Uncertainty in Crisis Bargaining with Multiple Policy Options*

Brenton Kenkel† Peter Schram‡

October 29, 2023

Word count: 9,647

Abstract

Formal models commonly characterize interstate bargaining as dichotomous, ending in either war or peace. But there are many forms of coercion—including supporting rebel groups, sanctions, and cyberattacks. How does the availability of intermediate policy options affect the incidence of war and peace? We present an analysis of crisis bargaining models with intermediate policy options that challenges conventional results about the relationship between private information and negotiation outcomes. In our "flexible-response" modeling framework, unlike in traditional crisis bargaining models, we find that greater private war payoffs may be associated with a lower probability of war or worse settlement values. When intermediate options are available, the relationship between the private efficacy of war and the private efficacy of these other options largely determines equilibrium outcomes. By utilizing the tools of mechanism design, we derive game-form free results on how private information shapes international conflict, regardless of the precise negotiating protocol.

*We thank Sandeep Baliga, Brett Benson, Peter Bils, Ethan Bueno de Mesquita, Josh Clinton, Andrew Coe, David Carter, Songying Fang, Mark Fey, Andrew Kydd, Chris Li, Adam Meirowitz, Jim Morrow, Kris Ramsay, Keith Schnakenberg, Brad Smith, William Spaniel, and Jessica Sun for helpful comments and discussions. We also thank participants at the Vanderbilt Conflict Workshop, the Virtual Formal Theory Seminar, and the 2022 European Political Science Association conference for their feedback on presentations of earlier drafts.

†Associate 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 central question in the study of international conflict is why wars occur even though they are costly to all sides. It is puzzling that states seek political gains through violence when the underlying stakes could be allocated costlessly at the bargaining table (Fearon 1995). Perhaps the most prominent explanation for this puzzle is that states have incomplete information about each other's capabilities or intentions (Jervis 1976), leading them to make incompatible demands about the shape of a negotiated settlement. To understand why inefficient conflict takes place, we must understand how private information shapes states' incentives and actions.

A long tradition of scholarship has used game-theoretical models of international crisis bargaining to understand the role of incomplete information in the outbreak of conflict. Building from foundational theories of bargaining under uncertainty (Morrow 1989; Fearon 1995), an impressive line of recent work has used crisis bargaining models to study the role of incentives, information, and capabilities in determining when and why states go to war (e.g., Leventoğlu and Tarar 2008; Bas and Schub 2017; Spaniel and Malone 2019; Morrow and Sun 2020). These crisis bargaining models have made valuable progress in the theoretical study of interstate war, but they are limited by a common simplifying assumption about the nature of state interactions—namely, that states either reach a peaceful and efficient bargain or else engage in a decisive full-scale war. In reality, states in crises possess a diverse range of coercive intermediate policy options besides war and peace: implementing sanctions or tariffs (Bapat and Kwon 2015; McCormack and Pascoe 2017; Di Lonardo and Tyson 2022*a*), offering third-party support to an adversary's enemies (Bapat 2006; Carter 2012; Berman and Lake 2019; Qiu 2022*b*), engaging in cyberattacks (Gartzke and Lindsay 2015; Baliga, Bueno de Mesquita and Wolitzky 2020), conducting drone strikes (Lin-Greenberg 2022*b*), or engaging in low-level conflict, hybrid warfare, or "hassling" (Lanoszka 2016; Gurantz and Hirsch 2017; Schram 2021; Lin-Greenberg 2022*a*). While simplifying assumptions are useful in allowing models to abstract away from irrelevant features and focus in on key mechanisms (Paine et al. 2020), the war-or-peace dichotomy in crisis bargaining models is overly reductive and far more consequential. As we demonstrate below, this assumption systematically limits our predictions about the relationship between private information and the outcomes of international disputes.

Crisis bargaining models that end either in peaceful settlement or a decisive war offer a strikingly consistent portrayal of the conditions that lead to conflict. Incomplete or asymmetric information is modeled as a private signal that a state receives about its ability, willingness, or political motivations to wage war, often called the state's "type." Within equilibria of crisis bargaining models with private information, states with stronger private types are more likely to take actions that result in war, and they receive greater payoffs from bargaining even when the game ends peacefully. These relationships between private type and crisis outcomes are not specific to any particular model, but rather occur within *every* equilibrium of *any* crisis bargaining game where the only possible outcomes are war or peace (Banks 1990; Fey and Ramsay 2011).[1] Consequently, traditional crisis bargaining models provide a consistent answer to the question of when states are most likely to resort to violence: when at least one state has a strong private signal of its battlefield ability or willingness to fight.[2]

We reexamine the relationship between private information and war in *flexible-response crisis bargaining games* that allow for forms of costly conflict short of all-out war. Once we allow for multiple forms of conflict, the traditional models' monotonic relationships between private type and bargaining outcomes no longer hold unconditionally. Within the flexible-response models, states with better private signals about their ability to wage war may be

[1] These monotonicity results may not hold for models with alternative forms of uncertainty (e.g., Slantchev 2011; Spaniel and Malone 2019).

