the transgression is some act that is beneficial to the challenger but hurts the defender. In response, the defender can enter into a decisive war over the transgression, can allow the transgression to come to fruition, or can "hassle." Hassling, following Schram's (Forthcoming) use of the term, undercuts the transgression through some destructive and non-decisive low-level response. How the defender plays the game is a function of their private type, which influences their war and hassling payoffs. Beyond this, we place no particular structure on the game. Within flexible-response crisis bargaining games, states could bargain, send costly or non-costly signals, make ultimatum offers, walk back or increase their selected hassling levels, or some combination of any of these actions before the game eventually ends.

As a proof of concept to illustrate how adding additional conflict options can undermine the canonical monotonicity results, consider the models in Figures 1 and 2. In both games, Nature moves first and designates D as type $\underline{\theta}$ with probability $\Pr(\underline{\theta}) \in (0, 1)$ and as type $\bar{\theta}$ with probability $1 - \Pr(\underline{\theta})$. D knows their own type, but C does not. The higher type of D has a higher war payoff. Next, C selects whether to transgress ($t = 1$) or not ($t = 0$). Finally, D observes C's choice, and can accept the transgression, go to war over the transgression, or conduct some limited response and hassle.

The flexible-response model in Figure 1 demonstrates a setting where, when D has a better private war payoff, D will go to war less. The equilibrium likelihood of war is decreasing in type with $\underline{\theta}$ D's going to war and $\bar{\theta}$ D's hassling. Why? Within this case, mirroring the cyber-capabilities example above, we assumed that moving from $\underline{\theta}$ to $\bar{\theta}$ not only makes war a better option for D, but also makes hassling a better option for D. More specifically, while war outcomes do improve in θ, hassling outcomes improve at a faster rate in θ. This allows the shift from $\underline{\theta}$ to $\bar{\theta}$ to have the hassling payoffs surpass the wartime payoffs, thereby reducing type $\bar{D}$'s willingness to go to war. It is worth highlighting that if D lacked the hassling option, the equilibrium would follow the standard monotonicity results, as then both types would go to war.

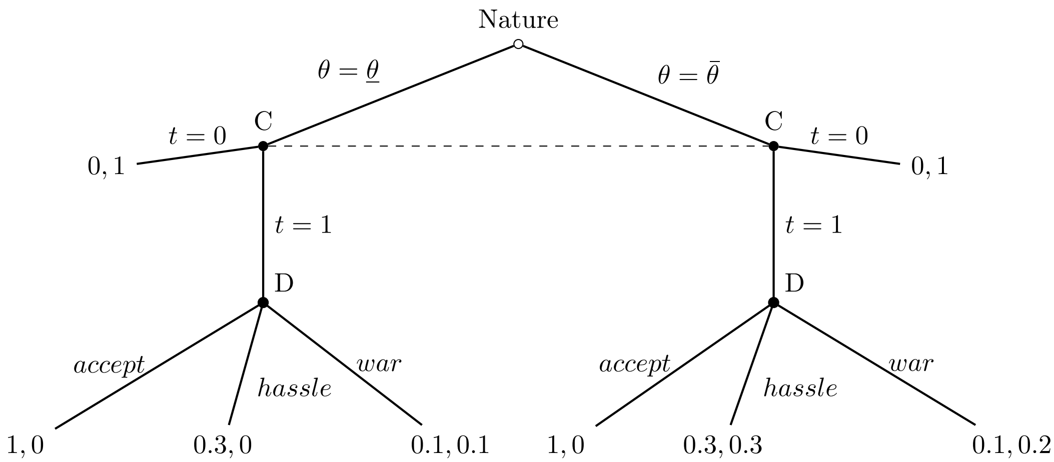

Figure 1: Greater θ implies less war.

C's payoffs are listed first. Here we assume that $\bar{\theta}$ has both greater wartime payoffs and hassling payoffs than $\underline{\theta}$. In equilibrium, C will transgress ($t = 1$), $\underline{\theta}$ D's will go to war and $\bar{\theta}$ D's will hassle.

The model in Figure 2 demonstrates a setting where D's equilibrium payoff is lower when its private war payoff is greater. Why? In this case, as in the rally-round-the-flag example above, moving from $\underline{\theta}$ to $\bar{\theta}$ not only makes war a better option for D, but also makes hassling a *worse* option for D. Specifically, the shift from $\underline{\theta}$ to $\bar{\theta}$ makes hassling much worse while making war only a little better. As a result, D's utility is decreasing in type as D's improvements in wartime payoffs fail to offset the losses from D's hassling payoffs. Again, notice that the standard results would hold if the hassling option were unavailable: the equilibrium would then entail both types going to war, with $\bar{\theta}$ receiving a greater payoff than $\underline{\theta}$.

While simple, these two models illustrate how introducing multiple related conflict options can undermine the monotonicity observed in standard crisis bargaining models. Of course, these models leave much to be desired. While it would be straightforward to create more generalized models than what is above, such an exercise would be inherently limited. A defining feature of the anarchic international order is that there is no clearly defined institution within which states interact and bargain (Waltz, 1979; Axelrod and Keohane, 1985; Niou and Ordeshook, 1990; Wendt, 1992; Mearsheimer *et al.*, 2001)—in other words, there is no clear game form. But any analysis of a game or a cluster of games has the possibility that the results are driven by the game form and would not exist if minor structural changes were made. Thus, to explain the relationship between private information and conflict in

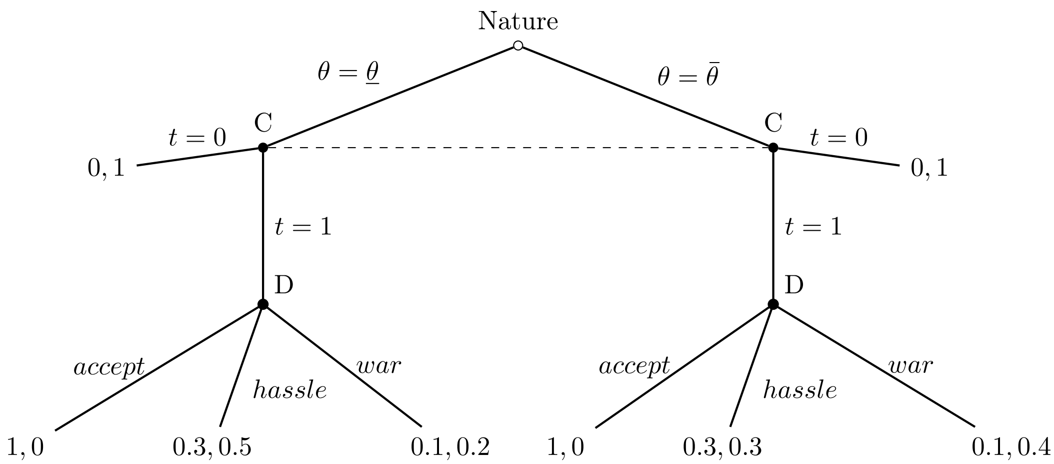

Figure 2: Greater θ implies lower utility

C's payoffs are listed first. Here we assume that $\bar{\theta}$ has greater wartime payoffs and lower hassling payoffs than $\underline{\theta}$. In equilibrium, C will transgress ($t = 1$), $\underline{\theta}$ D's will hassle and $\bar{\theta}$ D's will go to war.

an environment with multiple related conflict options—the central aim of this paper—the game form should ideally play as little role as possible. Our mechanism design approach insures that our results are robust to the broadest set of possible institutional protocols that fall within the flexible-response crisis bargaining framework. This paper is valuable because it can save future scholars the effort of establishing (and re-establishing) the comparative static results we have here on the relationship between private type and outcomes within specific models. Instead, future work can focus on alternate facets of flexible response crisis bargaining models, or on settings where some structure can be imposed.

Our primary contribution is a general characterization of the relationship between private war payoffs and equilibrium outcomes in flexible-response crisis bargaining games. We establish, based on model primitives and (at times) a partial equilibrium analysis, when there is a positive or negative relationship between a greater private willingness to go to war and a greater likelihood of war. At best, this means (for example) that if we know the basic properties of the defender's war payoff function and hassling cost function, we can identify the direction of the relationship between private type and war likelihood for any flexible-response crisis bargaining game. For a broad set of other cases, we can still identify the direction of the relationship, but we may also need to partially characterize equilibrium behavior. Similarly, we establish when there is a positive relationship or a non-monotonic relationship between a greater private willingness to go to war and the defender's utility.

Substantively, this paper is most similar to Schram (2020), which considers multiple conflict options and a publicly observed type that systematically determines conflict utilities. Importantly, while the model in Schram (2020) does include a private type, the private type in Schram (2020) only affects hassling costs rather than affecting both hassling and war costs as is the case here. As a result, all findings here for flexible-response crisis bargaining games that break the Banks (1990) and Fey and Ramsay (2011) monotonicity results are entirely novel.[2] This paper is also similar to a range of research within the crisis bargaining framework or within the deterrence literature that considers actors with multiple possible actions within a crisis (Schultz, 2010; Powell, 2015; McCormack and Pascoe, 2017; Coe, 2018; Spaniel and Malone, 2019; Schram, Forthcoming; Baliga *et al.*, 2020); notably, this is the first paper to systematically examine how the spillover effects of improvements in one kind of conflict capability can affect other conflict capabilities as well. This allows us to offer novel insights into a wide-range of previously unconsidered substantive settings. Methodologically, our paper is similar to a class of work on political science topics that embraces the tools of mechanism design to establish relationships between private information and outcomes. Outside of sources listed above, mechanism design has been featured in work on crisis bargaining and arbitration (Fey and Ramsay, 2009; Hörner *et al.*, 2015; Fey and Kenkel, 2019), bureaucracies and delegation (Ashworth and Sasso, 2019), firm regulation (Baron and Besanko, 1987, 1992), legislation and policy-making (Meirowitz *et al.*, 2006), voting (Aghion and Jackson, 2016), and many others.

1 What do Flexible-Response Crisis Bargaining Models Describe?

In 2006, Israel discovered that Syria was building a nuclear reactor. This discovery was monumental, as, internally, Israeli decision-makers viewed the possibility of a nuclear-armed Syria as an "existential threat" to the Israeli state (Opall-Rome, 2018). In the context of the Banks (1990) and Fey and Ramsay (2011) monotonicity results—where a high private willingness to go to war over an issue leads to a greater likelihood of war—Israel's predicament was the exact setting where we might expect war to arise. But, instead of going to war, Israel used an electronic warfare attack to disable Syrian air defenses and conducted an airstrike on the reactor (Katz, 2010). Those actions, which together are known as Operation Outside the Box, successfully destroyed a critical component of the Syrian nuclear program.

[2] While Schram (2020) also conducts a game-form free analysis of private type, this analysis finds that a greater private willingness to hassle (or lower private costs to hassling) always results in weakly greater utilities. This is different from the U-shaped relationship between private type and utilities that can arise here (see Lemma 9).

Flexible response crisis bargaining models can model political crises like those surrounding the Syrian reactor development. These models formalize the following interaction. First, a challenger—State C—undertakes some opportunistic and costly action that we will refer to as a "transgression." Transgressions are politically beneficial to State C but are detrimental to State D. For example, when Syria was building a nuclear reactor in 2008, this could have eventually led to Syria possessing a nuclear bomb, which could have strengthen Syria's bargaining leverage in the future. Transgressions like this have been proposed in scholarship considering enforcement problems in bargaining (Schultz, 2010), deterrence (Gurantz and Hirsch, 2017), or endogenous power shifts (Debs and Monteiro, 2014).[3] As examples, transgressions could be investing in conventional, nuclear, space, or cyber military technologies (Debs and Monteiro, 2014; Gartzke and Lindsay, 2017; Spaniel, 2019), forming alliances (Benson and Smith, 2020), securing geopolitically valuable territory (Fearon, 1996; Powell, 2006), or engaging in economic warfare or containment to weaken a target (McCormack and Pascoe, 2017; Joseph, 2020).

In response to the transgression, a defender—State D—has the choice between accepting the transgression, going to war to decisively resolve the political issues between the two states, or in engaging in some low-level actions that undercut the future impact of the transgressions. The first two options (acceptance and war) are standard in the crisis bargaining or deterrence frameworks. We will refer to the the last option (D's low-level response) as "hassling." As originally defined in Schram (Forthcoming), hassling is the use of limited conflict to degrade a challenger's rise. Our use of the term here is consistent with this definition, but also refers to any actions by the defender that undercuts the challenger's transgression. So long that D did not previously go to war, C and D then engage in a bargaining protocol that can end with a diplomatic resolution or war. In the Syria case, Israel detected Syria's nuclear reactor and destroyed it. This would be consistent with hassling because it was a destructive blow to Syria's nuclear program, but it was not a decisive military move that would prevent the Assad regime from every possessing a nuclear weapon.[4] Hassling can take the form of limited airstrikes, special operations, cyberattacks, supporting domestic (to State C) insurgent groups, or other military activities. Of course, while the Schram (Forthcoming) definition of hassling is limited to military activity, here it can apply to

[3]The "transgression" here is similar to how a large literature treats the challenger's first move in a deterrence game. See Huth (1999) for a review of early work on deterrence, as well as Fearon (1997); Kydd and McManus (2017); Baliga *et al.* (2020); Smith (1998); Di Lonardo and Tyson (2017); Chassang and Miquel (2010); Baliga and Sjöström (2020).

[4]To offer an example of a decisive military move, the 2003 U.S. invasion of Iraq decisively prevented the Ba'athist regime from attaining nuclear weapons by overthrowing the Ba'athist regime. This invasion is consistent with this model's treatment of war.

anything that can undermine the future political impact of C's transgressions, for example sanctions (McCormack and Pascoe, 2017).

Within these interactions, we assume that D possesses a private type that influences both war and hassling capabilities. In other words, if state D is (privately) very capable at conducting war or very willing to conduct war, we might also expect that state D is systematically better (or worse) at conducting hassling. To offer perhaps the simplest example, suppose a state is privately very hawkish on the matter of its neighbors developing nuclear weapons. Based on these private preferences, this state might be very willing to conduct a war and much less willing to issue sanctions or tariffs to weaken a state. Ultimately, whether better private capabilities or a greater willingness to go to war also makes low-level options more or less appealing is ultimately an empirical question; we discuss this later in subsection 4.4.

2 The Flexible-Response Crisis Bargaining Model

Here we present the flexible-response crisis bargaining framework, in which negotiations may end in war or in one of a continuum of possible inefficient non-war outcomes. We assume a state's private information may affect its payoffs from both kinds of outcome, and we use the tools of Bayesian mechanism design to obtain general results about the relationship between private types and the equilibrium properties of flexible-response crisis bargaining games.

2.1 Structure of the Interaction

At the outset of the interaction, Nature assigns D's type, $\theta \in \mathbb{R}$. Without loss of generality, we assume D's war payoff increases with D's type,[5] while D's cost of hassling may increase, decrease, or neither. The realized value of θ is known only to D, but the prior distribution from which it is drawn is common knowledge. Let F denote the CDF of this prior distribution, and let Θ denote its support.

The interaction between the states takes the familiar form of a crisis bargaining game, except each state may engage in activity that affects outcomes short of war. First, C selects transgression $t \in \mathcal{T} \subseteq \mathbb{R}_+$. Following this choices, C and D partake in a bargaining process that may end in war or in some hassling response by D. Like Banks (1990) and Fey and Ramsay (2011), we place no particular structure on the bargaining process. We simply

[5] In terms of the traditional "costly lottery" formulation of crisis bargaining games, greater θ may correspond to a greater probability of winning, a lower cost of fighting, or both. However, we do not impose a costly lottery model—all that matters for our purposes is that D's expected utility from war strictly increases with θ.

assume that each player chooses from a set of available bargaining actions; these choices determine whether the game ends in war, and, if not, how the prize is divided. In the negotiation stage, we let $b_C \in \mathcal{B}_C$ denote C's bargaining strategy (offers, counteroffers, accept-reject plans, etc.). D's strategy consists of an analogous bargaining strategy $b_D \in \mathcal{B}_D$, as well as a level of hassling, $h \in \mathcal{H} \subseteq \mathbb{R}_+$.[6] A *game form* G consists of the bargaining action spaces, $\mathcal{B}_A$ and $\mathcal{B}_D$, along with an *outcome function* g mapping from the choices (t, h, b_C, b_D) into the set of possible crisis bargaining outcomes.[7] We decompose the outcome function g into three components: whether war occurs, what C receives from bargaining, and what D receives from bargaining.[8] Whether war occurs or not depends solely on actions taken in bargaining. Let $\pi^g(b_C, b_D) \in \{0, 1\}$ be an indicator for whether the interaction ends without war.[9] Conditional on war not occurring, each player's payoff depends on the bargaining behavior, C's choice of transgression, and D's selection of hassling. Let $V_C^g(t, h, b_C, b_D)$ and $V_D^g(t, h, b_C, b_D)$ denote the benefits that C and D receive, respectively, in case war is avoided.

Payoffs depend on the outcome of bargaining, including the costs of the hassling or transgression when war does not occur.[10] War payoffs may depend on D's private information, but do not depend on any of the endogenous choices in the game, including transgressions and hassling.[11] We therefore write war payoffs as $W_A(\theta)$ and $W_D(\theta)$. We assume W_D is strictly increasing, so higher types of D can be interpreted as stronger in wartime. If war is avoided, each player receives their division of the spoils but must pay the cost of their transgression or hassling. Let $K_C(t, h)$ denote the cost to A, and let $K_D(h, \theta)$ denote the cost to D. We assume that K_C is strictly increasing in t and weakly decreasing in h, and we assume K_D is strictly increasing in h. We let $h = 0$ denote no hassling, which entails assuming $0 \in \mathcal{H}$ and $K_D(0, \theta) = 0$ for all θ. Putting these together, the players' utility functions in a given

[6] We place no restriction on whether hassling is chosen before, during, or after the bargaining process—all that matters for our purposes is that the cost of any given level $h \in \mathcal{H}$ is independent of b_C and b_D.

[7] The game form represents the elements of the model that are specific to a particular bargaining protocol. Implicitly, then, we take the type space, prior distribution, transgression action set, hassling action set, cost functions, and war payoff functions as primitives of the model rather than features of a game form G.

[8] Unlike in Banks (1990) and Fey and Ramsay (2011), we allow for inefficient settlements. This means D's value from bargaining cannot be immediately deduced from C's, and vice versa.

[9] By ruling out $\pi^g \in (0, 1)$, we are implicitly assuming the bargaining process has no exogenous random components (see Fey and Kenkel, 2019).

[10] We restrict attention to models in which the players' payoffs from any peaceful outcome do not depend on D's private information, except insofar as that private information affects the cost to D of the chosen hassling level h.

[11] We relax this in an extension below, in which transgressions and hassling may affect war payoffs.

game form are as follows:

$$u^g_D(t, h, b_C, b_D \mid \theta) = (1 - \pi^g(b_C, b_D))W_D(\theta) + \pi^g(b_C, b_D)[V^g_C(t, h, b_C, b_D) - K_C(t, h)], \quad (1)$$
$$u^g_D(t, h, b_C, b_D \mid \theta) = (1 - \pi^g(b_C, b_D))W_D(\theta) + \pi^g(b_C, b_D)[V^g_D(t, h, b_C, b_D) - K_D(h, \theta)]. \quad (2)$$

We restrict our attention to game forms in which neither player can force a settlement on the other. This assumption reflects the anarchic nature of international politics, in which states always have the option to resort to force if desired. A sufficient condition is that each player has an action $b_i \in \mathcal{B}_i$ such that $\pi^g(b_i, b_j) = 0$ (war is guaranteed) for all $b_j \in \mathcal{B}_j$. As we show below, this condition places important limits on what kinds of outcomes are sustainable as equilibria.

2.2 Solution Concept and Direct Mechanisms

We restrict attention to pure strategy perfect Bayesian equilibria of each flexible-response crisis bargaining game. Depending on the bargaining protocol and the equilibrium selected, the equilibrium path may be very complex, involving numerous offers and counteroffers before concluding, or it may be simple, ending quickly in war or a settlement. We will not dwell on the details of bargaining itself, as our primary concern is the *outcome* of the interaction: whether war prevails, and if not, what each party receives from a bargained outcome. Each player's bargaining behavior only affects their payoffs insofar as it affects these components of the outcome.

We will focus on the incentives of D, the player with private information. Given an equilibrium of a flexible-response crisis bargaining game, we can summarize the outcome of the game for each type of D with three functions:[12]

- Their hassling level, $h(\theta)$.
- Whether a bargained outcome prevails, $\pi(\theta)$.
- Their settlement value in case of a bargained resolution, $V_D(\theta)$.

A *direct mechanism* for D consists of these functions, (h, π, V_D). If type θ of D were to follow

[12] In the Appendix, we formally define an equilibrium and describe how a direct mechanism can be derived from it. See the discussion in Fey and Ramsay (2011).

the equilibrium bargaining strategy of type θ', D's expected utility from doing so would be:

$$\Phi_D(\theta' \mid \theta) = \underbrace{(1-\pi(\theta'))W_D(\theta)}_{\text{war}} + \underbrace{\pi(\theta')[V_D(\theta') - K_D(h(\theta'),\theta)]}_{\text{bargained outcome}}.$$

While mimicking another type's strategy may change the hassling level, the occurrence of conflict, and the settlement value for θ, it does not change D's war payoff, nor the cost D pays for any given hassling level.[13] In equilibrium, no type may have a strict incentive to mimic another type's bargaining strategy. We can phrase this requirement as an *incentive compatibility* condition on the direct mechanism. Let $U_D(\theta)$ denote each type's expected utility along the path of play, so that $U_D(\theta) = \Phi_D(\theta \mid \theta)$.

Definition 1. A direct mechanism (h, π, V_D) is *incentive compatible* if

$$U_D(\theta) \geq \Phi_D(\theta' \mid \theta) \qquad \text{for all } \theta, \theta' \in \Theta. \tag{IC}$$

To identify regularities in the outcomes of flexible-response crisis bargaining games, we will analyze incentive compatible direct mechanisms. We rely on the *revelation principle*: for any Bayesian Nash equilibrium of a particular game form, there is an incentive compatible direct mechanism that yields the same outcome (Myerson, 1979). Logically, this means that if we find that some property holds for all incentive compatible direct mechanisms, than it is true of all equilibria of all flexible-response crisis bargaining games. Without bogging ourselves down in the particulars of how crisis bargaining plays out in any particular game, we are still able to characterize robust properties of the outcomes of any flexible-response crisis bargaining game.

Recall that we only consider game forms in which neither player can impose a settlement on the other. This condition ensures that no type of D may receive less than its war payoff in equilibrium—if a settlement would yield less, then it would be profitable for D to deviate to fighting. This requirement amounts to a participation constraint in the language of mechanism design, or what Fey and Ramsay (2011) call voluntary agreements in the crisis bargaining context.

[13] The definition of Φ_D illustrates an important difference between the flexible-response framework and the environment studied by Fey and Ramsay (2009), who also allow for inefficient bargained settlements. In their model, the inefficiency loss due to reporting θ' is the same for all types θ. By contrast, we assume the cost of hassling is a function of the true type θ, which implies that different types may value the "same" settlement differently.

Definition 2. A direct mechanism (h, π, V_D) has *voluntary agreements* if

$$\pi(\theta)[V_D(\theta) - K_D(h(\theta), \theta)] \geq \pi(\theta)W_D(\theta) \qquad \text{for all } \theta \in \Theta. \tag{VA}$$

Naturally, the voluntary agreements condition is automatically satisfied for those types that fight on the path of play. The constraint only applies to the types that settle—the settlement must yield at least as much as their war payoff, even when accounting for the costs of the hassling. Throughout the analysis, we will restrict attention to direct mechanisms that satisfy both (IC) and (VA), as any equilibrium of a flexible-response crisis bargaining game with voluntary agreements must be outcome-equivalent to some such mechanism (Fey and Ramsay, 2011).

2.3 Example Game with Direct Mechanism

To offer a more grounded example of why analysis using direct mechanisms is so valuable, consider the game in figure 1 presented earlier. In this model, in equilibrium, when D was assigned a greater private war payoff, D went to war less. This was shown to hold for a single set of parameters and for a single game form, so natural questions remain. How robust is this equilibrium? What underlying assumptions are needed for such an equilibrium to exist?

To start addressing these questions, consider the game in figure 3. This game is a semi-parameterized version of the games in figures 1 and 2. To offer some explanation, the war outcome here is designed to look like the standard "war as a costly lottery" treatment of war, where there is uncertainty over the costs of war as captured in $\theta \in \{\underline{\theta}, \bar{\theta}\}$, and β and α are constants. For the hassling outcomes, hassling pushes the political settlement distance $1 - q$ while invoking costs $-k_C$ for C and $\theta \in \{\underline{\theta}, \bar{\theta}\}$ for D.

To understand the properties that produce equilibria where greater private types (i.e. D's with greater wartime payoffs) go to war less, we could solve for all possible equilibria in the game in figure 3 and analyze their properties. Alternatively, instead, we can use the revelation principle and conduct an analysis of these equilibria through direct mechanisms. To do so, we first posit the existence of an equilibrium (or equilibria) that has the properties that we are interested in. Specifically, these are equilibria where low-type D's go to war and high-type D's hassle, and equilibria where low-type D's go to war and high-type D's accept. We then construct equivalent direct mechanisms for these equilibria. The direct mechanisms are depicted in figures 4 and 5.

In figure 4, the direct mechanism represents an equilibrium where low-type D's go to war

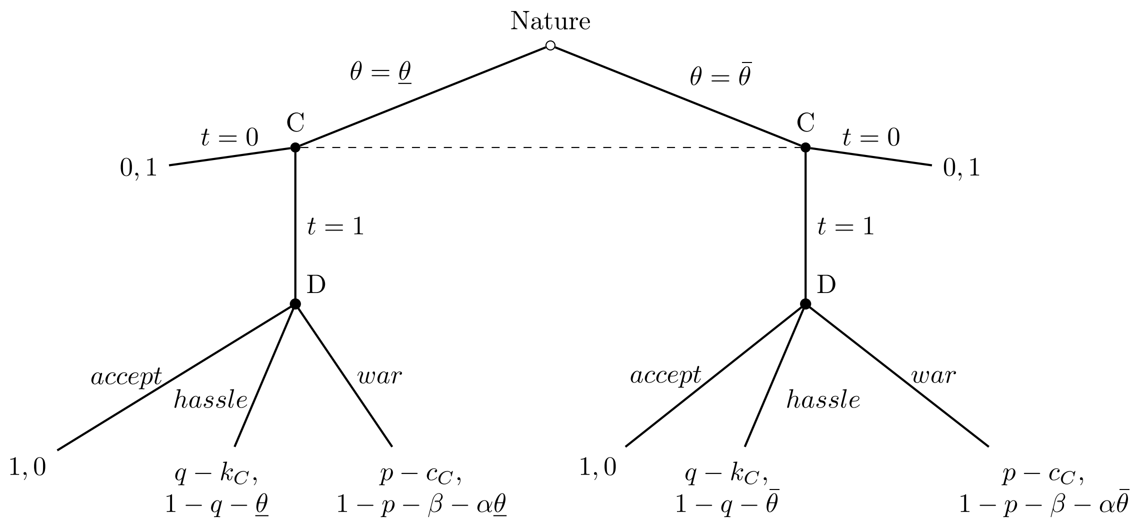

Figure 3: A semi-parameterized flexible response crisis bargaining model.

$\underline{\theta}$ is the "low" type, and $\bar{\theta}$ is the "high" type, and we assume throughout that low-types do worse in war than high types. If $\underline{\theta} > \bar{\theta}$ and $\alpha > 0$, then private type improves both hassling and war payoffs. If $\underline{\theta} < \bar{\theta}$ and $\alpha < 0$, then private type improves war payoffs and increases hassling costs. Letting $q = 0.4$, $k_C = 0.1$, $p = 0.3$, $c_C = 0.2$, $\underline{\theta} = 0.6$, $\bar{\theta} = 0.3$, $\alpha = \frac{1}{3}$, and $\beta = 0.4$ re-creates Figure 1. Letting $q = 0.4$, $k_C = 0.1$, $p = 0.3$, $c_C = 0.2$, $\underline{\theta} = 0.1$, $\bar{\theta} = 0.3$, $\alpha = -1$, and $\beta = 0.6$ re-creates Figure 2.

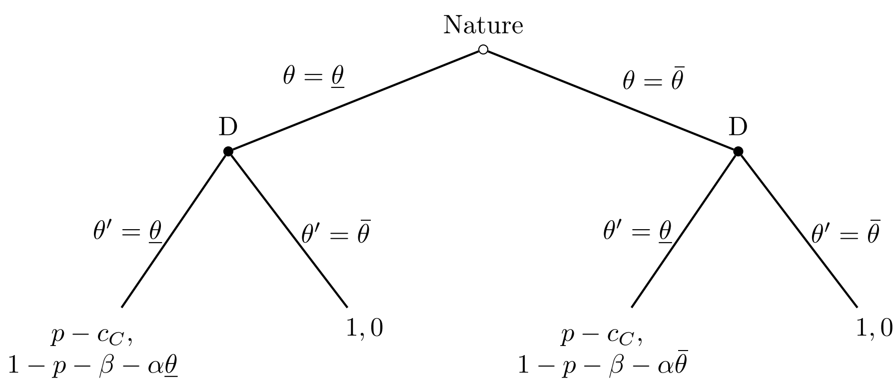

Figure 4: incentive compatible direct mechanism where type $\underline{\theta}$ D's go to war and type $\bar{\theta}$ D's accept.

C's payoffs are listed first. θ denotes D's true type (assigned by nature) and θ' denotes D's reported type (reported by D). For an equilibrium where type $\underline{\theta}$ D's go to war and type $\bar{\theta}$ D's hassle to exist, D must be willing to truthfully report their type ($\theta' = \theta$).

and high-type D's accept. After nature sets D's type $\theta \in \{\underline{\theta}, \bar{\theta}\}$, D's action is reporting their type $\theta' \in \{\underline{\theta}, \bar{\theta}\}$, where a report of $\theta' = \underline{\theta}$ results in each (true) type D's "war" payoff, and a report of $\theta' = \bar{\theta}$ results in each (true) type D's "accept" payoff. The direct mechanism in

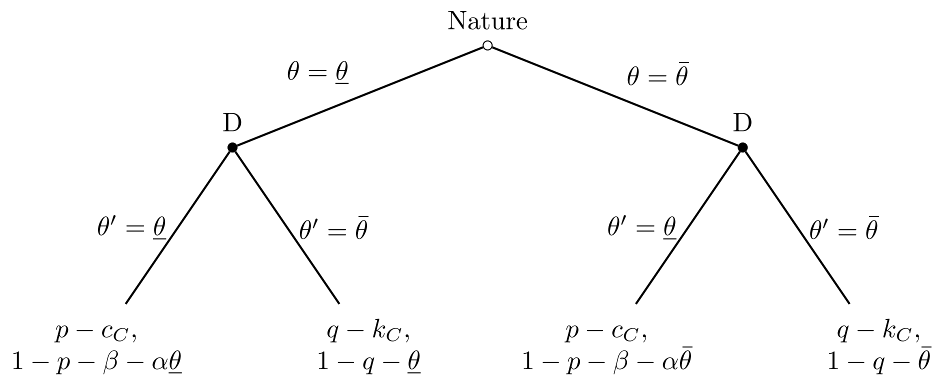

Figure 5: incentive compatible direct mechanism where type $\underline{\theta}$ D's go to war and type $\bar{\theta}$ D's hassle.

C's payoffs are listed first. θ denotes D's true type (assigned by nature) and θ' denotes D's reported type (reported by D). For an equilibrium where type $\underline{\theta}$ D's go to war and type $\bar{\theta}$ D's hassle to exist, D must be willing to truthfully report their type ($\theta' = \theta$).

figure 5 is similar, but it represents an equilibrium where high-type D's hassle, and a report of $\theta' = \bar{\theta}$ results in the "hassle" payoffs.

We first seek to understand the conditions under which an equilibrium exists where low-type D's go to war and high-type D's accept in the model in figure 3. The revelation principle implies that we can accomplish this by deriving the conditions for incentive compatibility within the direct mechanism in figure 4—in other words, the conditions for truthful revelation of types. For truthful revelation in figure 4, it must be that $0 \leq 1 - p - \beta - \alpha\underline{\theta}$ and $0 \geq 1 - p - \beta - \alpha\bar{\theta}$, or combined, $\alpha\bar{\theta} \geq \alpha\bar{\theta}$. But, this cannot hold, or else low-types ($\underline{\theta}$) would be better at conducting war. Thus, there is no incentive compatible direct mechanism within the scope of our analysis that allows for low-types to go to war and high-types to accept; by the revelation principle, this class of equilibria cannot exist in the game in figure 3.[14]

Second, we turn to the conditions under which an equilibrium exists where low-type D's go to war and high-type D's hassle in figure 3. To do so, we turn to the incentive-compatibility conditions in figure 5. For truthful revelation, it must be that $1 - q - \underline{\theta} \leq 1 - p - \beta - \alpha\underline{\theta}$ and $1 - q - \bar{\theta} \geq 1 - p - \beta - \alpha\bar{\theta}$, or when $(1 - \alpha)(\underline{\theta} - \bar{\theta}) > 0$. It is straightforward to see that whenever D's private type improves war payoffs and decreases hassling payoffs ($\underline{\theta} < \bar{\theta}$ and $\alpha < 0$), then this condition will never hold. Whenever D's private type improves war payoffs and increases hassling payoffs ($\underline{\theta} > \bar{\theta}$ and $\alpha > 0$), then this condition will hold so long that

[14]This result follows from Lemma 1 in Banks (1990).

$\alpha < 1$. Thus, whenever D's private type improves war payoffs and increases hassling payoffs and $0 < \alpha < 1$, then an equilibrium where low-types to go to war and high-types hassle can exist.

What was presented above with the direct mechanisms may be perceived as superfluous; after all, the game in 3 is straightforward enough to solve. However, consider a different game. Suppose, following C's initial move $t \in \{0, 1\}$, C and D simultaneously choose to accept, hassle, or go to war, where the greatest level of conflict selected leads to that outcome and payoff?[15] Or, what if escalation occurred sequentially, where, following C's selection of t, D could accept (terminating the game) or "escalate," and then following the escalation move C and D could simultaneously choose to hassle or go to war? Each of these games could individually be solved for. But, for each of these games—or for any other similar game[16]—the direct mechanism analysis of the equilibrium where low-type D's go to war and high-type D's do not *is the same*. In other words, the direct mechanism analysis in figures 4 and 5 allows us to identify the conditions for the equilibrium that we are interested in without needing to solve for the equilibria for each of these games.

This is particularly valuable when we move into a broader class of hassling games. The flexible-response crisis bargaining framework is quite general. Our framework allows for a continuum of types of D, a continuum of transgression options, a continuum of hassling options, and a broad set of possible bargaining moves that can happen before, after, or simultaneously with hassling and that can be used to screen or signal. For the task at hand—understanding the relationship between private information and outcomes—we can reduce a vast set of complex, indirect analyses of game forms and equilibria into a streamlined analysis of incentive compatible direct mechanism.

3 Game-Free Analysis Results

3.1 Connections to Earlier Work

The strategic environment we model is a generalization of the ordinary crisis bargaining environment studied in earlier work. If the options to transgress and to hassle were unavailable, then we would have an ordinary crisis bargaining game. In fact, we can go further—if these options are not exercised in equilibrium, then the equilibrium has the properties of an

[15] In other words, if C selects "accept" and D selects "war," war will occur.

[16] Similar games would be any game where D has a dichotomous private type, C selects $t \in \{0, 1\}$ first, and then C and D have a series of moves where the game can end in "accept," "war," or "hassling" with utilities as listed above.

equilibrium of an ordinary crisis bargaining game.

The headline monotonicity results from ordinary crisis bargaining games hold up in our environment if all types of D forego hassling. In this case, stronger types never have a lower probability of conflict than weaker types. The following result recovers Lemma 1 of Banks (1990) as a special case in our environment.[17]

Lemma 1. *If* $h = 0$ *and* $\theta < \theta'$, *then* $\pi(\theta) \geq \pi(\theta')$.

All proofs are in the Appendix. We also recover the monotonicity of utility in type, stated in Lemma 4 of Banks (1990), as a special case.

Lemma 2. *If* $h = 0$ *and* $\theta < \theta'$, *then* $U_D(\theta) \leq U_D(\theta')$.

Together these results confirm a key mechanism behind our results: the exceptions we find to the monotonicity of conflict probability and expected utility in D's private type are due exclusively to the availability and use of the hassling option. No other feature of the strategic environment is driving the difference in results.

3.2 General Patterns

Here we identify some general properties of D's private information, θ, when a non-war outcome is guaranteed. While higher types of D have greater war payoffs, we do not in general assume any particular relationship between θ and the cost or effectiveness of D's hassling. In order to compare equilibrium outcomes across types, we will place some additional structure on the relationship between D's type and D's cost of h. We will say θ' has *greater hassling effectiveness* than θ if $K_D(h, \theta') < K_D(h, \theta)$ for all $h \in \mathcal{H} \setminus \{0\}$. Depending on the strategic environment, hassling effectiveness may increase, decrease, or vary non-monotonically with D's type.

Our first general result is that if two types both reach a settlement in equilibrium, then the one with greater hassling effectiveness must end up no worse off than the other. All of our formal results, including the following lemma, apply only to incentive compatible direct mechanisms with voluntary agreements—those that satisfy (IC) and (VA). By the revelation

[17]Though we restrict to deterministic outcomes and thus $\pi \in \{0, 1\}$ in the bulk of our analysis, the proof of the following result holds even if we allow for probabilistic outcomes.

principle, this means that the properties identified in our results hold for *all* pure strategy equilibria of *all* flexible-response crisis bargaining games.

Lemma 3. *If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $U_D(\theta) \leq U_D(\theta')$.*

A boost in the private component of D's hassling capabilities therefore cannot make a player worse off unless war occurs. In fact, if D makes a nontrivial investment in the hassling activities on the path of play, we can strengthen the result: types with greater hassling effectiveness have *strictly* greater expected utilities in equilibrium.

Corollary 1. *If $\pi(\theta) = \pi(\theta') = 1$, $h(\theta) > 0$, and θ' has greater hassling effectiveness than θ, then $U_D(\theta) < U_D(\theta')$.*

Because hassling activities are costly, they must produce some benefit at the negotiating table. If D could get the same or better settlement while spending less on hassling, it would be strictly profitable to do so. Consequently, among the types that end up settling on the path of play, greater h must be associated with a more generous settlement.

Lemma 4. *If $\pi(\theta) = \pi(\theta') = 1$ and $h(\theta) \leq h(\theta')$, then $V_D(\theta) \leq V_D(\theta')$. Furthermore, if $h(\theta) < h(\theta')$, then $V_D(\theta) < V_D(\theta')$.*

Next, we examine how private types affect the equilibrium choice of hassling activities. When comparing two types of D, one might naturally suspect that the one with greater hassling effectiveness will choose a greater value of h. In order to guarantee this, we need to impose an additional (but reasonable) condition on the cost function for D's hassling. We will assume the following condition to ensure the function has the single-crossing property, as is common in formal analyses with monotone comparative statics (Ashworth and Bueno de Mesquita, 2006).

Definition 3. The cost function K_D has *decreasing differences* in h and θ if

$$\begin{aligned} &\theta' \text{ has greater hassling effectiveness than } \theta \Rightarrow \\ &\quad K_D(h', \theta') - K_D(h, \theta') < K_D(h', \theta) - K_D(h, \theta) \text{ for all } h < h'. \end{aligned} \tag{DD}$$

Substantively, this condition can be interpreted as more effective types having a lower marginal cost of hassling. For example, the ratio cost function $K_D(h,\theta) = \frac{h}{\theta}$ satisfies (DD) (with higher types having greater hassling effectiveness). More broadly, so does any function of the form $K_D(h,\theta) = \psi(h)\xi(\theta)$, where ψ is strictly increasing. As long as K_D has decreasing differences, more effective types of D choose greater hassling in equilibrium.

Lemma 5. *Assume (DD) holds. If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $h(\theta) \leq h(\theta')$.*

If more effective types choose higher levels of hassling, then they must also receive more favorable settlements in equilibrium. The next result is an immediate corollary of Lemma 4 and Lemma 5.