[2] As one interpretation of this result, states with stronger private types will behave more aggressively in bargaining.

less likely to fight and may have lower payoffs in equilibrium. These exceptions to the traditional patterns arise in the presence of multiple policy options because the factors that influence a state's ability or willingness to wage war may also shape its facility with alternative coercive instruments. Importantly, we do not simply present a single counterexample to established theories. Instead, we conduct a game-free analysis along the lines of previous mechanism design research (Banks 1990; Fey and Ramsay 2011; Akçay et al. 2012; Liu 2021), showing that the exceptions to standard patterns can arise in *any* interaction with flexible policy options, regardless of the specifics of the bargaining protocol. Our approach combines the methodological generality of previous game-free analyses with the substantive richness of newer models of coercive policies short of war, which allows us to characterize broad influences on the likelihood of conflict in more realistic strategic environments.

The key driver of our results is that the same features that affect a state's war effectiveness might also affect its effectiveness in using alternative policy choices, for better or worse.[3] The relationship between war payoffs and the payoffs from alternate policy choices is important because it means the private signal that a state receives about its wartime strength also shapes its ability to use other coercive policies. For example, if a state has a wide range of privately known cyber-exploits, then the state knows that it could not only perform well in a conventional war that uses cyberattacks but also perform well in a precise cyberattack against a target's infrastructure. Because this state possesses robust private capabilities, it is effective in both war and low-level conflict. Alternatively, if a state's leaders are privately concerned about losing popular support, then they may be more willing to fight a war to create a rally-around-the-flag effect than they would be to choose intermediate options such as tariffs or sanctions that might also damage the domestic economy (Baker and Oneal 2001). In this case, the same private signal that would make a leader more willing to go war—namely, perceived electoral vulnerability—would also increase their perceived cost of

[3]To clarify our terminology, a state with high war effectiveness is a state that has a greater capacity to go to war, a greater willingness to go to war, or both. Formally, states that are more effective at war attain higher payoffs from war. We define hassling effectiveness analogously.

economic coercion. The question of whether private war payoffs are associated with greater or lower payoffs from alternative policy options is an empirical one, and its answer varies across cases and contexts—but these linkages undoubtedly exist, and their effects on the outcomes of crisis bargaining have not been systematically examined. We demonstrate that the basic relationship between a state's signal of wartime effectiveness and the effectiveness of alternate coercive policies (i.e., whether types with higher war payoffs are advantaged or disadvantaged in the use of other instruments) is critical in predicting how private information affects the likelihood of conflict.

We provide a general characterization of the relationship between private war payoffs, the likelihood of conflict, and equilibrium payoffs in all flexible-response crisis bargaining games. The relationship between private signals about war payoffs and the effectiveness of low-level conflict is critical and refines existing understandings in several ways. First, the likelihood of war increases with the state's private war payoff (or type) when states with higher private war payoffs also have less capacity for or less willingness to use low-level responses, or when low-level responses only become slightly more effective as war payoffs increase—for example, as in the rally-round-the-flag scenario above. Second, a state's equilibrium payoff is only guaranteed to increase with its privately known strength if greater war payoffs are associated with greater effectiveness in executing flexible responses—for example, as in the cyber-conflict scenario. Third, unlike in traditional models, states with stronger private signals of wartime prowess do not necessarily receive better settlements when the interaction ends short of war. The value of settlement instead depends on the extent to which the state employs flexible coercive instruments. Finally, the extent to which a state uses intermediate policy instruments may depend greatly on idiosyncratic features of the negotiation process, even with no change in the underlying technological fundamentals.

Building on our formal analysis, we also offer new substantive insights into the value of specific military technological advancements. Existing research examines the value of certain coercive capabilities within specific military contexts, such as the value of airstrike capabilities in low-level conflicts and war (Horowitz and Reiter 2001; Kreps and Fuhrmann 2011; Allen and Martinez Machain 2019). Most, though not all,[4] of the previous research in this area only considers how these technologies fare within an active conflict or after a challenger has already transgressed. However, it is also valuable to know how these capabilities shape outcomes at the bargaining table. That is where our analysis comes in. Once we identify whether improved private wartime capabilities improve hassling capabilities or are detrimental to them, then we are able to make clear predictions about how changes in wartime capabilities affect the likelihood of war, as well as distributive outcomes at the bargaining table. Proceeding from our results, scholars can leverage existing empirical and public policy research to identify how improving specific capabilities or altering the factors that increase resolve can affect the outcomes in any crisis setting where states have options short of all-out war.