Corollary 2. *Assume (DD) holds. If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $V_D(\theta) \leq V_D(\theta')$.*

We can pin down more about the relationship between private information and equilibrium outcomes by imposing additional structure on the model. Our results so far are quite general—they allow for a discrete or convex type set, as well as discontinuous war payoff and hassling cost functions. However, if we introduce some additional continuity and differentiability conditions, we can yield "envelope theorem" results that more specifically characterize equilibrium payoffs and outcomes.

Definition 4. The model has *bounded variation* if W_D and K_D are differentiable and

$$\left.\begin{aligned} \Theta = [\underline{\theta}, \bar{\theta}] \qquad & \text{where } \underline{\theta} < \bar{\theta}, \\ |W_D(\theta) - W_D(\theta')| \leq M_W|\theta - \theta'| \qquad & \text{for all } \theta, \theta' \in \Theta, \text{ where } M_W < \infty, \\ |K_D(h,\theta) - K_D(h',\theta')| \leq \qquad & \text{for all } h, h' \in \mathcal{H} \\ M_D\|(h,\theta) - (h',\theta')\| \qquad & \quad \text{and } \theta, \theta' \in \Theta, \text{ where } M_D < \infty. \end{aligned}\right\} \tag{BV}$$

The bounded variation condition essentially reduces the number of degrees of freedom to consider in our analysis of incentive compatible direct mechanisms. It allows us to back out the settlement value for each type that ends up at when war does not occur, $V_D(\theta)$, from three components of the model: (1) the indicator for which types end up at war, $1 - \pi(\cdot)$; (2) the hassling of each type, $h(\cdot)$; and (3) the lowest type's equilibrium utility, $U_D(\underline{\theta})$. It becomes even simpler when the lowest type's equilibrium outcome is war, as then $U_D(\underline{\theta}) = W_D(\underline{\theta})$, an exogenous constant. The following proposition summarizes the relationship between these factors and each type's equilibrium utility.[18]

Proposition 1. *Assume* (BV) *holds. For all* $\theta_0 \in \Theta$*,*

$$U_D(\theta_0) = U_D(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_0} (1 - \pi(\theta)) \frac{dW_D(\theta)}{d\theta} \, d\theta - \int_{\underline{\theta}}^{\theta_0} \pi(\theta) \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta. \tag{3}$$

Obviously, if all types in a neighborhood of θ go to war in equilibrium, then the marginal utility at θ simply equals the marginal war utility, $dW_D(\theta)/d\theta$. This is reflected in Equation 3. The implication about the settlement value in case of a bargained outcome is more important. If all types in a neighborhood of θ settle in equilibrium, then the proposition implies that the settlement increases at a rate equal to the marginal cost of hassling.

3.3 When Stronger Types Are Better at Hassling

To obtain stronger results about equilibrium behavior in flexible-response crisis bargaining games, we must place additional structure on how D's private type—which represents how much D expects to gain from war—affects D's effectiveness in hassling. We now consider the case in which both types of effectiveness go hand-in-hand. We will say θ *improves hassling effectiveness* if D's cost of hassling strictly decreases with D's type: if $\theta < \theta'$, then $K_D(h, \theta) > K_D(h, \theta')$ for all $h > 0$.

When higher types are more effective at both war and hassling, then higher types of D must be better off in equilibrium regardless of which outcome prevails. The following result extends Lemma 3 to apply even when a bargained outcome is not guaranteed.

Lemma 6. *Assume θ improves hassling effectiveness. If $\theta < \theta'$, then $U_D(\theta) \leq U_D(\theta')$.*

In fact, in this case higher types of D should usually be strictly better off than lower types. If two types both go to war in equilibrium, then of course the higher type is better off. If the outcome is a bargained outcome with nonzero hassling for both types, then the higher type must also be better off, per Corollary 1. Finally, if the lower type fights and the higher type does not, the voluntary agreements condition ensures that the higher type must still receive

[18] The condition on K_D in (BV) is slightly stronger than necessary for Proposition 1. For this result, we only need $K_D(h, \cdot)$ to be Lipschitz in θ for each fixed $h \in \mathcal{H}$.

at least its war payoff, making it better off. Therefore, two types can receive the same payoff only if the lower type avoids war and invests nothing in hassling.

Lemma 7. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and $U_D(\theta) = U_D(\theta')$, then $\pi(\theta) = 1$ and $h(\theta) = 0$.*

Loosely speaking, Lemma 7 tells us that D's equilibrium utility is strictly increasing with D's private willingness to fight in the vast majority of cases, assuming private type enhances hassling capability. Exceptions to this rule occur only in atypical cases, where we have an interval of types of D that all end up at peace and all choose $h = 0$ in equilibrium.

In ordinary crisis bargaining games without the possibility of hassling, types with greater military power are always more likely to fight along the path of play (Banks, 1990, Lemma 1). The same is not necessarily true in flexible-response crisis bargaining games, particularly when greater military strength is associated with greater effectiveness in hassling. In general, it is possible for the occurrence of war to increase, to decrease, or to behave non-monotonically as D's type increases.

When military and hassling effectiveness go together, we can only pin down a monotonic relationship between D's type and the occurrence of conflict by making strong assumptions about its *relative* effects on each component. Loosely speaking, if the increase in war payoffs due to an increase in θ is greater in magnitude than the concomitant decrease in the hassling cost, then higher types will fight and lower types will hassle. When this is true, we say the war utility is relatively increasing (WURI). The opposite is that the settlement utility is relatively increasing (SURI), in which case low types fight and high types hassle.

Definition 5. In a direct mechanism, the *war utility is relatively increasing* if

$$\begin{aligned} W_D(\theta') - W_D(\theta) &> K_D(h(\theta'), \theta) - K_D(h(\theta'), \theta') \\ &\text{for all } \theta, \theta' \in \Theta \text{ such that } \theta < \theta' \text{ and } \pi(\theta') = 1. \end{aligned} \tag{WURI}$$

The *settlement utility is relatively increasing* if

$$\begin{aligned} W_D(\theta') - W_D(\theta) &< K_D(h(\theta), \theta) - K_D(h(\theta), \theta') \\ &\text{for all } \theta, \theta' \in \Theta \text{ such that } \theta < \theta' \text{ and } \pi(\theta) = 1. \end{aligned} \tag{SURI}$$

If either of these conditions holds, then we can obtain a monotonicity result for the equilibrium occurrence of war. High types are associated with war when the war effect is stronger (WURI), and with settlement when the war effect is weaker (SURI).

Lemma 8. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and* (WURI) *holds, then $\pi(\theta) \geq \pi(\theta')$. If $\theta < \theta'$ and* (SURI) *holds, then $\pi(\theta) \leq \pi(\theta')$.*

This result shows that the conventional relationship between private information and the likelihood of conflict is not robust to the introduction of hassling that affects payoffs from bargaining. Assuming that types with greater battlefield effectiveness are also more effective at hassling activities, the relationship between θ and the likelihood of conflict depends critically on the technology of hassling. If the marginal effect of D's type on the costs of hassling always outweighs its effect on the war payoffs, then we have the opposite of the usual result, with stronger types less likely to fight on the path of play.

The two conditions we have outlined here are mutually exclusive (except in the trivial case where all types end up fighting in equilibrium), but they are not mutually exhaustive. Depending on the functional forms of W_D and K_D, it is possible for the marginal effect of θ on the war payoff to be relatively strong for some types and relatively weak for others. If so, we cannot generally characterize the relationship between D's private type and which outcome prevails in equilibrium.

While Lemma 8 is useful for understanding how private information affects the occurrence of war in flexible response crisis bargaining games, its practical applicability is somewhat limited. Ideally, we would be able to say on the basis of the model primitives—the war payoff and hassling cost functions—whether stronger types will be associated with a greater likelihood of conflict in any given strategic environment. However, the WURI and SURI conditions do not exclusively concern model primitives, as they depend on the levels of hassling chosen on the path of play. This raises the possibility that the relationship between D's private type and the likelihood of conflict may vary depending on the exact bargaining protocol.

With additional conditions on the model primitives, we can ensure that the war utility is relatively increasing, meaning the likelihood of conflict increases with D's type. In particular, we need the cost function to have decreasing differences and for the marginal effect of θ on the war payoff to always exceed its marginal effect on the hassling cost when h is at its upper bound. Under these conditions, higher types are more likely to end up at conflict in the equilibria of all flexible response crisis bargaining games, regardless of the exact negotiating protocol employed.

Lemma 9. *Assume θ improves hassling effectiveness,* (DD) *holds, and* $\max \mathcal{H} = \bar{h} < \infty$. *If* $W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta')$ *for all* $\theta, \theta' \in \Theta$ *such that* $\theta < \theta'$, *then* (WURI) *holds.*

There is not an analogous sufficient condition for the settlement utility to be relatively increasing. The obstacle to such a condition is our assumption that $h = 0$ is always feasible at zero cost.[19] This means the marginal effect of θ on the cost of the hassling is zero for $h = 0$, ruling out any kind of sufficient condition for the marginal effect of θ on the hassling cost to always exceed its effect on the war payoff. At most, if we assume decreasing differences in the cost of hassling, we can make the SURI condition slightly less onerous to check. In this case, letting $\underline{h}$ denote the minimal hassling among types that end up in a bargained resolution, a sufficient condition is that $W_D(\theta') - W_D(\theta) < K_D(\underline{h}, \theta) - K_D(\underline{h}, \theta')$ for all $\theta < \theta'$. If this condition holds, then the equilibrium must entail low types of D fighting and high types of D settling.

3.4 When Stronger Types Are Worse at Hassling

We now consider situations where θ is negatively correlated with D's effectiveness in hassling. We say θ *degrades hassling effectiveness* if D's cost of hassling increases with D's type: if $\theta < \theta'$, then $K_D(h, \theta) < K_D(h, \theta')$ for all $h > 0$.

While this situation may seem counter-intuitive, this can occur whenever hassling technologies are not well designed for warfare (or vice versa). For example, while in the early 2000s the U.S. possessed an ability to conduct sophisticated precision strike attacks, it lacked a robust ability to conduct a prolonged counterinsurgency; thus, the U.S. could capably hassle Iraq, but could not as easily engage in a protracted counterinsurgency in Iraq. Note that section 1 includes additional examples.

If greater military strength is associated with lower hassling effectiveness, then the equilibrium will entail high types fighting and low types settling. Unlike when strength and hassling effectiveness are positively associated, in this case we need not place any additional restrictions on the technology of war or hassling to obtain monotonicity of equilibrium outcomes.

[19] The normalization of the cost to zero is immaterial, but the constancy of the cost of across types of D is important here.

Lemma 10. *Assume θ degrades hassling effectiveness. If $\theta < \theta'$, then $\pi(\theta) \geq \pi(\theta')$.*

This result corresponds to the monotonicity observed in traditional crisis bargaining games, in which stronger types are more likely to go to war (Banks, 1990; Fey and Ramsay, 2011). Viewed this way, the environment in which θ degrades hassling effectiveness is most like the ordinary crisis bargaining environment without flexibility in case of settlement. However, as we show below, the introduction of flexible responses still leads to substantial differences in equilibrium outcomes.

In equilibrium when θ degrades hassling effectiveness, low types settle and high types go to war. Consequently, there is a cutpoint, $\hat{\theta}$, summarizing which types settle and which fight. D's equilibrium utility is decreasing up to this cutpoint, as types with lower hassling effectiveness receive less from settling. Beyond the cutpoint, however, D's utility increases, as higher types expect to do better in wartime. If we place additional structure on the model, namely the bounded variation assumptions summarized by (BV), we can pin down two additional features of the equilibrium. First, the cutpoint type's utility must equal its war payoff, even if it chooses to settle in equilibrium. Second, the value of the bargained settlement is pinned down by the cutpoint type's war utility and the equilibrium hassling activity for each type up to the cutpoint. The following result summarizes these properties of the equilibrium.

Lemma 11. *Assume θ degrades hassling effectiveness.*

(a) There exists $\hat{\theta} \in \Theta$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$.

(b) If (BV) holds and $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$, then $U_D(\hat{\theta}) = W_D(\hat{\theta})$.

(c) If (BV) holds, then for all $\theta_0 < \hat{\theta}$,

$$V_D(\theta_0) = W_D(\hat{\theta}) + K_D(h(\theta_0), \theta_0) + \int_{\theta_0}^{\hat{\theta}} \frac{\partial K_D(h, \theta)}{\partial \theta}\bigg|_{h=h(\theta)} d\theta. \tag{4}$$

(d) U_D is non-increasing on $[\underline{\theta}, \hat{\theta}]$ and strictly increasing on $[\hat{\theta}, \bar{\theta}]$.

The fact that D's utility is U-shaped as a function of D's type is a noteworthy contrast with ordinary crisis bargaining models. In an environment without flexible responses, types that expect to receive more from fighting also receive more at the bargaining table (Banks, 1990; Fey and Ramsay, 2011). In the class of models we consider, however, the value of settlement is determined not only by one's threat to fight, but also by one's ability to revise the terms of agreement through hassling. When D's ability to do this is negatively related to D's battlefield strength, a marginal increase in strength may actually be associated with a worse bargaining outcome.

Lemma 11 does not pin down anything about $h(\theta)$, the relationship between D's type and D's final hassling level. We know from Lemma 5 that higher types will choose lower hassling levels if the decreasing differences property is satisfied. Besides that, our results so far place no restrictions on how types affect hassling. This is no accident. If higher types have higher marginal costs of hassling (i.e., decreasing differences is satisfied) and there is bounded variation in the model primitives, then we can rationalize a wide variety of relationships between types and the equilibrium level of hassling. More formally, for any mapping from types into hassling that is non-increasing and absolutely continuous, we can design a direct mechanism that is incentive compatible and has voluntary agreements in which the choices of revisionism follow this mapping. Under these conditions, the *necessary* conditions identified in Lemma 11 are *sufficient* for a bargaining equilibrium.

Proposition 2. *Assume θ degrades hassling effectiveness and* (BV) *and* (DD) *hold. If the direct mechanism (h, π, V_D) satisfies conditions (a)–(c) of Lemma 11 with $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ and h is absolutely continuous and non-increasing, then this direct mechanism satisfies* (IC) *and* (VA)*.*

The upshot of this finding is that incentive compatibility and voluntary agreements place few restrictions on the relationship between private types and D's equilibrium hassling levels in an environment where battlefield strength is associated with lower hassling effectiveness. For any non-decreasing and absolutely continuous relationship between θ and hassling costs, we can write down a flexible-response crisis bargaining game where D's equilibrium choice of hassling follows the specified relationship. Any further restrictions on the relationship between private information and D's equilibrium hassling must be traced to features of a particular bargaining protocol, not to the technology of warfare or even the cost function for hassling.

3.5 The Possibility of Peace

Previous game-free analyses of crisis bargaining have sought to identify sufficient conditions for peace to be an equilibrium outcome (Fey and Ramsay, 2009, 2011). The mechanism design approach is well suited for studying the possibility of peace. We may ask whether there is *any* crisis bargaining game that would lead to peace as an equilibrium outcome. Any given game form may result in a positive probability of war, but that does not mean war is inevitable—unlike bargaining in legislatures, states are not bound to follow any particular protocol.

When working with flexible-response crisis bargaining games, we must be explicit about what it means for an equilibrium to be peaceful. At a minimum, as in ordinary crisis bargaining games, the game must end with a negotiated settlement for all types $\theta \in \Theta$. Furthermore, because transgressions and hassling may be interpreted as forms of low-level conflict, we will focus on equilibria in which C chooses $t = 0$ and each type of D chooses $h = 0$. Mirroring the terminology of Fey and Ramsay (2011), we will call an equilibrium meeting these conditions *always peaceful*.

In our baseline flexible-response context, the sufficient condition for peace is virtually the same as in ordinary crisis bargaining models. In particular, it must be possible to divide the pie so as to simultaneously satisfy both C, assuming C's knowledge of D's type is limited to the prior distribution, and the strongest type of D. In what follows, let $\hat{W}_C = \mathbb{E}[W_C(\theta)] = \int_\Theta W_C(\theta)\, dF(\theta)$, C's prior expectation of its own war payoff.

Proposition 3. *If $\hat{W}_C + W_D(\bar{\theta}) \leq 1$, then there is a flexible-response crisis bargaining game form with voluntary agreements that has an always peaceful equilibrium.*

The condition of Proposition 3 is least likely to hold when the distribution of D's type is right-skewed. In this case, C's expected war payoff will be relatively high, since D's type is likely to be low. It will thus be impossible to satisfy the (rarely occurring) strongest type of D while giving C at least its expected war payoff.

If C's war payoff is independent of D's type (i.e., D's type only affects its cost of fighting, not its probability of victory), then the condition of Proposition 3 is sure to hold. A distribution of the pie following the probability of war will be acceptable both to C and to all types of D. The following result is a direct analogue of Proposition 2 in Fey and Ramsay (2011).

Corollary 3. *If $W_C(\theta) = p - c_C$ and $W_D(\theta) = 1 - p - c_D(\theta)$, where $c_D : \Theta \to \mathbb{R}_+$, then there is a flexible-response crisis bargaining game form with voluntary agreements that has an always peaceful equilibrium.*

One may wonder why our conditions for always peaceful equilibria do not depend on the costs of transgressions or hassling. This is due to our baseline assumption that t and h can only affect payoffs from negotiations, not from fighting. As we show in an extension below, if we make the war payoffs functions of these choices, then the players' reservation values for conflict depend on the marginal effects and costs of transgressions and hassling. In order for the flexible-response framework to materially affect the prospects for peace, these responses must shape payoffs in war as well as peace.

4 Discussion

The field of international relations has spent decades exploring the relationship between private information and war within the context of crisis bargaining models. A common shared assumption within these models is that states end the game either with a peaceful resolution to their crisis or with a war.[20] Empirically, this is perhaps too strong an assumption; after all, states can engage in a wide range of different destructive possible activities. But theoretically, before this paper, we did not know how innocuous (or problematic) the dichotomous peace-or-war assumption was. While some existing work has shown, for example, the possibility of low-level conflict can serve as an alternative to a greater war (McCormack and Pascoe, 2017; Coe, 2018; Schram, Forthcoming; Joseph, 2020) or that low-level conflict capabilities can perversely undermine efforts at deterrence (Bas and Coe, 2016; Baliga *et al.*, 2020; Schram, 2020), the topic of how private capabilities shape conflict outcomes has not been broadly examined.

As demonstrated in the games in figures 1 and 2, within some models with multiple policy options, the seminal monotonicity results in Banks (1990) and Fey and Ramsay (2011) no longer hold. These simple games raise questions about the relationship between private wartime payoffs and likelihood of conflict and state payoffs, but admittedly, are not particularly constructive. After all, it would be straightforward to consider different parameters and different game forms where the Banks and Fey and Ramsay monotonoicity results hold. To build results that can be used, we define a broad class of games with multiple conflict options—flexible-response crisis bargaining games—and conduct a Bayesian mechanism design analysis of equilibria within these games. In doing so, we establish several general findings on the relationship between private information and outcomes.

[20]More broadly, the existing literature typically assumes state are engaging in an efficient division of an asset or are engage in some fixed, inefficient action that divides the asset, like war or sanctions

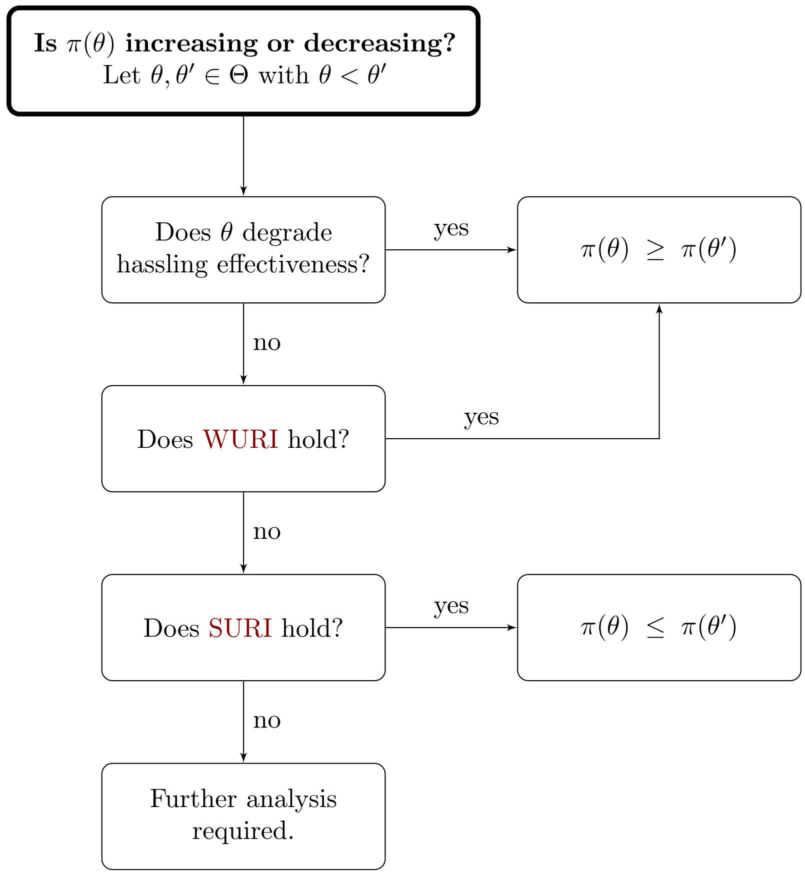

Figure 6: The relationship between type and war likelihood in all flexible-response crisis bargaining games. We assume that the $K(h, \theta)$ is monotonic in θ for all h.

4.1 On Private Type and Conflict Probability

Banks (1990) and Fey and Ramsay (2011) observed a positive monotonic relationship between private wartime payoffs and likelihood of conflict within any equilibrium of any crisis bargaining model. As we present in the game in 1, we do not always observe this monotonicity within flexible-response crisis bargaining games. The flow chart in figure 6 clarifies, for a given equilibrium to a flexible-response crisis bargaining model, what to expect and when. For any flexible-response crisis bargaining game considered here, we assume that the $K(h, \theta)$ is monotonic in θ for all h.

Consider the set of flexible-response crisis bargaining games. To understand whether or not a greater private wartime payoff (θ) implies a greater likelihood of war in equilibrium, one should first ask if hassling becomes more expensive for D as D's private wartime payoffs increase (i.e. if, for $\theta < \theta'$ and for all $h \in \mathcal{H}$ with $h > 0$, $K_D(h, \theta) < K_D(h, \theta')$). If θ degrades hassling effectiveness, then the positive relationship between private wartime payoffs and war likelihood described in Banks (1990) and Fey and Ramsay (2011) continues to hold

(Lemma 10). Intuitively, in both standard crisis bargaining games and flexible-response crisis bargaining games, C may undertake a series of bargaining moves that will provoke some types of D to go to war. As the simplest example, consider C making an ultimatum offer: sometimes, C prefers making a less-conciliatory offer that risks war with some types of D to an offer that will appease all possible types of D but that leaves more surplus on the negotiating table. The logic of the monotonicity result in crisis bargaining models is that if C is willing to risk war sometimes, the types of D that will be unsatisfied are the types of D that are privately the best at war. Introducing the possibility of transgressions and hassling and assuming that an improved private wartime payoff makes hassling more expensive does nothing to break the monotonicity. If C is willing to risk war with some types of D, now the types of D that are best at war are also the worst at hassling, thus insuring that if any types will go to war, it will be the highest types.

Now consider alternate settings where θ improves hassling effectiveness. If WURI holds for a given θ and θ' with $\theta < \theta'$, then the likelihood of war is weakly increasing in moving from θ to θ' (Lemma 8). Alternatively, if SURI holds for a given θ and θ', then the likelihood of war is weakly decreasing moving from θ to θ' (Lemma 8). If neither WURI nor SURI hold, then we are unable to say whether the likelihood of war is weakly increasing or decreasing in moving from θ to θ', and an equilibrium analysis of the game is needed.

It is worthwhile expanding on WURI and SURI here. To identify if WURI or SURI hold for a given θ and θ' in the broadest possible sense, a partial equilibrium analysis is needed. For WURI, one must check if $\pi(\theta') = 1$ and identify $h(\theta')$, and for (SURI), one must check if $\pi(\theta) = 1$ and identify $h(\theta)$. Once this analysis is conducted, identifying whether the likelihood of peace is weakly increasing or weakly decreasing becomes a matter of examining model primitives.

Alternatively, we can identify if WURI holds based on model primitives (Lemma 9), but this requires stricter assumptions. If the cost function K_D has decreasing differences in h and θ—satisfies DD—and, letting $\max \mathcal{H} = \bar{h} < \infty$, if $W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta')$ holds, then we can say that WURI holds and that the likelihood of war is weakly increasing in moving from θ to θ'. Unfortunately we are unable to establish an analogous sufficiency condition based entirely on model primitives for when SURI holds.

4.2 On Private Type and Utility

In Banks (1990) and Fey and Ramsay (2011), as war becomes privately better for a state, that state will, in expectation, perform weakly better within the equilibrium to the crisis

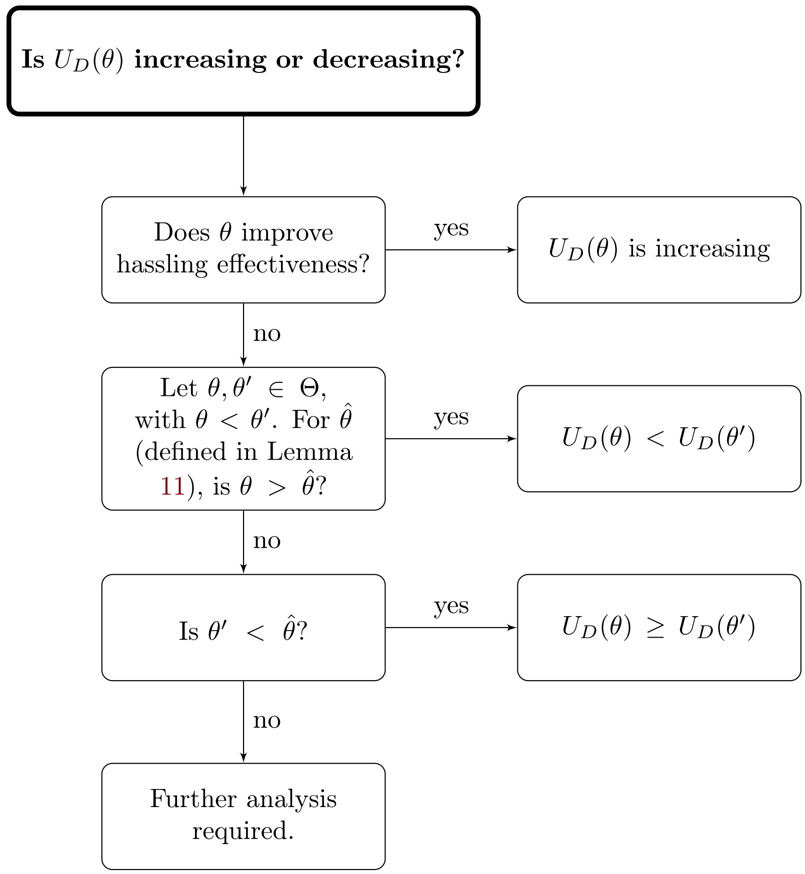

Figure 7: The relationship between type and D's utility in all flexible-response crisis bargaining games. We assume that the $K(h, \theta)$ is monotonic in θ for all h.

bargaining game. The game in 2 demonstrates that this may not hold when considering games with multiple policy options. For the set of equilibria to flexible response crisis bargaining games, we can in some cases describe whether D's utility is increasing or decreasing as their private wartime payoffs increase. The flow chart in figure 7 clarifies what to expect and when. For any flexible-response crisis bargaining game considered here, we assume that the $K(h, \theta)$ is monotonic in θ for all h.

If improvements in D's private wartime payoffs concurrently make hassling less expensive, then as war becomes privately better for a state, that state will, in expectation, perform weakly better within the flexible-response crisis bargaining game within an equilibrium. Essentially, because war and hassling are improving in type, a type θ' could always mimicking the equilibrium actions of type θ and attain a higher utility. Of course, this is not to say the the type θ' D will mimic the type θ D in equilibrium; rather, this suggests, at minimum, the type θ' D could mimic and perform better, thus establishing a kind of guaranteed baseline payoff that is better than type θ's payoff, and that can only be improved upon through alternate actions.

When θ degrades hassling effectiveness, then the relationship between D's private war payoffs and D's utility can be monotonically increasing, decreasing, or exhibit a U-shape. We use Lemma 11 to discern the shape of this relationship. Lemma 11 requires a partial equilibrium analysis to identify the type of D that represents a threshold type where all lower-types will not go to war and all higher-types will go to war (denoted $\theta = \hat{\theta}$). Once this is done, we can identify what effect a shift from θ to θ' has on D's utility. If both θ and θ' fall below the $\hat{\theta}$ type (which occurs when $\theta' \leq \hat{\theta}$, then in equilibrium war will not occur for either type, and the utility will be non-increasing across these types. If both types fall above the $\hat{\theta}$ type (which occurs when $\theta \geq \hat{\theta}$), then in equilibrium war will occur across both types the utility will be strictly decreasing across these types. If neither condition holds, which occurs when $\hat{\theta}$ falls between θ and θ', then a further analysis of the game is required.

4.3 On Always Peaceful Game Forms

While the monotonicity results for flexible-response crisis bargaining games are quite different from those for crisis bargaining models, the conditions for the existence of a game form that produces an always-peaceful equilibrium are essentially identical. Why is this the case? Fey and Ramsay (2011) describes a simple game: both players can "accept" and receive their smallest fixed payoff that will keep each of the privately-best-at-war types of D or C from going to war, or they can declare (and go to) "war". So long that such a division—where each player receives a fixed payoff that would keep their most-aggressive types from going to war—is possible, this game form can exist, and a peaceful equilibrium can occur. Proposition 3 functions for a similar equilibrium to a nearly identical game form within the flexible-response crisis bargaining framework. Simply, if C sets $t = 0$ and D sets $h = 0$, then both states receive the $\hat{W}_C$ and $W_D(\bar{\theta})$ payoffs, which is guaranteed to weakly exceed C's and D's expected wartime payoffs; if $t \neq 0$, $h \neq 0$, or either state declares war, then war occurs. Thus, the conditions for the existence of an always peaceful equilibrium to some game form is roughly the same.

4.4 Empirical Relevance

What are the empirical implications of our theory? Consider the ongoing debate over the usefulness of aerial bombing in conflict. To quickly summarize a subset of findings, existing research has shown that aerial bombings or precision strike capabilities can be useful in hassling operations (Kreps and Fuhrmann, 2011), can be useful in a conventional war (Pape, 1996; Horowitz and Reiter, 2001; Allen and Martinez Machain, 2019), and may not be as useful in a counterinsurgency (Lyall, 2013; Dell and Querubin, 2018). While this empirical research agenda is quite deep, the existing literature does not speak to, for example, what precision strike technology means for shaping deterrence or bargaining.[21] In every paper cited above, the value of these weapons were evaluated conditional upon a deterrence failure (or bargaining failure) having occurred. It would also be valuable to know how these weapons influence decision making in the lead-up to crises, whether they lead to more (or less) war, and whether they produce expected windfalls for the state that developing the bombing capabilities. But, these latter questions are difficult to answer empirically; empirically identifying or constructing a counterfactual world where a state did not develop precision strike capabilities is challenging, as is identifying how these capabilities (or their absences) shape international crisis initiation or behavior. Our theory is well suited to address these questions. The flowcharts in Figures 6 and 7 demonstrate that once we identify if improved private wartime capabilities improve or are detrimental to hassling capabilities—if hassling costs are increasing or decreasing in θ—then we are able to make strong statements about how changes in wartime capabilities increase or decrease utilities or the likelihood of war. Armed with our theory, future scholars can leverage past empirical and public policy research to identify how improving specific capabilities (or willingness mechanisms) can affect the outcomes in any setting where a flexible-response crisis bargaining model can apply.

To illustrate how this could work, consider the existing research on bombings and precision strike capabilities. Considered a crisis where (a) a defender state possesses private information about their ability to conduct precision strikes, (b) the most feasible low-level conflict option is a precision strike (along the lines of Operation Desert Fox), and (c) the war option would be a conventional war (along the lines of Operation Desert Storm). Because precision strike technology is dual-use for both hassling and war, then these private capabilities would have a positive effect on both war and hassling outcomes (hence the “ + ” symbols). Thus, if (a)-(c) held, as the flowcharts in figures 6 and 7 show, any flexible-response crisis bargaining model of this case would find that a defender with a better private ability to conduct precision strikes would end up with greater final payoffs, but not necessarily a greater likelihood of engaging in a conventional war. On the other hand, consider a crisis where (a) and (b) held, but, instead of (c), it was the case that (d) the war option would be a protracted counterinsurgency, where precision strike capabilities may be less effective (hence the “ - ” symbol).[22] If (a), (b), and (d) held, then any flexible-response crisis bargaining model of this

[21] One notable exception is Post (2019), which analyzes airpower events (i.e. military exercises, mobilizations, shows of force, deployments of military assets, or other military signal) as signals for use in compellence.

[22] Additionally, in a setting with budget restrictions, it is plausible that investments in precision strike

case would find that a defender state with a better private ability (or private willingness) to conduct precision strikes would be less likely to enter into a war, and may or may not end up with greater final expected payoffs.[23]

It is straightforward to identify a number of other capabilities that are dual-use for hassling and war. During Operation Outside the Box (2007), Israel disabled Syrian air defenses with an electronic warfare (EW) attack. While the full details of the EW attack are not disclosed, an attack that allowed multiple Israeli aircraft to enter Syria and conduct a raid without harassment plausibly could have been used to conduct a more extensive conventional attack as well (Katz, 2010). Additionally, assorted cyberattacks have been used both independently (Stuxnet and Estonian cyber attacks), as a part of a cluster of operations against a target state (the NotPetya attacks targeting Ukraine), or as part of a conventional war (Russia-Georgia War, 2008) (Buchanan, 2020; Gannon *et al.*, 2020). Furthermore, developments in anti-satellite technologies have opened the possibility for disruption of GPS signals, which could both create low-level disruptions and create serious problems for modern air and sea warfare (Harrison *et al.*, 2020). While we do not have the space to discuss every dual-use technology case at length, a number of other capabilities positively affect hassling and war in straightforward ways: continued development of airlift capabilities can facilitate special operations for use in low-level conflict, conventional war, and irregular warfare (Bolkcom, 2007; Pietrucha and Renken, 2019); coordinating with or providing support to violent non-state actors found use in hassling and war (Schultz, 2010; Schram, Forthcoming); and an ability to control the civilian information environment have been used in hassling operations (Russia in Eastern Ukraine, 2015-on) and hearts-and-minds counterinsurencies.

Of course, a number of capabilities are not dual use. Just as precision strike capabilities may not be well suited for a protracted hearts-and-minds COIN campaign, the electronic, cyber, or anti-satellite tools can be very expensive and may not be as well suited for a COIN environment;[24] to the extent that, for example, states over-invest in precision strike capabilities, those states may be less able to conduct a war requiring extensive COIN operations. Alternatively, while information operations can help in a hassling or COIN context, these operations may not be as well-suited for a conventional conflict. Furthermore, the U.S. Navy, Marine Force, and Coast Guard has jointly issued a report suggesting that conventional navy vessels designed for lethality in conventional warfare may not be as effective in deft handling of gray zone attacks in the pacific theater, and a more agile fleet could better serve in "gray

capabilities were made at the expense of counterinsurgency capabilities.

[23] While throughout the text we treat θ as improving war capabilities, we can draw the conclusions stated here due to 10.

[24] There are some notable exceptions, see Shachtman (2011).

Capability/Willingness Mechanism	**Hassling Type**	**War Type**	**Rationale**
Precision Strikes/Drones	Targeted strikes against military facilities (+)	Conventional War (+)	Dual-Use Technology
Electronic/Cyber/Anti-Satellite Attacks	Disruption of military systems or weapons facilities (+)	Conventional War (+)	Dual-Use Technology
Airlift capabilities	Special operations (+)	Conventional/ Irregular War (+)	Dual-Use Technology
Militants on Retainer	Supporting low-level conflict/insurgency (+)	Conventional/ Irregular War (+)	Dual-Use Technology
Domestic Information Operations	Undermining domestic authority/promoting discord (+)	Irregular War (+)	Dual-Use Technology
Domestic Information Operations	Undermining domestic authority/promoting discord	Conventional War (-)	Specialized Technology/Budgetary Issues
Precision, Electronic, Cyber, Anti-Satellite Attacks	Targeted strikes against military facilities (+)	Irregular War/COIN (-)	Specialized Technology/Budgetary Issues
Conventional Navy gray hulls optimized for high-end naval warfighting	Countering naval gray zone operations (-)	Conventional War (+)	Specialized Technology/Budgetary Issues
Domestic Electoral Considerations	Sanctions, hassling, or any other low-level response (-)	Conventional War (+)	Rally 'round the Flag
Domestic Political Economy Considerations	Sanctions or low-level conflict (+/-)	Conventional War (+/-)	Political Elite Disconnect

Table 1: How various capabilities or willingness mechanisms affect hassling and war capacities. The "+" symbol ("-" symbol) indicates that the capability or willingness mechanism improves (is detrimental to) the listed hassling operations or war type.

zone" operations (Berger *et al.*, 2020; Owen, 2021).

Everything described above is related to capabilities; other factors could shape state leadership's incentives in such a way that alters their willingness to engage in some forms of conflict over others. Statistical analyses of the "rally-round-the-flag" phenomenon suggest that larger military operations, especially wars, generate an increase in public support for domestic leadership (Baker and Oneal, 2001; Chapman and Reiter, 2004). This effect could alter a leader's preferences in such a way to prefer war to hassling operations. Alternatively, domestic political economy considerations can shape incentives of leadership. In 1954 the United States provided arms, funds, and training to Guatemalan rebels who overthrew Jacobo Árbenz and installed right-wing dictator Castillo Armas; this move benefited the politically connected United Fruit Company (Kinzer, 2007, 125–147). Similarly, while a "blood for oil" hypothesis may not fully explain 2003 Iraq invasion (Paul, 2003; Stokes, 2007), it could have still shaped a private willingness for U.S. leadership to go to war. In both cases, it is not clear that sanctions or non-decisive, low-level conflict options would have benefited private interests as effectively. Of course, while this is not to say that relevant domestic actors always prefer war, as the use of international sanctions or tariffs to weaken a regime can also create domestic winners and losers.

4.5 Limitations of the Flexible-Response Crisis Bargaining Framework

Our framework captures a crisis environment where states can form fully peaceful resolutions, states can conduct costly, low-level revisions to the peaceful resolution, or states can go to war. While the flexible-response crisis bargaining framework embraces a broader set of possible actions than what is allowed in standard crisis-bargaining models, it also has some subtleties and limits worth expanding on.

Our framework is distinct from what is covered in the Fey and Ramsay (2011) discussion of "war as a bargaining process." This alternate treatment of crisis bargaining models allows for some inefficiencies within bargained outcomes, which could be interpreted as allowing for low-level conflict to occur within bargaining. We concede that it could, but only if low-level conflict follows a very specific structure: namely, that the final payoffs engaging in some fighting and then settling are a convex-combination of the payoffs from fighting forever and the (efficient) payoff from the negotiated settlement. The setup presented in Fey and Ramsay follows naturally from second-wave "fighting while bargaining" models creates costs as it progresses (Wagner, 2000; Filson and Werner, 2002; Slantchev, 2003a; Smith and

Stam, 2004; Powell, 2004; Leventoğlu and Slantchev, 2007; Fearon, 2007). In this setup, each round of fighting has an identical structure, and where one round of fighting is identical to ten rounds of fighting, only the ten-rounds of fighting has ten times the costs. In the context of low-level conflict and war, this could be appropriate if low-level conflict used the same technology as a decisive war but was (proportionally) smaller or conducted over a shorter period. In most cases, this seems like too strong an assumption: the force posture used on Operation Desert Fox (a limited strike against Saddam's weapons facilities) and Operation Iraqi Freedom (what became a protracted counterinsurgency) were quite different, which implies that, even when considering kinetic operations, hassling and war do not share a common per-round structure.