Our work builds on a recent line of research on crisis bargaining and deterrence in which states are assumed to have multiple coercive options available to respond to a threat (Schultz 2010; McCormack and Pascoe 2017; Coe 2018; Spaniel and Malone 2019; Qiu 2022*b*; Baliga, Bueno de Mesquita and Wolitzky 2020; Schram 2021; Di Lonardo and Tyson 2022*a*). This paper is the first to systematically examine how the spillover effects of improvements in one kind of conflict capability can also affect other response options, allowing us to offer novel insights into a wide range of previously neglected substantive settings. The most closely related work to ours is by Schram (2022), who considers a deterrence game with multiple conflict options and a publicly observed type that determines payoffs from conflict. Our key innovation is to assume that a state's private type affects both its war payoff and its low-level conflict costs, allowing us to show that prior monotonicity results (Banks 1990; Fey and Ramsay 2011) only hold under particular conditions.

[4]See Post (2019) as an exception.

1 Flexible Responses in International Crises

Our goal is to understand crisis bargaining in cases where states have intermediate coercive options short of all-out war. We begin by illustrating the concept with an example. In 2006, Israel discovered that Syria was building a nuclear reactor. Internally, Israeli decision-makers viewed the possibility of a nuclear-armed Syria as an "existential threat" to the Israeli state (Opall-Rome 2018). Additionally, Israel possessed a covert capability at its disposal: an electronic warfare attack that could (at least temporarily) disable Syria's integrated air defense system, impairing Syria's ability to track incoming aircraft and act (Katz 2010). Traditional crisis bargaining models—where possessing a high private willingness or a robust private capability to fight a war always leads to a greater likelihood of war (Banks 1990; Fey and Ramsay 2011)—would portray this as the exact setting in which we might expect Israel to resort to war. But instead, Israel used the electronic warfare attack as part of a limited airstrike on the reactor (Katz 2010). Israel's airstrike, known as Operation Outside the Box, successfully destroyed a critical component of the Syrian nuclear program while negating the need for a more expansive response.

Flexible-response crisis bargaining models capture political interactions like those surrounding the Syrian reactor. In these models, a challenger state may undertake some opportunistic and costly action, which we call a "transgression," against a defender. Transgressions are beneficial to the challenger but harm the defender's interests. In our example, Syria's construction of the reactor was a transgression that could have eventually led to Syria possessing a nuclear bomb, thus strengthening its future leverage against Israel. Transgressions like this arise in theories of enforcement problems in bargaining (Schultz 2010), deterrence (Fearon 1997; Gurantz and Hirsch 2017), and endogenous power shifts (Debs and Monteiro 2014).[5] Examples of transgressions include states investing in new military technologies (Debs and Monteiro 2014; Gartzke and Lindsay 2017; Spaniel 2019), forming alliances (Benson and

[5] The "transgression" here is similar to the challenger's first move in a deterrence game (e.g., Chassang and Miquel 2010; Di Lonardo and Tyson 2022*b*; Kydd and McManus 2017; Baliga and Sjöström 2020).

Smith 2022), or securing valuable territory (Powell 2006).

In response, the defender's options are to settle the issue peacefully through negotiations, go to war to resolve the issue decisively, or engage in some low-level action to undercut the transgression. The first two options—settlement or war—are the two standard outcomes in models of crisis bargaining and deterrence. We will refer to the the defender's low-level response option as "hassling." As originally defined in Schram (2021), 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 expands it to include any actions by the defender that are costly, coercive, and fall outside of war, like sanctions, low-level conflict, limited airstrikes, or cyberattacks. In the Syrian case, Israel detected Syria's nuclear reactor and destroyed it. This was a form of hassling because of its destructive impact on Syria's nuclear program, but it was not a decisive military move that would prevent the Assad regime from ever possessing a nuclear weapon. By contrast, the 2003 U.S. invasion of Iraq decisively prevented the Ba'athist regime from ever attaining nuclear weapons by overthrowing it, consistent with our treatment of war. Hassling can take the form of limited airstrikes (Reiter 2005; Fuhrmann and Kreps 2010), hybrid conflict (Lanoszka 2016), aspects of gray-zone conflict (Mazarr 2015; Gannon et al. 2020), (limited) preventive war (Levy 2011), *fait accompli* (Tarar 2016), sanctions (McCormack and Pascoe 2017), or arming (Coe 2018).[6]

We assume the defender possesses a private type that influences war and hassling payoffs. When a state knows that they are very capable at conducting war or very willing to conduct a war, that state may be more (or less) capable or willing to conduct hassling. As one example, suppose a state possessed a covert capability where it could use an electronic warfare attack or a cyberattack to disable a rival's air defense systems. Based on these private capabilities, this state might be especially effective at conducting a war—but also capable of implementing

[6] In some ways, hassling resembles a restrained, "instrumental" style of coercive action conducted for the pursuit of limited goals and with the intent to not violate important escalation thresholds; in many ways, this echoes discussions in Kahn (2017) about how conflict will evolve in the Twenty-first Century.

an effective low-level hassling attack to weaken a target. As a second example, suppose a different state is privately very concerned about the domestic political costs from fighting a war. Based on these private preferences, this state may be more willing to use low-salience hassling techniques—like drone strikes—within a crisis. Ultimately, the question of whether better private capabilities or the factors that result in a greater willingness to go to war also make low-level options more or less effective is an empirical one, as we discuss below following the main results.

2 The Impact of Flexible Responses on Standard Crisis Bargaining Results

Before we present our general results, we present two stylized examples of how introducing intermediate conflict options can undermine the canonical relationships between private information and outcomes in crisis bargaining games. Both games are between a challenger (C) and defender (D). Nature moves first and decides whether D's war payoff is low ($\theta = \underline{\theta}$) or high ($\theta = \bar{\theta}$), with each outcome having positive probability.

First, D observes their private type, while C only knows the prior probability that D is a high or low type. This type represents D's privately known abilities, willingness, or political motivations for waging a war against C. Next, C selects whether to transgress ($t = 1$) or not ($t = 0$). Finally, D can accept the transgression, go to war over it, or conduct some limited response via hassling (where hassling could refer to sanctions, cyberattacks, gray-zone conflict, etc.).

Figure 1 illustrates a setting where only the weak type of D goes to war, contrary to the standard result that the probability of conflict increases with D's private value of war (Banks 1990; Fey and Ramsay 2011). This happens because the type with a greater war payoff is also more effective at hassling, as in the cyber-capability example above. Specifically, while

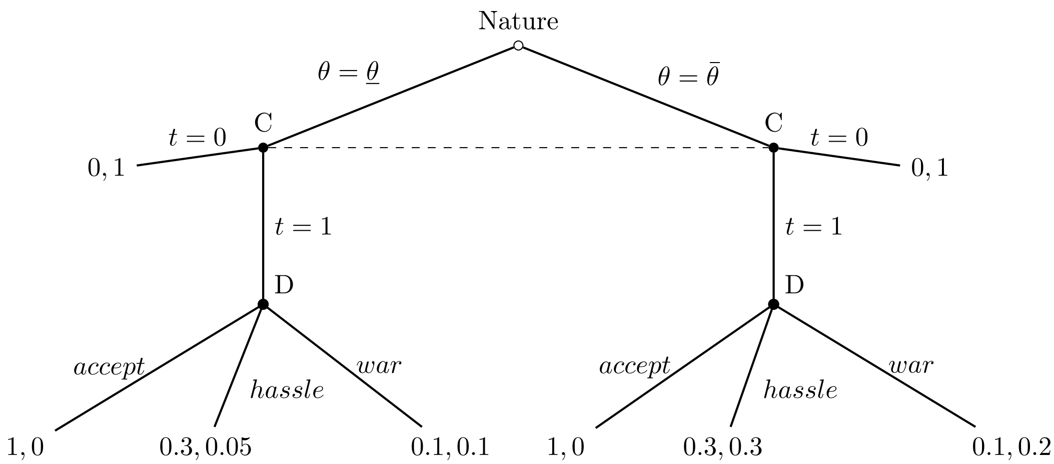

Figure 1: Greater private type θ implies less war.

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

war outcomes improve with θ, hassling outcomes improve even more. In equilibrium, the stronger type (or the type with greater war efficacy) opts not to fight a war because hassling is better.[7] Note that if the hassling option were not available, then the equilibrium would conform to the usual pattern, with both types going to war.

Next, we consider an interaction in which the stronger type of D receives a lower payoff, illustrated in Figure 2. This stands in contrast to the standard result that actors with higher private war payoffs end up better off (Banks 1990; Fey and Ramsay 2011). In this example, greater private payoffs from fighting a war corresponds to a lower hassling payoff, as in the rally-round-the-flag example above. In equilibrium, weaker types ($\underline{\theta}$) hassle and stronger types ($\overline{\theta}$) go to war. Crucially, the increase in the stronger type's war payoff does not sufficiently compensate for the decrease in its hassling payoff. Once more, if the hassling option were unavailable, we would return to the standard pattern: both types would go to war in equilibrium, with the stronger one yielding a greater utility.

[7] Unless otherwise specified, we will define "stronger" types as the types that possess a greater war payoff, which is how Fey and Ramsay (2011) use the term. That said, a key take-away from this paper is that this notion of strength is less informative in environments with multiple policy options. We discuss this further in the Conclusion.