One limitations of the above flexible-response crisis bargaining framework pertains to how we model the non-war outcomes: we assume that type only influences non-war payoffs through the decisions surrounding hassling. This modeling choice excludes the possibility, for example, that the selected transgressions and hassling influences future wartime payoffs, and that states have the option to go to war before or after the selected transgressions and hassling levels are implemented. As modeled in the text, our non-war outcomes describe settings where some political bargain is settled upon, and states push that bargain in one direction or another through transgressions and hassling. We interpret this as being consistent with treatments of gray zone conflict or hybrid wars as tools of "measured revisionism" (Mazarr, 2015a,b), However, adding the layer where these revisionist activities also affect future wartime payoffs could more thoroughly capture scenarios where transgressions or hassling today affects future wartime capabilities (Fearon, 1996; Hirsch and Shotts, 2015; Schram, Forthcoming). We include an analysis of such a class of models in the Extensions section.

We highlight two shortcomings of our framework as avenues for future research. First, the flexible-response crisis bargaining framework does not speak to scenarios where transgressions or hassling decisions are private but stochastically observed. This undercuts our framework's ability to describe, for example, settings where a transgression is imperfectly observed or where identifying attribution is hard. As one interpretation of these limitations, we admit that our framework is not well suited to describe cyberwarfare when attribution problems are present (Baliga *et al.*, 2020), or when there is a hidden development of technological capabilities as the transgression (Meirowitz *et al.*, 2008; Baliga and Sjöström, 2008; Schultz, 2010; Debs and Monteiro, 2014; Bas and Coe, 2016; Spaniel, 2019; Meirowitz *et al.*, 2019). Second, while introducing the possibility of a continuum of transgression and hassling options constitutes a step in the direction of better describing international interactions with multiple policy options, our transgression and hassling options are both uni-dimensional. Building out a more sophisticated framework that allows for more dimensions of policy responses could better describe the world.

5 Extension: When Flexible Responses Affect War

Our main model assumes no spillovers between flexible responses and war. Transgressions and hassling affect payoffs only in case of peace, and their costs are not realized if war occurs. In this section, we extend the baseline framework to allow for these spillovers, and we analyze how this extension alters the conditions for always peaceful equilibria to exist.

We extend the original model environment by making both players' war payoffs functions of the level of transgressions and hassling. We assume that the cost of any given choice of t or h is the same as if a settlement takes place, but the benefits may differ. Specifically, if war occurs in the extended model, C receives a payoff of $W_C(t, h, \theta) - K_C(t)$, where W_C is weakly increasing in t and weakly decreasing in h and θ. Similarly, D receives $W_D(t, h, \theta) - K_D(h, \theta)$, where W_D is weakly decreasing in t, weakly increasing in h, and strictly increasing in θ.[25]

In an environment where transgressions and hassling affect war payoffs as well, it is more difficult to sustain peace than in the baseline model. Remember that an always peaceful equilibrium entails zero transgressions, zero hassling, and no war. The additional difficulty for peace in this context comes from the temptation to exploit the other player through conflict when the other player's strategy involves spending nothing on the flexible response.

To illustrate the greater barriers to peace, we impose some specific functional forms for the relationship between transgressions, hassling, and war payoffs.[26] Let the type space for D be an interval, $\Theta = [\underline{\theta}, \bar{\theta}]$. We assume C's baseline war payoff is $p - \kappa_C$, where $p \in [0, 1]$ and $\kappa_C > 0$ is a constant representing a fixed cost of war. Similarly, the baseline war payoff for type θ of D is $1 - p - \kappa_D(\theta)$, where $\kappa_D(\underline{\theta}) = \underline{\kappa}_D$ and $\kappa_D(\bar{\theta}) = \bar{\kappa}_D$, scaling linearly in between. As D's war payoff increases with type, higher types must have lower costs: $\underline{\kappa}_D > \bar{\kappa}_D$.

Transgressions shift this baseline war payoff in favor of C, while hassling causes a shift toward

[25]Throughout this section we assume $\mathcal{T}$ and $\mathcal{H}$ are compact, that W_C and K_C are continuous in t, and that W_D and K_D are continuous in h. Together, these assumptions imply that C's war payoff always attains a maximum with respect to t, and similarly for D's with respect to h.

[26]Full formal details in subsection C.2.

D. Letting $S(t,h)$ denote the magnitude of the payoff shift, overall war payoffs are thus

$$W_C(t,h,\theta) = p - \kappa_C + S(t,h),$$
$$W_D(t,h,\theta) = 1 - p - \kappa_D(\theta) - S(t,h).$$

Notice that D's type has no direct effect on C's war payoff here. Nonetheless, in contrast with Corollary 3, we will see that this independence alone is not sufficient for peace. We assume a simple functional form for the payoff shift toward C, $S(t,h) = (1-h)t - \sigma t$, so that hassling has both a direct effect ($\sigma > 0$) and dampens the effect of transgressions.[27] Finally, we assume quadratic costs, $K_C(t) = \lambda_C t^2$ and $K_D(h,\theta) = \lambda_D(\theta)h^2$. As with D's cost of war, we have $\lambda_D(\underline{\theta}) = \underline{\lambda}_D$ and $\lambda_D(\bar{\theta}) = \bar{\theta}$, varying linearly in between.

First consider the case in which θ enhances hassling effectiveness, so $\underline{\lambda}_D < \bar{\lambda}_D$ (lower marginal cost of hassling for higher types). A peaceful equilibrium requires that the pie be large enough for a settlement that each player, including every type of D, prefers over war. If D is committed to no hassling, $h = 0$, then C attains its maximal war payoff at $t = \frac{1}{2\lambda_C}$, for a total war payoff of

$$\bar{R}_C = p - \kappa_C + \frac{1}{4\lambda_C}.$$

To characterize the analogous reservation value for D, it will suffice to consider solely the highest type, which has the lowest costs of both war and hassling. If C is committed to no transgressions, this type attains its maximal war payoff at $h = \frac{\sigma}{2\bar{\lambda}_D}$, resulting in a reservation value of

$$\bar{R}_D = 1 - p - \bar{\kappa}_D + \frac{\sigma^2}{4\bar{\lambda}_D}.$$

As in Proposition 3, a sufficient condition for an always peaceful equilibrium is $\bar{R}_C + \bar{R}_D \leq 1$. Here that is equivalent to

$$\kappa_C + \bar{\kappa}_D \geq \frac{1}{4}\left(\frac{\sigma^2}{\bar{\lambda}_D} + \frac{1}{\lambda_C}\right).$$

Notice that non-negative costs of war alone are not necessarily sufficient for peace here. If the marginal costs of transgressions and hassling are too low, or the direct effect of hassling on the payoff shift is too great, then this condition will fail.

Now suppose θ degrades hassling effectiveness, so $\underline{\lambda}_D > \bar{\lambda}_D$. The basic strategy to find a sufficient condition for peace is the same as before—identify each player's reservation value, then characterize when the pie is large enough to satisfy both. In fact, because C's war payoff

[27] We assume $\mathcal{H} = [0,1]$ and impose some boundary conditions to ensure the upper constraint never binds. See the appendix.

is independent of D's type, C's reservation value $\bar{R}_C$ is exactly the same as in the previous case. However, it is more complicated for D, because now the types with the lowest war costs have the highest hassling costs, so $\bar{\theta}$ might not attain the maximal war payoff across types of D. Whether that is the case depends on the relative marginal effects of type on war costs versus hassling costs. If $\underline{\kappa}_D - \bar{\kappa}_D$ is large enough relative to $\bar{\lambda}_D - \underline{\lambda}_D$, then the sufficient condition for peace is exactly the same as when θ enhances hassling capabilities. Otherwise, the lowest type of D can attain the maximal war payoff, and the sufficient condition for peace becomes

$$\kappa_C + \underline{\kappa}_D \geq \frac{1}{4}\left(\frac{\sigma^2}{\underline{\lambda}_D} + \frac{1}{\lambda_C}\right).$$

The upshot remains the same as before: low costs of transgressions and hassling, as well as high direct effects of hassling on the payoff shift, may threaten the possibility of peace.

We should note that the conditions derived here are not tight, in that we cannot definitively say that an always peaceful equilibrium fails to exist if the conditions are violated. The obstacle to obtaining a necessary condition is that each player's reservation value—the most they could expect to obtain from forcing a conflict—is dependent on the game form. For example, D's reservation value will be relatively high if D moves last and can choose h to respond optimally to C's choice of t; it will be relatively low if instead C moves last.

6 Conclusion

We examine a new class of games that fall within the flexible-response crisis bargaining framework. Through our approach, we are able to establish general results on the relationship between private information and conflict. Rather than fix a single game form and solve for a set of equilibria, we are able identify the properties shared by all equilibria within the class of flexible-response crisis bargaining games, regardless of how the bargaining actually plays out. The value to this approach is that we can identify relationships between private capabilities and outcomes like war and low-level conflict without the concern that a specific equilibrium or modeling specification is driving the result. While typically a cluster of models or papers are needed to identify the general relationships within a class of models, the results here skirts that process, and allows for researchers to focus on new classes of models or the specifics of the game-play.

Our most surprising results are the results that present relationships that are different from what Banks (1990) and Fey and Ramsay (2011) find. While these papers show that improved private war capabilities or an improved private willingness to go to war *always* results in weakly more war, we find the results are more nuanced when war capabilities can also benefit low-level conflict capabilities or hassling capabilities. Similarly, while Banks (1990) and Fey and Ramsay (2011) show that an improved private ability to conduct war *always* produces a greater utility, we find the results do not necessarily hold when a robust ability to go to war can hurt an actor's ability to effectively sanction or hassle.

A central concern of international relations is understanding the drivers of costly and destructive conflict. While this topic has been well examined through models of war and peace, much of what occurs in international relations falls outside of what could easily be classified as a peaceful bargain or a decisive war. While naturally any model must make some simplifying assumptions on how the world works, this paper shows that neglecting the possibility for low-level responses, we may be misunderstanding how what actually drives war. More work is needed on this topic.

References

Acharya, A., Grillo, E. *et al.* (2015) War with crazy types, *Political Science Research and Methods*, **3**, 281–307. (document)

Aghion, P. and Jackson, M. O. (2016) Inducing leaders to take risky decisions: dismissal, tenure, and term limits, *American Economic Journal: Microeconomics*, **8**, 1–38. (document)

Allen, S. H. and Martinez Machain, C. (2019) Understanding the impact of air power, *Conflict management and peace science*, **36**, 545–558. 4.4

Ashworth, S. and Bueno de Mesquita, E. (2006) Monotone comparative statics for models of politics, *American Journal of Political Science*, **50**, 214–231. 3.2

Ashworth, S. and Sasso, G. (2019) Delegation to an overconfident expert, *The Journal of Politics*, **81**, 692–696. (document)

Axelrod, R. and Keohane, R. O. (1985) Achieving cooperation under anarchy: Strategies and institutions, *World politics*, **38**, 226–254. (document)

Baker, W. D. and Oneal, J. R. (2001) Patriotism or opinion leadership? the nature and origins of the "rally'round the flag" effect, *Journal of conflict resolution*, **45**, 661–687. 4.4

Baliga, S., De Mesquita, E. B. and Wolitzky, A. (2020) Deterrence with imperfect attribution, *American Political Science Review*, **114**, 1155–1178. (document), 3, 4, 4.5

Baliga, S. and Sjöström, T. (2008) Strategic ambiguity and arms proliferation, *Journal of political Economy*, **116**, 1023–1057. 4.5

Baliga, S. and Sjöström, T. (2020) The strategy and technology of conflict, *Journal of Political Economy*, **128**, 3186–3219. 3

Banks, J. S. (1990) Equilibrium behavior in crisis bargaining games, *American Journal of Political Science*, pp. 599–614. (document), 1, 2.1, 8, 14, 3.1, 3.1, 3.3, 3.4, 3.4, 4, 4.1, 4.1, 4.2, 6

Baron, D. P. and Besanko, D. (1987) Commitment and fairness in a dynamic regulatory relationship, *The Review of Economic Studies*, **54**, 413–436. (document)

Baron, D. P. and Besanko, D. (1992) Information, control, and organizational structure, *Journal of Economics & Management Strategy*, **1**, 237–275. (document)

Bas, M. A. and Coe, A. J. (2016) A dynamic theory of nuclear proliferation and preventive war, *International Organization*, **70**, 655–685. 4, 4.5

Benson, B. and Smith, B. C. (2020) Do alliances cause commitment problems and war?, *Working Paper*. 1

Berger, D. H., Gilday, M. M. and Schultz, K. L. (2020) *Advantage at Sea: Prevailing with Integrated All-Domain Naval Power*, United States Department of Defense. 4.4

Bolkcom, C. (2007) Military airlift: C-17 aircraft program, LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE. 4.4

Buchanan, B. (2020) *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics*, Harvard University Press. 4.4

Chapman, T. L. and Reiter, D. (2004) The united nations security council and the rally'round the flag effect, *Journal of Conflict Resolution*, **48**, 886–909. 4.4

Chassang, S. and Miquel, G. P. I. (2010) Conflict and deterrence under strategic risk, *The Quarterly Journal of Economics*, **125**, 1821–1858. 3

Chassang, S. and Padró i Miquel, G. (2009) Defensive weapons and defensive alliances, *American Economic Review*, **99**, 282–86. (document)

Coe, A. J. (2014) Sanctions as instruments of regime change: A case study of civil conflict in iraq during containment, Tech. rep., Working Paper. (document)

Coe, A. J. (2018) Containing rogues: A theory of asymmetric arming, *The Journal of Politics*, **80**, 000–000. (document), 4

Coe, A. J. and Vaynman, J. (2020) Why arms control is so rare, *American Political Science Review*, pp. 1–14. (document)

Debs, A. and Monteiro, N. P. (2014) Known unknowns: power shifts, uncertainty, and war, *International Organization*, **68**, 1–31. (document), 1, 4.5

Dell, M. and Querubin, P. (2018) Nation building through foreign intervention: Evidence from discontinuities in military strategies, *The Quarterly Journal of Economics*, **133**, 701–764. 4.4

Di Lonardo, L. and Tyson, S. (2017) Political instability and the failure of deterrence, *Available at SSRN 3094929*. 3

Dunford, J. (2016) Gen. Dunford's Remarks and Q&A. (document)

Fearon, J. D. (1995) Rationalist explanations for war, *International organization*, **49**, 379–414. (document)

Fearon, J. D. (1996) Bargaining over objects that influence future bargaining power. 1, 4.5

Fearon, J. D. (1997) Signaling foreign policy interests tying hands versus sinking costs, *Journal of Conflict Resolution*, **41**, 68–90. 3

Fearon, J. D. (2007) Fighting rather than bargaining, in *Annual Meetings of the American Political Science Association*. 4.5

Fey, M. and Kenkel, B. (2019) Is an ultimatum the last word on crisis bargaining? (document), 9

Fey, M. and Ramsay, K. W. (2009) Mechanism design goes to war: peaceful outcomes with interdependent and correlated types, *Review of Economic Design*, **13**, 233. (document), 13, 3.5

Fey, M. and Ramsay, K. W. (2011) Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict, *American Journal of Political Science*, **55**, 149–169. (document), 1, 1, 2.1, 8, 12, 2.2, 2.2, 3.4, 3.4, 3.5, 3.5, 4, 4.1, 4.1, 4.2, 4.3, 4.5, 6, C

Filson, D. and Werner, S. (2002) A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war, *American Journal of Political Science*, pp. 819–837. 4.5

Gannon, A. J., Gartzke, E., Lindsay, J. R. and Schram, P. (2020) The shadow of deterrence, *Manuscript*. (document), 4.4

Gartzke, E. and Lindsay, J. R. (2015) Weaving tangled webs: offense, defense, and deception in cyberspace, *Security Studies*, **24**, 316–348. (document)

Gartzke, E. and Lindsay, J. R. (2017) Thermonuclear cyberwar, *Journal of cybersecurity*, **3**, 37–48. 1

Gurantz, R. and Hirsch, A. V. (2017) Fear, appeasement, and the effectiveness of deterrence, *The Journal of Politics*, **79**, 1041–1056. (document), 1

Harrison, T., Johnson, K., Roberts, T. G., Way, T. and Young, M. (2020) *Space threat assessment 2020*, Center for Strategic and International Studies (CSIS). 4.4

Hirsch, A. V. and Shotts, K. W. (2015) Competitive policy development, *The American Economic Review*, **105**, 1646–1664. 4.5

Hörner, J., Morelli, M. and Squintani, F. (2015) Mediation and peace, *The Review of Economic Studies*, **82**, 1483–1501. (document)

Horowitz, M. and Reiter, D. (2001) When does aerial bombing work? quantitative empirical tests, 1917-1999, *Journal of Conflict Resolution*, **45**, 147–173. 4.4

Huth, P. K. (1999) Deterrence and international conflict: Empirical findings and theoretical debates, *Annual Review of Political Science*, **2**, 25–48. 3

Joseph, M. (2020) The secret failure of nuclear containment, *Working Paper*. (document), 1, 4

Katz, Y. (2010) And they struck them with blindness, *The Jerusalem Post*. 1, 4.4

Kinzer, S. (2007) *Overthrow: America's century of regime change from Hawaii to Iraq*, Macmillan. 4.4

Kreps, S. E. and Fuhrmann, M. (2011) Attacking the atom: Does bombing nuclear facilities affect proliferation?, *Journal of Strategic Studies*, **34**, 161–187. 4.4

Kydd, A. H. and McManus, R. W. (2017) Threats and assurances in crisis bargaining, *Journal of conflict resolution*, **61**, 325–348. 3

Lanoszka, A. (2016) Russian hybrid warfare and extended deterrence in eastern europe, *International Affairs*, **92**, 175–195. (document)

Leventoğlu, B. and Slantchev, B. L. (2007) The armed peace: a punctuated equilibrium theory of war, *American Journal of Political Science*, **51**, 755–771. 4.5

Lyall, J. (2013) Airpower and coercion in counterinsurgency wars: Evidence from afghanistan, Tech. rep., Working Paper. 4.4

Mazarr, M. J. (2015a) *Mastering the Gray Zone: Understanding a Changing Era of Conflict*, United States Army War College. 4.5

Mazarr, M. J. (2015b) Struggle in the gray zone, *War on the Rocks*. (document), 4.5

McCormack, D. and Pascoe, H. (2017) Sanctions and preventive war, *Journal of Conflict Resolution*, **61**, 1711–1739. (document), 1, 4

Mearsheimer, J. J. *et al.* (2001) *The tragedy of great power politics*, WW Norton & Company. (document)

Meirowitz, A., Morelli, M., Ramsay, K. W. and Squintani, F. (2019) Dispute resolution institutions and strategic militarization, *Journal of Political Economy*, **127**, 378–418. 4.5

Meirowitz, A., Sartori, A. E. *et al.* (2008) Strategic uncertainty as a cause of war, *Quarterly Journal of Political Science*, **3**, 327–352. (document), 4.5

Meirowitz, A. *et al.* (2006) Designing institutions to aggregate preferences and information, *Quarterly Journal of political science*, **1**, 373–392. (document)

Milgrom, P. and Segal, I. (2002) Envelope theorems for arbitrary choice sets, *Econometrica*, **70**, 583–601. B.6

Myerson, R. B. (1979) Incentive compatibility and the bargaining problem, *Econometrica*, pp. 61–73. 2.2

Niou, E. M. and Ordeshook, P. C. (1990) Stability in anarchic international systems, *American Political Science Review*, **84**, 1207–1234. (document)

Opall-Rome, B. (2018) Declassified: How an israeli operation derailed syria's nuclear weapons drive, *Defense News*. 1

Owen, D. (2021) Time for a fully integrated dual-fleet force model, *War on the Rocks*. 4.4

Pape, R. A. (1996) *Bombing to win: Air power and coercion in war*, Cornell University Press. 4.4

Paul, J. A. (2003) Oil companies in iraq: a century of rivalry and war, in *Global Policy Forum*, pp. 25–26. 4.4

Pietrucha, M. and Renken, J. (2019) Uplifted: The case for small tactical airlift, *War on the Rocks*. 4.4

Post, A. (2019) Flying to fail: Costly signals and air power in crisis bargaining, *Journal of Conflict Resolution*, **63**, 869–895. 21