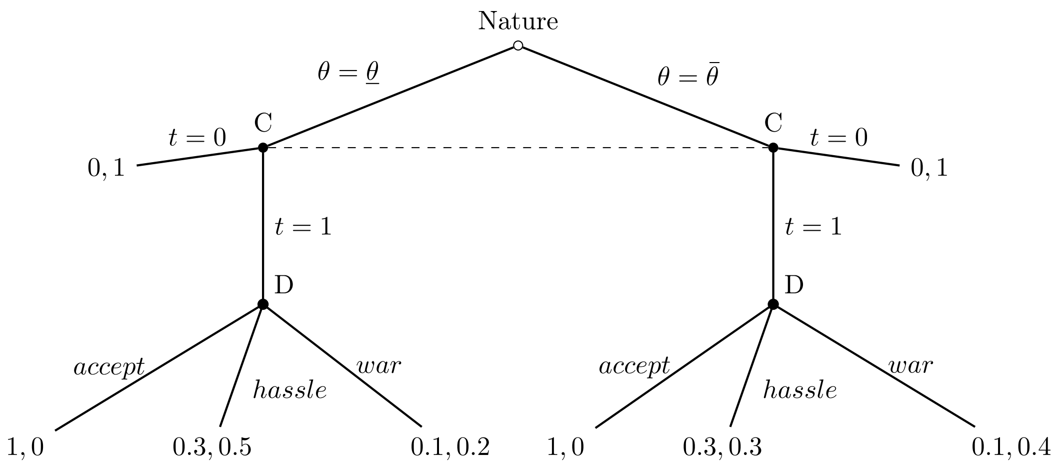

Figure 2: Greater private type θ implies lower utility

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.

The two models we have presented here are deliberately sparse, so as to illustrate the simplest possible settings in which games with flexible responses depart from ordinary crisis bargaining games. Because these models lack many important elements of crisis bargaining theory, such as pre-negotiation signaling and incentives to misrepresent, one might wonder whether the lessons we have drawn from them would carry over to richer, more realistic settings.[8] To identify broad regularities in flexible response models, we employ the game-free methodology of Banks (1990) and Fey and Ramsay (2011), identifying properties that arise from foundational requirements of equilibrium rather than idiosyncratic features of any given game tree. We find, in fact, that the outcomes in Figures 1 and 2 are typical of crisis bargaining with flexible responses.

3 Formal Framework

We consider a class of models where crisis negotiations may end in pure peace, total war, or some costly non-war outcome.

[8]In Appendix C.6, we present a signaling model where the probability of war decreases with D's type.

3.1 Structure of the Interaction

The players are a *C*hallenger and *D*efender in dispute over a prize of size 1. At the outset of the game, Nature assigns D's private type, $\theta \in \mathbb{R}$. D's war payoff increases with θ, while D's cost of hassling may increase, decrease, or neither. The realized value of θ is known only to D, but its prior distribution is common knowledge. Let F denote the cumulative distribution function of this prior distribution, and let Θ denote its support.

The interaction takes the familiar form of a crisis negotiation, except each state may engage in activity that affects outcomes short of war. As in typical crisis games, the states have bargaining strategies, denoted $b_C \in \mathcal{B}_C$ and $b_D \in \mathcal{B}_D$, which may represent offers, counteroffers, accept-reject plans, costly or cheap talk signals, and so on. Each state also has a special action that shifts the outcome in case of peace, but at a cost. For C we call this action a "transgression," denoted $t \in \mathcal{T} \subseteq \mathbb{R}_+$. Similarly, for D the action is "hassling," denoted $h \in \mathcal{H} \subseteq \mathbb{R}_+$. We place no restriction on whether transgressions and hassling are chosen before, during, or after the bargaining process—we simply assume that they do not directly affect whether war occurs, and that their cost is independent of the other components of the bargaining strategy.

A game form G consists of the bargaining action spaces, $\mathcal{B}_A$ and $\mathcal{B}_D$, along with an outcome function g that maps the choices (t, h, b_C, b_D) into the set of possible crisis bargaining outcomes.[9] We decompose the outcome function g into three components: whether war occurs, what C receives from bargaining, and what D receives from bargaining. Whether war occurs 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.[10] Conditional on war not occurring, each player's payoff depends on the bargaining behavior, C's choice of transgression, and D's selection of

[9] The game form represents the elements of the model that are specific to a particular bargaining protocol. Thus, we treat the type space, prior distribution, transgression and hassling action sets, cost functions, and war payoff functions as primitives of the interaction rather than features of a specific game form.

[10] By ruling out $\pi^g \in (0, 1)$, we are implicitly assuming that the bargaining process has no exogenous random components (see Fey and Kenkel 2021).

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, if war is avoided.

If war occurs, war payoffs depend on D's private information, but not on the endogenous bargaining choices, including transgressions and hassling.[11] We therefore write war payoffs as $W_C(\theta)$ and $W_D(\theta)$. We assume W_D is strictly increasing, so higher types of D can be interpreted as more effective in war, or as "stronger" types.

If war is avoided, each player receives their division of the spoils but must pay the cost of their transgression or hassling. Let $K_C(t)$ denote the cost to C, and let $K_D(h,\theta)$ denote the cost to D. We assume that K_C is strictly increasing in t, and we assume that K_D is strictly increasing in h. For now, we are agnostic whether K_D is increasing or decreasing in θ. We let $h = 0$ denote no hassling, which entails assuming that $0 \in \mathcal{H}$ and $K_D(0,\theta) = 0$ for all θ. Putting these together, the players' utility functions in a given game form are:

$$u_C^g(t,h,b_C,b_D \,|\, \theta) = (1-\pi^g(b_C,b_D))W_C(\theta) + \pi^g(b_C,b_D)[V_C^g(t,h,b_C,b_D) - K_C(t)],$$
$$u_D^g(t,h,b_C,b_D \,|\, \theta) = (1-\pi^g(b_C,b_D))W_D(\theta) + \pi^g(b_C,b_D)[V_D^g(t,h,b_C,b_D) - K_D(h,\theta)].$$

We restrict our attention to games 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. A sufficient condition is that each player has an action $b_i' \in \mathcal{B}_i$ to guarantee war: $\pi^g(b_i',b_j) = 0$ for all $b_j \in \mathcal{B}_j$. As we demonstrate below, this condition places important limits on what kinds of outcomes are sustainable as equilibria.

The relationship between D's private type and hassling ability is central to our analysis. We will show that the effects of θ on equilibrium outcomes depend critically on whether

[11] This assumption may appear to exclude games where power shifts over time as a function of transgressions and hassling. However, in Appendix C.7, we demonstrate how our results apply to a game with endogenously shifting power. The important condition for application is that war can only occur in the first period on the equilibrium path, as is typically the case in shifting-power models (e.g., Powell 2006; Spaniel 2019).

greater payoffs from war are associated with higher or lower costs of hassling. In general, when comparing types θ' and θ'', we say that θ'' has greater hassling effectiveness than θ' if $K_D(h, \theta'') < K_D(h, \theta')$ for all $h > 0$. We say that θ improves hassling effectiveness if higher types always have greater hassling effectiveness than lower types. In the opposite case, when K_D strictly increases with D's type, we say θ degrades hassling effectiveness. Note that this means "stronger" types—those that attain greater war payoffs—could attain lower hassling payoffs.

3.2 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 discuss the details of bargaining itself, as our primary concern is the outcome of the interaction: whether war occurs, and if not, what each party receives from a bargained outcome.

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

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

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

While mimicking another type's strategy may change the hassling level, the occurrence of conflict, and the settlement value, it does not change D's war payoff, nor the cost D pays for any given hassling level. The key requirement of Bayesian equilibrium is that no type can increase its payoff by mimicking 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

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

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 implies that if we find that some property holds for all incentive-compatible direct mechanisms, then it is true of all equilibria of all flexible-response crisis bargaining games. Without fully characterizing the particulars of how crisis bargaining plays out in any particular game, we are still able to characterize robust properties of the equilibrium 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 a war. In the language of mechanism design, this requirement amounts to a participation constraint, or what Fey and Ramsay (2011) call voluntary agreements in the crisis bargaining context.

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 go to war in equilibrium. 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 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).

4 Private Type and the Probability of War

In crisis bargaining games without flexible responses, private signals of wartime effectiveness or strength are associated with a greater equilibrium probability of war (e.g., Morrow 1989; Fearon 1995; Tarar 2021). A simple intuition drives this result. If some type finds it worthwhile to risk war to receive a better deal at the bargaining table, then all stronger types must be willing to run at least as great a risk—after all, by mimicking weaker types, strong types could attain the same bargained settlement while doing better if bargaining fails and war occurs. In other words, a country will behave more aggressively (or at least no less so) at the bargaining table when its private signal about its war payoff is relatively strong.

For equilibria with no hassling, our analysis recovers the classic positive relationship between private type and the likelihood of conflict. As long as $h(\theta) = 0$ for all types, weaker types never have a greater chance of conflict than higher types. The following result recovers

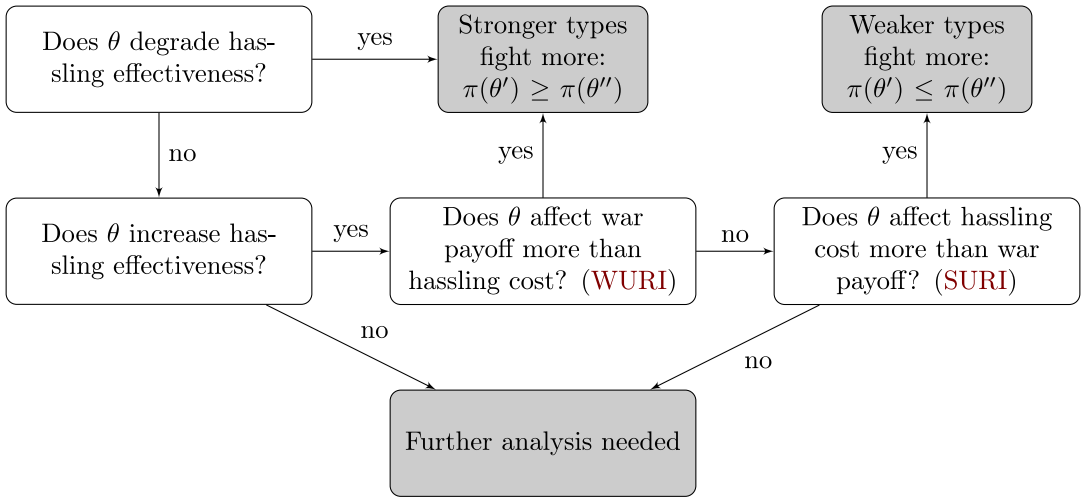

Figure 3: The relationship between type ($\theta' < \theta''$) and war likelihood in all flexible-response crisis bargaining games.