Powell, R. (1989) Nuclear deterrence and the strategy of limited retaliation, *The American Political Science Review*, pp. 503–519. (document)

Powell, R. (1999) *In the shadow of power: States and strategies in international politics*, Princeton University Press. (document)

Powell, R. (2004) Bargaining and learning while fighting, *American Journal of Political Science*, **48**, 344–361. 4.5

Powell, R. (2006) War as a commitment problem, *International organization*, **60**, 169–203. 1

Powell, R. (2015) Nuclear brinkmanship, limited war, and military power, *International Organization*, **69**, 589–626. (document)

Schram, P. (2020) When capabilities backfire, *Manuscript*. (document), 2, 4

Schram, P. (Forthcoming) Hassling: How states prevent preventive wars, *American Journal of Political Science*. (document), 1, 4, 4.4, 4.5

Schultz, K. A. (1999) Do democratic institutions constrain or inform? contrasting two institutional perspectives on democracy and war, *International Organization*, **53**, 233–266. (document)

Schultz, K. A. (2010) The enforcement problem in coercive bargaining: Interstate conflict over rebel support in civil wars, *International Organization*, **64**, 281–312. (document), 1, 4.4, 4.5

Shachtman, N. (2011) The secret history of iraq's invisible war, *Wired, April*, **11**. 24

Slantchev, B. L. (2003a) The power to hurt: Costly conflict with completely informed states, *American Political Science Review*, **97**, 123–133. 4.5

Slantchev, B. L. (2003b) The principle of convergence in wartime negotiations, *American Political Science Review*, pp. 621–632. (document)

Smith, A. (1998) International crises and domestic politics, *American Political Science Review*, **92**, 623–638. 3

Smith, A. and Stam, A. C. (2004) Bargaining and the nature of war, *Journal of Conflict Resolution*, **48**, 783–813. 4.5

Spaniel, W. (2019) *Bargaining over the bomb: the successes and failures of nuclear negotiations*, Cambridge University Press. 1, 4.5

Spaniel, W. and Bils, P. (2018) Slow to learn: Bargaining, uncertainty, and the calculus of conquest, *Journal of Conflict Resolution*, **62**, 774–796. (document)

Spaniel, W. and Malone, I. (2019) The uncertainty trade-off: Reexamining opportunity costs and war, *International Studies Quarterly*, **63**, 1025–1034. (document)

Stokes, D. (2007) Blood for oil? global capital, counter-insurgency and the dual logic of american energy security, *Review of International Studies*, pp. 245–264. 4.4

Wagner, R. H. (2000) Bargaining and war, *American Journal of Political Science*, pp. 469–484. (document), 4.5

Waltz, K. N. (1979) *Theory of international politics*, Waveland Press. (document)

Wendt, A. (1992) Anarchy is what states make of it: the social construction of power politics, *International Organization*, **46**, 391–425. (document)

A Equilibrium and Direct Mechanism

Consider a game form $G = (\mathcal{B}_A, \mathcal{B}_D, g)$, where $g = (\pi^g, V_C^g, V_D^g)$. A pure strategy perfect Bayesian equilibrium of G consists of:

- C's transgression, $t^* \in \mathcal{T}$.
- Each type of D's response to C's transgression, $h^*(\theta, t)$, where $h^* : \Theta \times \mathcal{T} \to \mathcal{H}$.
- C's beliefs following D's response, the collection $(\mu_h)_{h \in \mathcal{H}}$, where each μ_h is a probability measure on Θ.
- C's bargaining action, $b_C^*(t, h)$, where $b_C^* : \mathcal{T} \times \mathcal{H} \to \mathcal{B}_A$.

- Each type of D's bargaining action, $b_D^*(\theta, t, h)$, where $b_D^* : \Theta \times \mathcal{T} \times \mathcal{H} \to \mathcal{B}_D$.

We can define an outcome-equivalent direct mechanism for D in terms of three functions:

- Hassling level:
$$h(\theta) = h^*(\theta, t^*).$$

- Whether bargaining prevails:
$$\pi(\theta) = \pi^g\left(b_C^*(t^*, h^*(\theta, t^*)), b_D^*(\theta, t^*, h^*(\theta, t^*))\right).$$

- D's spoils from bargaining in case bargaining prevails:
$$V_D(\theta) = V_D^g\left(t^*, h^*(\theta, t^*), b_C^*(t^*, h^*(\theta, t^*)), b_D^*(\theta, t^*, h^*(\theta, t^*))\right).$$

Treating C's equilibrium strategy as given, the expected utility to type θ of following the strategy of type θ' is:

$$\begin{aligned}
U_D(\theta' \,|\, \theta) &= u_G^g\left(t^*, h^*(t^*, \theta'), b_C^*(t^*, h^*(\theta', t^*)), b_D^*(\theta', t^*, h^*(\theta', t^*)) \,|\, \theta\right) \\
&= \left[1 - \pi^g(b_C^*(t^*, h^*(\theta', t^*)), b_D^*(\theta', t^*, h^*(\theta', t^*)))\right] W_D(\theta) \\
&\quad + \pi^g(b_C^*(t^*, h^*(\theta', t^*)), b_D^*(\theta', t^*, h^*(\theta', t^*))) \\
&\quad \times \left[V_D^g(t^*, h^*(\theta', t^*), b_C^*(t^*, h^*(\theta', t^*)), b_D^*(\theta', t^*, h^*(\theta', t^*))) - K_D(h^*(\theta', t^*), \theta)\right]. \\
&= [1 - \pi(\theta')] W_D(\theta) + \pi(\theta') \left[V_D(\theta') - K_D(h(\theta'), \theta)\right].
\end{aligned}$$

Consequently, incentive compatibility is equivalent to

$$[1 - \pi(\theta)] W_D(\theta) + \pi(\theta) \left[V_D(\theta) - K_D(h(\theta), \theta)\right] \geq [1 - \pi(\theta')] W_D(\theta) + \pi(\theta') \left[V_D(\theta') - K_D(h(\theta'), \theta)\right]$$

for all $\theta, \theta' \in \Theta$.

B Proofs of Results in Text

B.1 Proof of Lemma 1 and Lemma 2

Both results depend on the following auxiliary result.

Lemma 12. *If $h = 0$ and $\theta < \theta'$,*

$$\Phi_D(\theta' \mid \theta) \leq \Phi_D(\theta \mid \theta) \leq \Phi_D(\theta \mid \theta') \leq \Phi_D(\theta' \mid \theta').$$

Proof. The first and third inequalities follow from (IC). The second follows because W_D is increasing and $h = 0$:

$$\begin{aligned}
\Phi_D(\theta \mid \theta) &= \pi(\theta)[V_D(\theta) - K_D(0, \theta)] + (1 - \pi(\theta))W_D(\theta) \\
&= \pi(\theta)V_D(\theta) + (1 - \pi(\theta))W_D(\theta) \\
&\leq \pi(\theta)V_D(\theta) + (1 - \pi(\theta))W_D(\theta') \\
&= \pi(\theta)[V_D(\theta) - K_D(0, \theta')] + (1 - \pi(\theta))W_D(\theta') \\
&= \Phi_D(\theta' \mid \theta).
\end{aligned}$$

□

We can now prove the results from the text.

Lemma 1. *If $h = 0$ and $\theta < \theta'$, then $\pi(\theta) \geq \pi(\theta')$.*

Proof. Lemma 12 implies

$$\Phi_D(\theta' \mid \theta') - \Phi_D(\theta' \mid \theta) \geq \Phi_D(\theta \mid \theta') - \Phi_D(\theta \mid \theta),$$

which is equivalent to

$$(1 - \pi(\theta'))[W_D(\theta') - W_D(\theta)] \geq (1 - \pi(\theta))[W_D(\theta') - W_D(\theta)].$$

As $W_D(\theta') > W_D(\theta)$, this in turn implies $\pi(\theta) \geq \pi(\theta')$. □

Lemma 2. *If $h = 0$ and $\theta < \theta'$, then $U_D(\theta) \leq U_D(\theta')$.*

Proof. Immediate from Lemma 12. □

B.2 Proof of Lemma 3

Lemma 3. *If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $U_D(\theta) \leq U_D(\theta')$.*

Proof. By (IC) and the fact that K_D is decreasing in θ, we have

$$U_D(\theta') \geq \underbrace{V_D(\theta) - K_D(h(\theta), \theta')}_{\Phi_D(\theta\,|\,\theta')} \geq V_D(\theta) - K_D(h(\theta), \theta) = U_D(\theta). \tag{5}$$

□

B.3 Proof of Corollary 1

Corollary 1. *If $\pi(\theta) = \pi(\theta') = 1$, $h(\theta) > 0$, and θ' has greater hassling effectiveness than θ, then $U_D(\theta) < U_D(\theta')$.*

Proof. Follows because the second inequality in Equation 5 is strict whenever $h(\theta) > 0$. □

B.4 Proof of Lemma 4

Lemma 4. *If $\pi(\theta) = \pi(\theta') = 1$ and $h(\theta) \leq h(\theta')$, then $V_D(\theta) \leq V_D(\theta')$. Furthermore, if $h(\theta) < h(\theta')$, then $V_D(\theta) < V_D(\theta')$.*

Proof. (IC) implies

$$V_D(\theta') - K_D(h(\theta'), \theta') \geq V_D(\theta) - K_D(h(\theta), \theta'),$$

which is equivalent to

$$V_D(\theta') - V_D(\theta) \geq K_D(h(\theta'), \theta') - K_D(h(\theta), \theta').$$

If $h(\theta) \leq h(\theta')$, then the RHS is non-negative, and the first claim follows. If $h(\theta) < h(\theta')$, then the RHS is strictly positive, and the second claim follows. □

B.5 Proof of Lemma 5

Lemma 5. *Assume (DD) holds. If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $h(\theta) \leq h(\theta')$.*

Proof. (IC) implies:

$$V_D(\theta) - K_D(h(\theta), \theta) \geq V_D(\theta') - K_D(h(\theta'), \theta),$$
$$V_D(\theta') - K_D(h(\theta'), \theta') \geq V_D(\theta) - K_D(h(\theta), \theta').$$

A rearrangement of terms gives

$$K_D(h(\theta'), \theta') - K_D(h(\theta), \theta') \leq V_D(\theta') - V_D(\theta) \leq K_D(h(\theta'), \theta) - K_D(h(\theta), \theta).$$

(DD) therefore implies $h(\theta) \leq h(\theta')$. $\square$

B.6 Proof of Proposition 1

We first state a helpful lemma.

Lemma 13. *Assume* (BV) *holds. For all* $\theta, \theta' \in \Theta$, Φ_D *is differentiable with respect to* θ, *and*

$$\frac{\partial \Phi_D(\theta' \,|\, \theta)}{\partial \theta} = (1 - \pi(\theta')) \frac{dW_D(\theta)}{d\theta} - \pi(\theta') \frac{\partial K_D(h(\theta'), \theta)}{\partial \theta}.$$

Proof. The existence of $\partial \Phi_D / \partial \theta$ follows from (BV). The expression in the lemma then follows immediately from the definition of Φ_D. $\square$

We then rely on standard mechanism design arguments to establish the proposition.

Proposition 1. *Assume* (BV) *holds. For all* $\theta_0 \in \Theta$,

$$U_D(\theta_0) = U_D(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_0} (1 - \pi(\theta)) \frac{dW_D(\theta)}{d\theta} \, d\theta - \int_{\underline{\theta}}^{\theta_0} \pi(\theta) \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h = h(\theta)} d\theta. \tag{3}$$

Proof. (IC) implies $U_D(\theta) = \sup_{\theta' \in \Theta} \Phi_D(\theta' \,|\, \theta)$ for all $\theta \in \Theta$. Therefore, by Milgrom and Segal (2002, Theorem 1),

$$\frac{dU_D(\theta)}{d\theta} = \left. \frac{\partial \Phi_D(\theta' \,|\, \theta)}{\partial \theta} \right|_{\theta' = \theta}$$

at each point where U_D is differentiable. Furthermore, (BV) implies Φ_D is Lipschitz continuous in θ. The claim then follows from Lemma 13 and Milgrom and Segal (2002, Corollary 1). $\square$

B.7 Proof of Lemma 6

Lemma 6. *Assume* θ *improves hassling effectiveness. If* $\theta < \theta'$, *then* $U_D(\theta) \leq U_D(\theta')$.

Proof. When θ improves hassling effectiveness, (IC) implies

$$\begin{aligned}
U_D(\theta') &= (1-\pi(\theta'))W_D(\theta') + \pi(\theta')[V_D(\theta') - K_D(h(\theta'),\theta')] \\
&\geq (1-\pi(\theta))W_D(\theta') + \pi(\theta)[V_D(\theta) - K_D(h(\theta),\theta')] \\
&\geq (1-\pi(\theta))W_D(\theta) + \pi(\theta)[V_D(\theta) - K_D(h(\theta),\theta)] \\
&= U_D(\theta).
\end{aligned}$$

□

B.8 Proof of Lemma 7

Lemma 7. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and $U_D(\theta) = U_D(\theta')$, then $\pi(\theta) = 1$ and $h(\theta) = 0$.*

Proof. We will prove $U_D(\theta) < U_D(\theta')$ in all other cases. If $\pi(\theta) = 0$, then (VA) implies

$$U_D(\theta') \geq W_D(\theta') > W_D(\theta) = U_D(\theta).$$

If $\pi(\theta) = 1$ and $h(\theta) > 0$, then (IC) implies

$$U_D(\theta') \geq V_D(\theta) - K_D(h(\theta),\theta') > V_D(\theta) - K_D(h(\theta),\theta) = U_D(\theta).$$

□

B.9 Proof of Lemma 8

Lemma 8. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and (WURI) holds, then $\pi(\theta) \geq \pi(\theta')$. If $\theta < \theta'$ and (SURI) holds, then $\pi(\theta) \leq \pi(\theta')$.*

Proof. We will prove the claims by contraposition. Let $\theta < \theta'$, and suppose $\pi(\theta) < \pi(\theta')$ (i.e., $\pi(\theta) = 0$ and $\pi(\theta') = 1$). We want to prove that this implies (WURI) does not hold. Note that (VA) implies

$$V_D(\theta') - K_D(h(\theta'),\theta') \geq W_D(\theta'),$$

while (IC) implies

$$W_D(\theta) \geq V_D(\theta') - K_D(h(\theta'),\theta).$$

Combining these gives

$$W_D(\theta) + K_D(h(\theta'),\theta) \geq V_D(\theta') \geq W_D(\theta') + K_D(h(\theta'),\theta'),$$

which in turn implies

$$W_D(\theta') - W_D(\theta) \leq K_D(h(\theta'), \theta) - K_D(h(\theta'), \theta).$$

Because $\pi(\theta') = 1$, this means (WURI) cannot hold, establishing the first claim of the lemma. An analogous argument establishes that (SURI) cannot hold if $\pi(\theta) > \pi(\theta')$ (i.e., $\pi(\theta) = 1$ and $\pi(\theta') = 0$). $\square$

B.10 Proof of Lemma 9

Lemma 9. *Assume θ improves hassling effectiveness, (DD) holds, and $\max \mathcal{H} = \bar{h} < \infty$. If $W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta')$ for all $\theta, \theta' \in \Theta$ such that $\theta < \theta'$, then (WURI) holds.*

Proof. For all $h < \bar{h}$ and $\theta < \theta'$, (DD) implies

$$K_D(\bar{h}, \theta') - K_D(h, \theta') < K_D(\bar{h}, \theta) - K_D(h, \theta),$$

which is equivalent to

$$K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta') > K_D(h, \theta) - K_D(h, \theta').$$

Therefore, under the hypothesis of the lemma, we have

$$W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta') \geq K_D(h, \theta) - K_D(h, \theta')$$

for all $h \in \mathcal{H}$, which implies (WURI). $\square$

B.11 Proof of Lemma 10

Lemma 10. *Assume θ degrades hassling effectiveness. If $\theta < \theta'$, then $\pi(\theta) \geq \pi(\theta')$.*

Proof. For a proof by contradiction, suppose $\theta < \theta'$ and $\pi(\theta) < \pi(\theta')$. This implies $\pi(\theta) = 0$ and $\pi(\theta') = 1$. (VA) implies

$$V_D(\theta') - K_D(h(\theta'), \theta') \geq W_D(\theta').$$

(IC), combined with the assumption that θ degrades hassling effectiveness, implies

$$W_D(\theta) \geq V_D(\theta') - K_D(h(\theta'), \theta) \geq V_D(\theta') - K_D(h(\theta'), \theta').$$

Combining these inequalities gives $W_D(\theta) \geq W_D(\theta')$, a contradiction. $\square$

B.12 Proof of Lemma 11

We first prove a lemma demonstrating equivalence between the characterization of V_D in Lemma 11 and that of U_D in Proposition 1.

Lemma 14. *Assume (BV) holds and consider any direct mechanism (h, π, V_D). Suppose there exists $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$. Then Equation 3 holds for all $\theta_0 \in \Theta$ if and only if $U_D(\hat{\theta}) = W_D(\hat{\theta})$ and Equation 4 holds for all $\theta_0 < \hat{\theta}$.*

Proof. To prove the "only if" direction, suppose Equation 3 holds for all $\theta_0 \in \Theta$. This implies U_D is (absolutely) continuous, so $U_D(\hat{\theta}) = \lim_{\theta \to \hat{\theta}+} U_D(\theta) = \lim_{\theta \to \hat{\theta}+} W_D(\theta) = W_D(\hat{\theta})$. Then, for all $\theta_0 < \hat{\theta}$, Equation 3 implies

$$U_D(\theta_0) - U_D(\hat{\theta}) = \int_{\theta_0}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta.$$

Substituting $U_D(\theta_0) = V_D(\theta_0) - K_D(h(\theta_0), \theta_0)$ and $U_D(\hat{\theta}) = W_D(\hat{\theta})$ into this expression and rearranging terms yields Equation 4.

To prove the "if" direction, suppose $U_D(\hat{\theta}) = W_D(\hat{\theta})$ and Equation 4 holds for all $\theta_0 < \hat{\theta}$. We will now take an arbitrary $\theta' \in \Theta$ and show that it satisfies Equation 3. If $\theta' < \hat{\theta}$, then

$$\begin{aligned} U_D(\theta') - U_D(\underline{\theta}) &= V_D(\theta') - V_D(\underline{\theta}) + K_D(h(\underline{\theta}), \underline{\theta}) - K_D(h(\theta'), \theta') \\ &= \int_{\theta'}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta - \int_{\underline{\theta}}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta \\ &= - \int_{\underline{\theta}}^{\theta'} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta. \end{aligned}$$

If $\theta' \geq \hat{\theta}$, then

$$\begin{aligned}
U_D(\theta') - U_D(\underline{\theta}) &= W_D(\theta') - V_D(\underline{\theta}) + K_D(h(\underline{\theta}), \underline{\theta}) \\
&= W_D(\hat{\theta}) + \int_{\hat{\theta}}^{\theta'} \frac{dW_D(\theta)}{d\theta} \, d\theta - V_D(\underline{\theta}) + K_D(h(\underline{\theta}), \underline{\theta}) \\
&= V_D(\underline{\theta}) - K_D(h(\underline{\theta}), \underline{\theta}) - \int_{\underline{\theta}}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta \\
&\quad + \int_{\hat{\theta}}^{\theta'} \frac{dW_D(\theta)}{d\theta} \, d\theta - V_D(\underline{\theta}) + K_D(h(\underline{\theta}), \underline{\theta}) \\
&= \int_{\hat{\theta}}^{\theta'} \frac{dW_D(\theta)}{d\theta} - \int_{\underline{\theta}}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta,
\end{aligned}$$

where the penultimate equality follows from a rearrangement of Equation 4 at $\theta_0 = \underline{\theta}$. Therefore, Equation 3 is satisfied for all $\theta' \in \Theta$. □

The result is immediate from this lemma and from earlier results.

Lemma 11. *Assume θ degrades hassling effectiveness.*

(a) There exists $\hat{\theta} \in \Theta$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$.

(b) If (BV) holds and $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$, then $U_D(\hat{\theta}) = W_D(\hat{\theta})$.

(c) If (BV) holds, then for all $\theta_0 < \hat{\theta}$,

$$V_D(\theta_0) = W_D(\hat{\theta}) + K_D(h(\theta_0), \theta_0) + \int_{\theta_0}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta. \tag{4}$$

(d) U_D is non-increasing on $[\underline{\theta}, \hat{\theta}]$ and strictly increasing on $[\hat{\theta}, \bar{\theta}]$.

Proof. (a) follows from Lemma 10. (b) and (c) follow from Lemma 14. (d) follows from Lemma 3. □

B.13 Proof of Proposition 2

Proposition 2. *Assume θ degrades hassling effectiveness and (BV) and (DD) hold. If the direct mechanism (h, π, V_D) satisfies conditions (a)–(c) of Lemma 11 with $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ and h*

is absolutely continuous and non-increasing, then this direct mechanism satisfies (IC) *and* (VA)*.*

Proof. As a preliminary, note that because K_D is Lipschitz and h is absolutely continuous, $K_D(h(\theta), \theta)$ is absolutely continuous when viewed as a function of θ (**?**, Corollary 3.3.9). Consequently, V_D is absolutely continuous and thus differentiable almost everywhere on $[\underline{\theta}, \hat{\theta})$.

Now take any $\theta, \theta' \in \Theta$. If $\theta' < \hat{\theta}$, then

$$\begin{aligned}
\Phi_D(\theta' \,|\, \theta) &= V_D(\theta') - K_D(h(\theta'), \theta) \\
&= W_D(\hat{\theta}) + K_D(h(\theta'), \theta') - K_D(h(\theta'), \theta) + \int_{\theta'}^{\hat{\theta}} \frac{\partial K_D(h, \theta'')}{\partial d\theta''}\bigg|_{h=h(\theta'')} d\theta''.
\end{aligned}$$

Therefore, for almost all $\theta' < \hat{\theta}$, we have

$$\begin{aligned}
\frac{\partial \Phi_D(\theta' \,|\, \theta)}{\partial \theta'} &= \frac{\partial K_D(h(\theta'), \theta')}{\partial h} \frac{dh(\theta')}{d\theta'} + \frac{\partial K_D(h(\theta'), \theta')}{\partial \theta'} \\
&\quad - \frac{\partial K_D(h(\theta'), \theta)}{\partial h} \frac{dh(\theta')}{\theta'} - \frac{\partial K_D(h(\theta'), \theta')}{\partial \theta'} \\
&= \underbrace{\frac{dh(\theta')}{\theta'}}_{\leq 0} \left[\frac{\partial K_D(h(\theta'), \theta')}{\partial h} - \frac{\partial K_D(h(\theta'), \theta)}{\partial h} \right].
\end{aligned}$$

Because θ degrades hassling effectiveness, (DD) implies that the term in brackets is non-negative if $\theta \leq \theta'$ and non-positive if $\theta \geq \theta'$. Next, notice that

$$\begin{aligned}
&\lim_{\theta' \to \hat{\theta}^-} \Phi_D(\theta' \,|\, \theta) \\
&\quad = \lim_{\theta' \to \hat{\theta}^-} \left[W_D(\hat{\theta}) + K_D(h(\theta'), \theta') - K_D(h(\theta'), \theta) + \int_{\theta'}^{\hat{\theta}} \frac{\partial K_D(h, \theta'')}{\partial d\theta''}\bigg|_{h=h(\theta'')} d\theta'' \right] \\
&\quad = W_D(\hat{\theta}) + K_D(h(\hat{\theta}), \hat{\theta}) - K_D(h(\hat{\theta}), \theta).
\end{aligned}$$

Therefore, if $\theta \leq \hat{\theta}$, then

$$\lim_{\theta' \to \hat{\theta}^-} \Phi_D(\theta' \,|\, \theta) \geq W_D(\hat{\theta}) \geq W_D(\theta) = \lim_{\theta' \to \hat{\theta}^+} \Phi_D(\theta' \,|\, \theta).$$

Conversely, if $\theta \geq \hat{\theta}$, then

$$\lim_{\theta' \to \hat{\theta}^-} \Phi_D(\theta' \mid \theta) \leq W_D(\hat{\theta}) \leq W_D(\theta) = \lim_{\theta' \to \hat{\theta}^+} \Phi_D(\theta' \mid \theta).$$

Finally, we have $\Phi_D(\theta' \mid \theta) = W_D(\theta)$ for all $\theta' > \hat{\theta}$. Altogether, these findings imply $\Phi_D(\theta' \mid \theta)$ is non-decreasing in θ' if $\theta' \in [\underline{\theta}, \theta]$ and non-increasing in θ' if $\theta' \in [\theta, \bar{\theta}]$. Therefore, (IC) holds. These results also imply $U_D(\theta) \geq \lim_{\theta \to \hat{\theta}^-} \Phi_D(\theta' \mid \theta) = W_D(\theta)$ for all $\theta \leq \hat{\theta}$, so (VA) also holds. □

B.14 Proof of Proposition 3

Proposition 3. *If $\hat{W}_C + W_D(\bar{\theta}) \leq 1$, then there is a flexible-response crisis bargaining game form with voluntary agreements that has an always peaceful equilibrium.*

Proof. We prove the result by construction. Consider the following extensive form game:

1. C chooses $t \in \mathcal{T}$. If $t > 0$, then the game ends in war. Otherwise, if $t = 0$, the game moves to the next step.

2. D chooses $h \in \mathcal{H}$. If $h > 0$, then the game ends in war. If $h = 0$, then C receives $1 - W_D(\bar{\theta})$ and D receives $W_D(\bar{\theta})$.

This game has voluntary agreements, as each player has a strategy available that guarantees them their war payoff. Clearly it is a best response for every type of D to choose $h = 0$ if its move is reached (strictly so for each $\theta < \bar{\theta}$). Moving up the game tree, if every type of D's strategy is to choose $h = 0$, then C's expected utility from choosing $t = 0$ is $1 - W_D(\bar{\theta})$, compared to $\hat{W}_C$ for any $t > 0$. Therefore, under the condition of the proposition, it is a best response for C to choose $t = 0$. □

B.15 Proof of Corollary 3

Corollary 3. *If $W_C(\theta) = p - c_C$ and $W_D(\theta) = 1 - p - c_D(\theta)$, where $c_D : \Theta \to \mathbb{R}_+$, then there is a flexible-response crisis bargaining game form with voluntary agreements that has an always peaceful equilibrium.*

Proof. The result follows from Proposition 3, as $\hat{W}_C + W_D(\bar{\theta}) = 1 - c_C - c_D(\bar{\theta}) < 1$. □

C Extension

We extend the baseline model to allow transgressions and hassling to affect war payoffs. The basic structure of the analysis follows the baseline model, except with an additional option for D to preempt C's choice of transgressions.

In the interaction, C first selects transgressions $t \in \mathcal{T} \subseteq \mathbb{R}_+$. D may then choose to proceed directly to war, which preempts C's choice of t (in effect forcing $t = 0$) and precludes D from responding with hassling. If D chooses not to go to war at this point, the game then proceeds as in the baseline model. C and D each choose bargaining actions $b_C \in \mathcal{B}_C$ and $b_D \in \mathcal{B}_D$, and D chooses hassling $h \in \mathcal{H} \subseteq \mathbb{R}_+$.

We now write each player's war payoff as a function of function of the flexible responses and D's type. C's war payoff is denoted $W_C(t, h, \theta)$, which is weakly increasing in t, weakly decreasing in h, and weakly decreasing in θ. Following Fey and Ramsay (2011), we say the game has *independent values* if W_C is everywhere constant in θ. D's war payoff is denoted $W_D(t, h, \theta)$, which is weakly decreasing in t, weakly increasing in h, and strictly increasing in θ. We assume throughout the extension that $\mathcal{T}$ and $\mathcal{H}$ are compact, that W_C and K_C are continuous in t, and that W_D and K_D are continuous in h. These assumptions ensure that the players' expected utilities attain their maximum with respect to the player's own transgression or hassling.

We must update our definition of a direct mechanism to account for the possibility of war that preempts transgressions. A direct mechanism now consists of four functions expressing the outcome for each type of D:

- Their hassling level, $h(\theta)$.
- Whether the game ends immediately in war, $\omega_0(\theta)$.
- Whether the game ends in war after hassling, $\omega_1(\theta)$.
- Whether a bargained outcome prevails, $\pi(\theta)$.
- Their settlement value in case of a bargained resolution, $V_D(\theta)$.

Note that $\omega_0 + \omega_1 + \pi = 1$ by construction. As in the baseline setting, we will focus on pure strategy equilibria of deterministic models, so that each of these will take a value of exactly 0 or 1. Letting t^* denote the equilibrium transgression choice, the expected utility for type

θ from adopting the strategy of type θ' is now

$$\begin{aligned}\Phi_D(\theta' \mid \theta) = {} & \omega_0(\theta')W_D(0,0,\theta) \\ & + \omega_1(\theta')[W_D(t^*, h(\theta'), \theta) - K_D(h(\theta'), \theta)] \\ & + \pi(\theta')[V_D(\theta') - K_D(h(\theta'), \theta)].\end{aligned}$$

Though the expression for Φ_D has changed, incentive compatibility still requires $\Phi_D(\theta \mid \theta) \geq \Phi_D(\theta' \mid \theta)$ for all $\theta, \theta' \in \Theta$, so we will continue to use (IC) to refer to this condition. By contrast, the dependence of war payoffs on transgressions and hassling, along with the "short circuit" option for D to start war before t takes effect, introduces additional complications in voluntary agreements. For comprehensibility, we break voluntary agreements in the extended model into two conditions. First, any equilibrium outcome must leave D no worse off than the short-circuit option:

$$U_D(\theta) \geq W_D(0,0,\theta) \qquad \text{for all } \theta \in \Theta. \tag{VA$'$}$$

Second, any equilibrium outcome must leave D no worse off than if D allowed C's transgressions to proceed and then fought with an optimal hassling response:

$$U_D(\theta) \geq \max_{h \in \mathcal{H}} [W_D(t^*, h, \theta) - K_D(h, \theta)] \qquad \text{for all } \theta \in \Theta. \tag{VA$''$}$$

C.1 Equilibrium Properties

We now characterize, where possible, general properties of equilibria in games where transgressions and hassling affect war payoffs. We are particularly interested in finding analogues of our baseline results.

C.1.1 Montonicity of Utility

In the baseline setting, if two types of D both end up at peace, then the more effective one has at least as great a payoff (Lemma 3). That result carries over here. In fact, we can strengthen it: if the less effective of two types settles in equilibrium, then the more effective one has at least a great a payoff regardless of whether it fights or settles.

Lemma 15. *If $\pi(\theta) = 1$ and θ' has greater hassling effectiveness than θ, then $U_D(\theta) \leq U_D(\theta')$.*

Proof. We have

$$\begin{aligned} U_D(\theta') &\geq V_D(\theta) - K_D(h(\theta), \theta') \\ &\geq V_D(\theta) - K_D(h(\theta), \theta) \\ &= U_D(\theta), \end{aligned} \tag{6}$$

where the first inequality follows from (IC) and the second follows from the definition of hassling effectiveness. □

We can strengthen this result similar to how Corollary 1 strengthens Lemma 3 in the baseline setting. Between two types that both proceed to the hassling stage, if the less effective one chooses $h > 0$ and settles, then the more effective one has strictly greater utility in equilibrium.

Corollary 4. *If $\pi(\theta) = 1$, $h(\theta) > 0$, and θ' has greater hassling effectiveness than θ, then $U_D(\theta) < U_D(\theta')$.*

Proof. Follows because the second inequality in Equation 6 is strict whenever $h(\theta) > 0$. □

Without additional conditions, we cannot prove an analogous result for when the less effective type goes to war. The monotonicity result carries over straightforwardly if θ enhances hassling effectiveness, as then the higher type benefits along multiple dimensions. (See below for an even more general result in this case.) But if θ degrades hassling effectiveness, then there are countervailing effects. We cannot in general guarantee that the more effective (lower) type will do better overall, as its lower costs of hassling may be offset by its lower war payoff. A sufficient condition for monotonicity is for θ to affect the costs of hassling more than the war payoff—an analogue of the (SURI) condition from the baseline setting, which we now restate for the extended model:

$$\begin{aligned} W_D(t^*, h(\theta'), \theta') - W_D(t^*, h(\theta'), \theta) &< K_D(h(\theta'), \theta') - K_D(h(\theta'), \theta) \\ &\text{for all } \theta, \theta' \in \Theta \text{ such that } \theta < \theta' \text{ and } \omega_1(\theta') = 1. \end{aligned} \tag{SURI†}$$

As in the baseline setting, it is impossible to derive a sufficient condition for (SURI†) from model primitives, as the LHS is positive by construction while the RHS equals zero at $h(\theta') = 0$. In any case, this condition is sufficient to guarantee that payoffs decrease with type (and thus increase with hassling effectiveness) when type degrades effectiveness and the less effective type goes to war.

Lemma 16. *Assume θ degrades hassling effectiveness and* (SURI†) *holds. If $\omega_1(\theta') = 1$ and $\theta < \theta'$, then $U_D(\theta) \geq U_D(\theta')$.*

Proof. We have

$$\begin{aligned} U_D(\theta) &\geq W_D(t^*, h(\theta'), \theta) - K_D(h(\theta'), \theta) \\ &\geq W_D(t^*, h(\theta'), \theta') - K_D(h(\theta'), \theta') \\ &= U_D(\theta'), \end{aligned}$$

where the first inequality follows from (IC) and the second follows from (SURI†). □

The conclusion of Lemma 16 holds strictly if $h(\theta') > 0$. The logic is the same as that of Corollary 4, so we omit the tedious statement of this result and its proof.

We can extend the monotonicity result much further if θ enhances hassling effectiveness. In an analogue of Lemma 6, we find that U_D is always non-decreasing in θ in this case.

Lemma 17. *Assume θ improves hassling effectiveness. If $\theta < \theta'$, then $U_D(\theta) \leq U_D(\theta')$.*

Proof. When θ improves hassling effectiveness, (IC) implies

$$\begin{aligned} U_D(\theta') \geq\ & \omega_0(\theta) W_D(0, 0, \theta') + \omega_1(\theta) W_D(t^*, h(\theta), \theta') + \pi(\theta) V_D(\theta) \\ & - (1 - \omega_0(\theta)) K_D(h(\theta), \theta') \\ \geq\ & \omega_0(\theta) W_D(0, 0, \theta) + \omega_1(\theta) W_D(t^*, h(\theta), \theta) + \pi(\theta) V_D(\theta) \\ & - (1 - \omega_0(\theta)) K_D(h(\theta), \theta) \\ =\ & U_D(\theta). \end{aligned}$$

□

Once again, the conclusion here holds unless the weaker type settles and chooses zero hassling.

Corollary 5. *Assume θ improves hassling effectiveness. Let $\theta < \theta'$. If $\pi(\theta) = 0$ or $h(\theta) > 0$, then $U_D(\theta) < U_D(\theta')$.*

Proof. Holds because the final inequality in the proof of Lemma 17 is strict as long as $\omega_0(\theta) = 1$, $\omega_1(\theta) = 1$, or $h(\theta) > 0$. □

C.1.2 Monotonicity of War Probability

When θ improves hassling effectiveness. In the baseline model, there is no general relationship between private type and the occurrence of conflict when θ improves hassling effectiveness. The probability of war increases with type when θ has a stronger effect on war payoffs than on the cost of hassling; it decreases with type in the reverse case (Lemma 8). To obtain analogous results for the extended model, we must extend the Equation WURI and Equation SURI conditions to this setting. Here the war utility is relatively increasing if

$$\begin{gathered} W_D(0,0,\theta') - W_D(0,0,\theta) > K_D(h(\theta'),\theta) - K_D(h(\theta'),\theta') \\ \text{for all } \theta,\theta' \in \Theta \text{ such that } \theta < \theta' \text{ and } \pi(\theta') = 1. \end{gathered} \tag{WURI$'$}$$

The settlement utility is relatively increasing if

$$\begin{gathered} W_D(0,0,\theta') - W_D(0,0,\theta) < K_D(h(\theta),\theta) - K_D(h(\theta),\theta') \\ \text{for all } \theta,\theta' \in \Theta \text{ such that } \theta < \theta' \text{ and } \pi(\theta) = 1. \end{gathered} \tag{SURI$'$}$$

If type affects war payoffs more strongly than the costs of hassling (WURI′), then the likelihood of choosing immediate war over settlement increases as D becomes stronger.

Lemma 18. *Assume θ improves hassling effectiveness and (WURI′) holds. If $\pi(\theta') = 1$, then $\omega_0(\theta) = 0$ for all $\theta < \theta'$.*

Proof. For a proof by contradiction, suppose $\omega_0(\theta) = 1$ for some $\theta < \theta'$. Then (IC) implies

$$W_D(0,0,\theta) \geq V_D(\theta') - K_D(h(\theta'),\theta);$$
$$V_D(\theta') - K_D(h(\theta'),\theta') \geq W_D(0,0,\theta').$$

In combination, these imply

$$W_D(0,0,\theta') - W_D(0,0,\theta) \leq K_D(h(\theta'),\theta) - K_D(h(\theta'),\theta'),$$

contradicting (WURI′). □

The opposite is true when type affects the costs of hassling more strongly (SURI′): higher types are less likely to choose immediate war over settlement.

Lemma 19. *Assume θ improves hassling effectiveness and (SURI′) holds. If $\omega_0(\theta') = 1$, then $\pi(\theta) = 0$ for all $\theta < \theta'$.*

Proof. For a proof by contradiction, suppose $\pi(\theta) = 1$ for some $\theta < \theta'$. Then (IC) implies

$$V_D(\theta) - K_D(h(\theta), \theta) \geq W_D(0, 0, \theta);$$
$$W_D(0, 0, \theta') \geq V_D(\theta) - K_D(h(\theta), \theta').$$

In combination, these imply

$$W_D(0, 0, \theta') - W_D(0, 0, \theta) \geq K_D(h(\theta), \theta) - K_D(h(\theta), \theta'),$$

contradicting (SURI′). □

When θ degrades hassling effectiveness. In the baseline model, we find that the likelihood of conflict increases with D's private type as long as θ degrades hassling effectiveness (Lemma 10). We obtain a similarly unconditional result about immediate conflict in the extended model.

Lemma 20. *Assume θ degrades hassling effectiveness. If $\pi(\theta') = 1$, then $\omega_0(\theta) = 0$ for all $\theta < \theta'$.*

Proof. For a proof by contradiction, suppose $\omega_0(\theta) = 1$ for some $\theta < \theta'$. Then (IC) implies

$$W_D(0, 0, \theta) \geq V_D(\theta') - K_D(h(\theta'), \theta);$$
$$V_D(\theta') - K_D(h(\theta'), \theta') \geq W_D(0, 0, \theta').$$

Together these imply

$$W_D(0, 0, \theta) + K_D(h(\theta'), \theta) \geq W_D(0, 0, \theta') + K_D(h(\theta'), \theta').$$

This is a contradiction, as $W_D(0, 0, \theta) < W_D(0, 0, \theta')$ and $K_D(h(\theta'), \theta) \leq K_D(h(\theta'), \theta')$. □

C.1.3 Montonicity of Hassling

In the baseline setting, if we compare two types that both settle in equilibrium, then the more effective one hassles more—as long as the decreasing differences condition (DD) on hassling costs is satisfied (Lemma 5). The same is true in the extended model for two types that both settle, and the proof carries over unchanged. We obtain a similar result for types that fight after hassling, so long as we impose an analogous single crossing condition on the relationship between types and war payoffs. We say the war payoff function W_D has *increasing differences* in h and θ if

$$\begin{aligned}&\theta' \text{ has greater hassling effectiveness than } \theta \Rightarrow \\ &\quad W_D(t^*, h', \theta') - W_D(t^*, h, \theta') > W_D(t^*, h', \theta) - W_D(t^*, h, \theta) \text{ for all } h < h'.\end{aligned} \tag{ID}$$

Substantively, this means the marginal effect of hassling on war payoffs is greater for types that have lower hassling costs. We state the requirement at equilibrium transgressions t^* in order to obtain the result under the weakest possible conditions. However, to make (ID) into a condition on model primitives, it would suffice to replace t^* with t throughout and to specify that the condition hold for all $t \in \mathcal{T}$.