Lemma 1 of Banks (1990) as a special case in our environment.[13]

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

This result confirms that the exceptions we find to the classic monotonicity results are due exclusively to our introduction of flexible responses, not to any other feature of our formal framework. Once states begin to employ alternative instruments for altering the balance of bargaining power, this straightforward relationship between private strength and the risk of war holds only under special conditions.

The flow chart in Figure 3 summarizes our findings on the relationship between private information and the occurrence of conflict. If private strength degrades hassling effectiveness, then stronger types are more likely to fight a war, just as in models without flexible responses. The same is true if private strength increases hassling effectiveness and has a stronger effect on war payoffs than on the cost of hassling (the WURI condition, formally defined below). However, we find that weaker types are more likely to fight a war—the opposite of the traditional result—when private strength improves hassling effectiveness but

[13] All proofs are in the Appendix.

affects war payoffs less than hassling costs (SURI, also defined below).

When hassling takes place, the effect of private type on the chance of war depends on whether a state's private war capability improves or degrades its hassling effectiveness. If private type degrades hassling effectiveness—i.e., if types with greater war payoffs also have greater costs of hassling—then higher types are weakly more likely to go to conflict.

Proposition 1. *Assume θ degrades hassling effectiveness. If $\theta' < \theta''$, then $\pi(\theta') \geq \pi(\theta'')$.*

Because the value of an efficient settlement without hassling is the same regardless of θ, higher types have the most incentive to choose war over an efficient settlement. If private type degrades hassling effectiveness, then this logic carries over to settlements involving hassling as well. If some type of D prefers war over a settlement with hassling level $h \geq 0$, then all stronger types must have the same preference: they have an even higher war payoff and would pay a greater cost from the same level of hassling.

When private type is instead associated with lower costs of hassling, we need more conditions to characterize its effect on the likelihood of war. In this case, not only is war more attractive to stronger types of D, but so is any given settlement with hassling. Because war payoffs and settlement payoffs are now moving in the same direction as D's type increases, the critical question for our purposes is which rate of increase is quicker. We say that the war utility is relatively increasing (WURI) when θ has a greater marginal effect on war payoffs than on the cost of hassling. In the opposite case, we say the settlement utility is relatively increasing (SURI).

Definition 3. In a direct mechanism, the *war utility is relatively increasing* if

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

The *settlement utility is relatively increasing* if

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

If either of these holds and private strength increases hassling effectiveness, then we can characterize the direction of the type's effect on the equilibrium chance of conflict.

Proposition 2. *Assume θ improves hassling effectiveness, and let $\theta' < \theta''$. If (WURI) holds, then $\pi(\theta') \geq \pi(\theta'')$. If (SURI) holds, then $\pi(\theta') \leq \pi(\theta'')$.*

Consequently, 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 war 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 a war on the path of play.

While Proposition 2 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 refer to the levels of hassling chosen on the path of play. In Appendix C.1, we derive sufficient conditions on model primitives to ensure WURI holds: the hassling cost function must have decreasing differences (see Ashworth and Bueno de Mesquita 2006), and the marginal effect of θ on the war payoff must be greater than its effect on hassling cost at

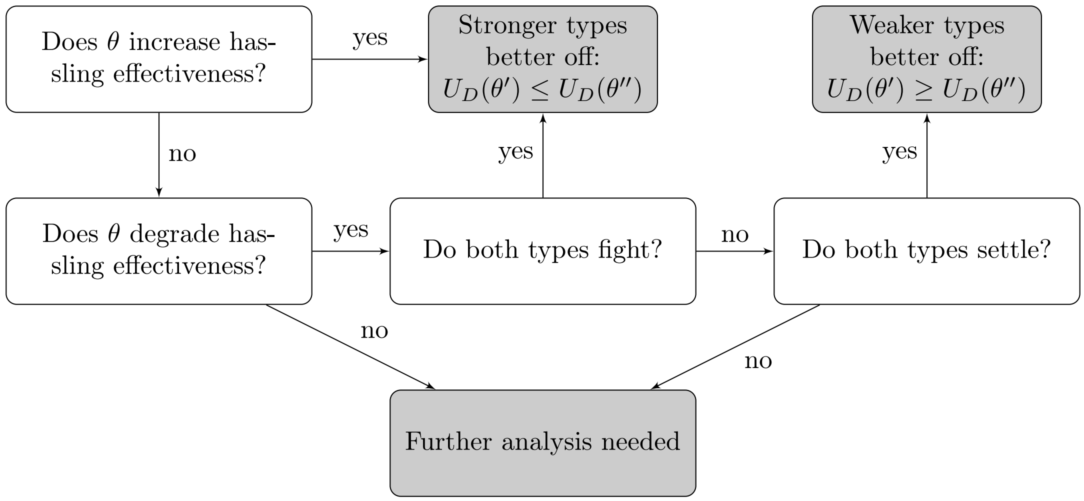

Figure 4: The relationship between type ($\theta' < \theta''$) and equilibrium payoffs in all flexible-response crisis bargaining games.

the highest feasible level of hassling.