Lemma 21. *Assume* (DD) *and* (ID) *hold. If* $\omega_1(\theta) = \omega_1(\theta') = 1$ *and* θ' *has greater hassling effectiveness than* θ*, then* $h(\theta) \leq h(\theta')$*.*

Proof. For a proof by contradiction, suppose $h(\theta) > h(\theta')$. (IC) implies:

$$\begin{aligned}W_D(t^*, h(\theta), \theta) - K_D(h(\theta), \theta) &\geq W_D(t^*, h(\theta'), \theta) - K_D(h(\theta'), \theta);\\ W_D(t^*, h(\theta'), \theta') - K_D(h(\theta'), \theta') &\geq W_D(t^*, h(\theta), \theta') - K_D(h(\theta), \theta').\end{aligned}$$

Combining these and rearranging terms gives

$$\begin{aligned}&[W_D(t^*, h(\theta), \theta) - W_D(t^*, h(\theta'), \theta)] - [W_D(t^*, h(\theta), \theta') - W_D(t^*, h(\theta'), \theta')]\\ &\qquad\qquad \geq [K_D(h(\theta), \theta) - K_D(h(\theta'), \theta)] - [K_D(h(\theta), \theta') - K_D(h(\theta'), \theta')].\end{aligned}$$

But (ID) implies that the LHS of this condition is strictly negative, while (DD) implies that the RHS is strictly positive, a contradiction. □

To completely analyze the monotonicity of h, we need to compare across types that settle and those that fight after hassling. We can only obtain this kind of characterization under particular conditions. First we consider the case where θ improves hassling effectiveness. If we compare a stronger type that settles to a weaker type that fights after hassling, we find that the stronger type must hassle weakly more than the weaker one.

Lemma 22. *Assume* θ *improves hassling effectiveness and* (DD) *holds. If* $\theta < \theta'$, $\omega_1(\theta) = 1$, *and* $\pi(\theta') = 1$*, then* $h(\theta) \leq h(\theta')$*.*

Proof. For a proof by contradiction, suppose $h(\theta) > h(\theta')$. (IC) implies

$$\begin{aligned} W_D(t^*, h(\theta), \theta) - K_D(h(\theta), \theta) \geq V_D(\theta') - K_D(h(\theta'), \theta); \\ V_D(\theta') - K_D(h(\theta'), \theta') \geq W_D(t^*, h(\theta), \theta') - K_D(h(\theta), \theta'). \end{aligned} \tag{7}$$

As θ' has greater hassling effectiveness than θ, (DD) implies

$$K_D(h(\theta), \theta') - K_D(h(\theta'), \theta') < K_D(h(\theta), \theta) - K_D(h(\theta'), \theta). \tag{8}$$

Combining the above inequalities, we have

$$\begin{aligned} W_D(t^*, h(\theta), \theta) &\geq V_D(\theta') + K_D(h(\theta), \theta) - K_D(h(\theta'), \theta) \\ &> V_D(\theta') + K_D(h(\theta), \theta') - K_D(h(\theta'), \theta') \\ &\geq W_D(t^*, h(\theta), \theta'), \end{aligned}$$

which implies $\theta > \theta'$, a contradiction. □

Next we obtain a mirror-image result for the case where θ degrades hassling effectiveness. Now if the stronger type settles and the weaker type fights after hassling, it is the weaker type (the one more effective at hassling) that must hassle weakly more.

Lemma 23. *Assume* θ *degrades hassling effectiveness and* (DD) *holds. If* $\theta < \theta'$*,* $\pi(\theta) = 1$*, and* $\omega_1(\theta') = 1$*, then* $h(\theta) \geq h(\theta')$*.*

Proof. For a proof by contradiction, suppose $h(\theta) < h(\theta')$. (IC) once again implies that the inequalities in Equation 7 hold. As θ has greater hassling effectiveness than θ', (DD) implies

$$K_D(h(\theta'), \theta) - K_D(h(\theta), \theta) < K_D(h(\theta'), \theta') - K_D(h(\theta), \theta'),$$

which is equivalent to Equation 8. Therefore, we obtain the same contradiction as in the proof of Lemma 22. □

C.2 Always Peaceful Equilibria

We now consider when there exists a game form with voluntary agreements that has a peaceful equilibrium. To begin, we will define bounds on each player's reservation value.

Specifically, these will be the maximum each player would expect to receive from fighting if the other player's transgressions or hassling were fixed at 0. These are thus worst-case bounds, as they pertain to the situation in which each player has the greatest possible incentive to deviate to fighting. For C, the worst-case reservation value is

$$\bar{R}_C = \sup_{\theta \in \Theta} \left\{ \max_{t \in \mathcal{T}} \{W_C(t, 0, \theta) - K_C(t)\} \right\}.$$

If $\Theta = [\underline{\theta}, \bar{\theta}]$, then the outer supremum will be attained at $\theta = \underline{\theta}$, as W_C is weakly decreasing in θ. For D, the worst-case reservation value is

$$\bar{R}_D = \sup_{\theta \in \Theta} \left\{ \max_{h \in \mathcal{H}} \{W_D(0, h, \theta) - K_D(h, \theta)\} \right\}.$$

If $\Theta = [\underline{\theta}, \bar{\theta}]$ and θ enhances hassling effectiveness, then the outer supremum will be attained at $\theta = \bar{\theta}$. However, the location of the maximizer (if any) is ambiguous if θ does not enhance hassling effectiveness.

A simple condition on these worst-case bounds guarantees the existence of an always peaceful mechanism.

Proposition 4. *Assume transgressions and hassling affect war payoffs. If $\bar{R}_C + \bar{R}_D \leq 1$, then there is a flexible-response crisis bargaining game form with voluntary agreements that has an always peaceful equilibrium.*

Proof. We prove the result by construction. Consider the following extensive form game:

1. C chooses $t \in \mathcal{T}$.

2. D chooses $h \in \mathcal{H}$ and selects whether to fight immediately. If D chooses to fight, then t is not realized, but h is.[28] War payoffs are thus $W_C(t, 0, \theta) - K_C(t)$ and $W_D(t, 0, \theta)$ respectively.

3. C chooses whether to fight. If C chooses to fight, war occurs with payoffs $W_C(t, h, \theta) - K_C(t)$ and $W_D(t, h, \theta) - K_D(h, \theta)$ respectively. If C chooses not to fight, then the pie is split, with C and D receiving $\bar{R}_C + K_C(t)$ and $1 - \bar{R}_C - K_C(t)$ respectively. Overall payoffs are thus $\bar{R}_C$ and $1 - \bar{R}_C - K_C(t) - K_D(h, \theta)$ respectively.

[28] This is stronger than our requirement that D be allowed to short-circuit transgressions, as here we do not restrict D to choose $h = 0$ if D preempts transgressions.

If the condition of the proposition holds, the following constitutes a perfect Bayesian equilibrium:

1. C chooses $t = 0$.

2. Following any choice of $t \in \mathcal{T}$, type θ of D chooses to fight if

$$\max_{h \in \mathcal{H}} \{W_D(t, h, \theta) - K_D(h, \theta)\} > 1 - \bar{R}_C - K_C(t). \tag{9}$$

   In this case, D selects a value of h that maximizes the above expression. Otherwise, D chooses not to fight and selects $h = 0$.

3. C updates beliefs in accordance with Bayes' rule wherever possible, maintaining them at the prior otherwise. Following all histories, C chooses not to fight.

To verify that this is an equilibrium, we will work backward. C's beliefs in the final stage are consistent by construction. Now take any t and h, and let $F(\cdot \,|\, t, h)$ denote the CDF corresponding to C's updated belief after a history where C chooses t and D chooses h and not to fight. C's expected utility from fighting is

$$\begin{aligned}
\int_\Theta W_C(t, h, \theta)\, dF(\theta \,|\, t, h) - K_C(t) &\leq \int_\Theta \max_{t \in \mathcal{T}} \{W_C(t, h, \theta) - K_C(t)\}\; dF(\theta \,|\, t, h) \\
&\leq \int_\Theta \max_{t \in \mathcal{T}} \{W_C(t, 0, \theta) - K_C(t)\}\; dF(\theta \,|\, t, h) \\
&\leq \sup_{\theta \in \Theta} \left\{ \max_{t \in \mathcal{T}} \{W_C(t, 0, \theta) - K_C(t)\} \right\} \\
&= \bar{R}_C.
\end{aligned}$$

Therefore, C's strategy of choosing not to fight is a best response.

Now consider D's choice of h and whether to fight. As an intermediate step, we verify that if D chooses not to fight, then $h = 0$ must be optimal. Given that C will also choose not to fight if D does so, D's expected utility from not fighting is $1 - \bar{R}_C - K_C(t) - K_D(h, \theta)$, which is strictly decreasing in h for all types θ. So now to characterize D's best response, we must simply compare D's expected utility from choosing not to fight with $h = 0$ to D's expected utility from fighting with an optimal choice of h (given t). The given strategy is thus a best response by construction.

Finally, consider C's initial choice of $t = 0$. As a first step to demonstrate that this is optimal, we will show that it results in every type of D choosing not to fight. Under the condition of the proposition, for all $\theta \in \Theta$ we have

$$\begin{aligned}
\max_{h \in \mathcal{H}} \{W_D(0, h, \theta) - K_D(h, \theta)\} &\leq \sup_{\theta \in \Theta} \max_{h \in \mathcal{H}} \{W_D(0, h, \theta) - K_D(h, \theta)\} \\
&= \bar{R}_D \\
&\leq 1 - \bar{R}_C \\
&= 1 - \bar{R}_C - K_C(0),
\end{aligned}$$

so Equation 9 fails. To complete the result, we show that it is optimal for C to choose t that results in peace for certain. Taking C and D's strategies in subsequent steps as given, any such choice of t (including $t = 0$, as just shown) results in an expected utility of $\bar{R}_C$ for C. Now consider a deviation $t > 0$ that results in war with some types of D. For each $\theta \in \Theta$, C's utility in case war with θ occurs is

$$\begin{aligned}
W_C(t, h^*(\theta), \theta) - K_C(t) &\leq W_C(t, 0, \theta) - K_C(t) \\
&\leq \max_{t' \in \mathcal{T}} \{W_C(t', 0, \theta) - K_C(t')\} \\
&\leq \sup_{\theta' \in \Theta} \left\{ \max_{t' \in \mathcal{T}} \{W_C(t', 0, \theta') - K_C(t')\} \right\} \\
&= \bar{R}_C,
\end{aligned}$$

where $h^*(\theta)$ solves the maximization problem in Equation 9. Therefore, such a deviation is unprofitable. $\square$

While conceptually straightforward, the condition $\bar{R}_C + \bar{R}_D \leq 1$ is difficult to check in practice. In the following subsections, we work through some parameterized examples to highlight the condition's practical use.

C.2.1 When θ Enhances Hassling Effectiveness

To further simplify the conditions—and to connect our work to mechanism design analyses of ordinary crisis bargaining games—we consider a parameterized model with *independent values*, in which C's war payoff is not a function of D's type. We give the problem additional structure by imposing linear and quadratic functional forms where convenient.

For the model primitives, we assume $\Theta = [\underline{\theta}, \bar{\theta}]$, where $0 < \underline{\theta} < \bar{\theta}$. The feasible spaces for transgressions and hassling are $\mathcal{T} = [0, \bar{t}]$ and $\mathcal{H} = [0, 1]$ respectively. We assume $\bar{t}$ is large enough that the boundary condition never binds for C.

We assume that C's baseline probability of victory is $p \in [0,1]$, which may shift in favor of either player depending on the level of transgressions and hassling. Letting $S(t,h)$ denote the net shift in favor of C, we assume

$$S(t,h) = t(1-h) - \sigma h,$$

where $\sigma > 0$ controls the direct marginal effect of hassling. War payoffs are therefore

$$W_C(t,h,\theta) = p - \kappa_C + S(t,h),$$
$$W_D(t,h,\theta) = 1 - p - \kappa_D(\theta) - S(t,h),$$

where κ_C is a positive constant and

$$\kappa_D(\theta) = \underline{\kappa}_D \left(\frac{\bar{\theta} - \theta}{\bar{\theta} - \underline{\theta}} \right) + \bar{\kappa}_D \left(\frac{\theta - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right).$$

In other words, the cost of war is $\underline{\kappa}_D$ for the lowest type and $\bar{\kappa}_D$ for the highest type, varying linearly in between. To reflect higher types having higher war payoffs, we assume $0 < \bar{\kappa}_D < \underline{\kappa}_D$.

To close the model, we must specify the cost functions for transgressions and hassling. For C's transgressions, we impose a quadratic form:

$$K_C(t) = \lambda_C t^2,$$

where $\lambda_C > 0$. For D, we impose a similar form,

$$K_D(h,\theta) = \lambda_D(\theta) h^2,$$

where $\lambda_D(\theta)$ is similar to the function used for the cost of war:

$$\lambda_D(\theta) = \underline{\lambda}_D \left(\frac{\bar{\theta} - \theta}{\bar{\theta} - \underline{\theta}} \right) + \bar{\lambda}_D \left(\frac{\theta - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right).$$

To reflect the assumption that θ enhances hassling capability, we assume $0 < \bar{\lambda}_D < \underline{\lambda}_D$. Furthermore, to ensure interior solutions, we assume $\bar{\lambda}_D > \frac{\sigma}{2}$.

To characterize D's reservation payoff, let $h^*(t,\theta)$ denote the best response of type θ to t if war were certain to occur. As D's utility is strictly concave in h, we can solve via the first-order condition,

$$\frac{\partial W_D(t,h,\theta)}{\partial h} - \frac{\partial K_D(h,\theta)}{\partial h} = t + \sigma - 2\lambda_D(\theta)h = 0.$$

Solving for h yields

$$h^*(t,\theta) = \min\left\{\frac{t+\sigma}{2\lambda_D(\theta)}, 1\right\}.$$

Note that $h^*(0,\theta) \in (0,1)$ for all θ, per our assumption that $\bar{\lambda}_D > \frac{\sigma}{2}$. Therefore, we have

$$\begin{aligned}
\bar{R}_D &= \sup_{\theta\in\Theta} \{W_D(0, h^*(0,\theta), \theta) - K_D(h^*(0,\theta),\theta)\} \\
&= W_D(0, h^*(0,\bar{\theta}), \bar{\theta}) - K_D(h^*(0,\bar{\theta}),\bar{\theta}) \\
&= 1 - p - \bar{\kappa}_D + \frac{\sigma^2}{2\bar{\lambda}_D} - \bar{\lambda}_D \frac{\sigma^2}{4\bar{\lambda}_D^2} \\
&= 1 - p - \bar{\kappa}_D + \frac{\sigma^2}{4\bar{\lambda}_D}.
\end{aligned}$$

To characterize C's reservation payoff, note that W_C is everywhere constant in θ, and so we have

$$\bar{R}_C = \max_{t\in\mathcal{T}} \{p - \kappa_C + S(t,h) - K_C(t)\}.$$

As the objective is strictly concave, we may solve for the best response $t^*(h)$ via the first-order condition,

$$\frac{\partial S(t,h)}{\partial t} - \frac{dK_C(t)}{dt} = 1 - h - 2\lambda_C t = 0.$$

Consequently, we have

$$t^*(h) = \frac{1-h}{2\lambda_C},$$

so $t^*(h) > 0$ for all $h < 1$. C's reservation value is therefore

$$\begin{aligned}
\bar{R}_C &= p - \kappa_C + S(t^*(0), 0) - K_C(t^*(0)) \\
&= p - \kappa_C + \frac{1}{2\lambda_C} - \lambda_C \frac{1}{4\lambda_C^2} \\
&= p - \kappa_C + \frac{1}{4\lambda_C}.
\end{aligned}$$

Recall from Proposition 4 that the sufficient condition for an always peaceful equilibrium is

$\bar{R}_C + \bar{R}_D \leq 1$. Here that condition is equivalent to

$$\kappa_C + \bar{\kappa}_D \geq \frac{1}{4}\left(\frac{\sigma^2}{\bar{\lambda}_D} + \frac{1}{\lambda_C}\right).$$

The greater the cost of fighting for C or for the highest type of D, the easier it is for this condition to hold. Similarly, the condition is also more likely to hold as the marginal cost of transgressions and the marginal cost of hassling for the highest type of D increase. On the other hand, the greater the direct effect of hassling on the war outcome, the less likely the condition is to hold. Most importantly, unlike in ordinary crisis bargaining models, positive costs of fighting alone are not sufficient to ensure a peaceful mechanism exists, even in the independent values case.

C.2.2 When θ Degrades Hassling Effectiveness

We now consider the more difficult case, where types of D with lower war costs have higher hassling costs. We use the same parameterized model as in the previous section, with a single change: we now assume $\bar{\lambda}_D > \underline{\lambda}_D$, so that the marginal hassling cost now increases with θ. Additionally, we must now modify our assumption to ensure interior solutions to $\underline{\lambda} > \frac{\sigma}{2}$.

C's reservation payoff is the same as in the previous case. To characterize D's reservation payoff, note that for any θ we have

$$\begin{aligned} W_D(0, h^*(0,\theta), \theta) - K_D(h^*(0,\theta), \theta) &= 1 - p - \kappa_D(\theta) + \frac{\sigma^2}{2\lambda_D(\theta)} - \lambda_D(\theta)\frac{\sigma^2}{4\lambda_D(\theta)^2} \\ &= 1 - p - \kappa_D(\theta) + \frac{\sigma^2}{4\lambda_D(\theta)}. \end{aligned}$$

We want to find the θ that maximizes this expression. Taking its derivative with respect to θ, we have

$$\begin{aligned} \frac{d}{d\theta}\left[W_D(0, h^*(0,\theta), \theta) - K_D(h^*(0,\theta), \theta)\right] &= -\kappa_D'(\theta) - \frac{\sigma^2}{4\lambda_D(\theta)^2}\lambda_D'(\theta) \\ &= \frac{\underline{\kappa}_D - \bar{\kappa}_D}{\bar{\theta} - \underline{\theta}} - \frac{\sigma^2}{4\lambda_D(\theta)^2}\left(\frac{\bar{\lambda}_D - \underline{\lambda}_D}{\bar{\theta} - \underline{\theta}}\right). \end{aligned}$$

Because $\bar{\lambda}_D > \underline{\lambda}_D$ and $\lambda_D(\theta)$ is positive and strictly increasing in θ, this expression is strictly increasing in θ. Therefore, $W_D(0, h^*(0,\theta), \theta) - K_D(h^*(0,\theta), \theta)$ is strictly convex in θ, meaning it attains its maximum at a boundary point. We therefore have

$$\begin{aligned}\bar{R}_D &= \sup_{\theta\in\Theta}\{W_D(0,h^*(0,\theta),\theta) - K_D(h^*(0,\theta),\theta)\}\\ &= 1 - p + \max\left\{\frac{\sigma^2}{4\bar{\lambda}_D} - \bar{\kappa}_D, \frac{\sigma^2}{4\underline{\lambda}_D} - \underline{\kappa}_D\right\}.\end{aligned}$$

The condition for the maximum to be attained at the values for $\bar{\theta}$ is

$$\underline{\kappa}_D - \bar{\kappa}_D \geq \frac{\sigma^2}{4\underline{\lambda}_D\bar{\lambda}_D}(\bar{\lambda}_D - \underline{\lambda}_D),$$

i.e., the difference that the type makes in the cost of war is large enough relative to the difference it makes in the marginal cost of hassling.

We can now restate the sufficient condition of Proposition 4. We have $\bar{R}_C + \bar{R}_D \leq 1$ if and only if

$$\begin{cases}\kappa_C + \bar{\kappa}_D \geq \dfrac{1}{4}\left(\dfrac{\sigma^2}{\bar{\lambda}_D} + \dfrac{1}{\lambda_C}\right) & \text{if } \underline{\kappa}_D - \bar{\kappa}_D \geq \dfrac{\sigma^2}{4\underline{\lambda}_D\bar{\lambda}_D}(\bar{\lambda}_D - \underline{\lambda}_D),\\ \kappa_C + \underline{\kappa}_D \geq \dfrac{1}{4}\left(\dfrac{\sigma^2}{\underline{\lambda}_D} + \dfrac{1}{\lambda_C}\right) & \text{otherwise.}\end{cases}$$