5 Private Type and Payoffs

A second key regularity in crisis bargaining games without flexible responses is that a better private war capability corresponds to a better expected equilibrium payoff, even when war does not occur (Banks 1990; Fey and Ramsay 2011). In this class of models, greater private strength increases an actor's willingness to run the risk of war, allowing the actor to extract more from negotiations if the interaction does not end in war and granting the actor a greater war payoff should the interaction end in war.

We recover the same relationship between private strength and expected equilibrium payoffs in the flexible response environment when hassling does not take place along the path of play. The following result is our analogue of Lemma 4 from Banks (1990).

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

Once we introduce additional ways for states to operate outside of war and peaceful negotiations, the tight relationship between private war capability and equilibrium payoffs no longer holds. The flowchart in Figure 4 summarizes the relationship between private strength and utility in flexible-response crisis bargaining games. Obviously, among types that end up going to war in equilibrium, stronger types are always better off. Outside of that case, however, private strength is only guaranteed to increase payoffs when it is also associated with greater hassling capability. If private strength instead degrades hassling capability, we find the opposite: types with lower private strength have greater payoffs when war does not occur.

We recover the positive relationship between private strength and equilibrium payoffs when θ improves hassling effectiveness. In this case, stronger types have an advantage in both channels of bargaining leverage—hassling and the threat of war—and therefore never yield lower payoffs in equilibrium. In fact, a stronger type has a strictly greater payoff than all weaker types whenever it goes to war in equilibrium or engages in non-zero hassling.

Proposition 3. *Assume θ improves hassling effectiveness. If $\theta' < \theta''$, then $U_D(\theta') \leq U_D(\theta'')$. The inequality is strict if $\pi(\theta') = 0$ or $h(\theta') > 0$.*

If private strength is instead associated with lower hassling effectiveness, then we find an exception to the traditional positive relationship between private type and equilibrium payoff. In this case, the relationship is U-shaped. Low types, which have poor war effectiveness but relatively low costs of hassling, choose to settle rather than to fight a war in equilibrium. Among these types, lower private strength is associated with greater hassling ability, and thus a greater equilibrium payoff. At a certain level of strength, however, it becomes profitable to fight a war rather than settle. After this point, greater military strength leads to a greater payoff.

Proposition 4. *Assume θ degrades hassling effectiveness. There exists $\hat{\theta}$ such that $\pi(\theta) = 1$*

for all $\theta < \hat{\theta}$ *and* $\pi(\theta) = 0$ *for all* $\theta > \hat{\theta}$. *If* $\theta' < \theta'' < \hat{\theta}$, *then* $U_D(\theta') \geq U_D(\theta'')$ *(strictly if* $h(\theta'') > 0$*). If* $\hat{\theta} < \theta' < \theta''$, *then* $U_D(\theta') < U_D(\theta'')$.

This result illustrates the new sources of bargaining leverage that arise in crisis bargaining games with flexible responses. In ordinary crisis bargaining games, a state's sole source of bargaining power is its threat to resort to war. In our framework, hassling provides another means of shifting the balance of spoils. There is a direct effect of hassling effectiveness, where more effective types end up better off because they can afford to hassle more, thereby shifting the balance of goods in their favor. There is also an indirect effect: more effective types can pay a lower cost to undertake the same amount of hassling. Both effects contribute to the negative relationship between private strength and payoffs from non-war outcomes when θ degrades hassling effectiveness.

In analyzing private strength's relationship with both the probability of war and equilibrium payoffs in flexible-response crisis bargaining games, we have identified some conditions under which our results diverge with the traditional patterns and other conditions under which they agree. Interestingly, we find only one case in which *both* the chance of conflict and the equilibrium utility increase with private strength, as in the traditional pattern—namely, when θ increases hassling effectiveness but has an even stronger effect on war payoffs (WURI). The ordinary crisis bargaining framework may, in a sense, be considered a special case of these more general conditions. If private strength increases hassling effectiveness while having a relatively low effect on war payoffs (SURI), then equilibrium payoffs increase with θ as in traditional models, but the probability of war decreases. We see the converse pattern if private strength degrades hassling effectiveness: the probability of war increases with θ as usual, but the equilibrium utility is U-shaped.

6 Terms of Settlement

In ordinary crisis bargaining games, states can only receive better terms at the bargaining table by running a greater risk of war (Banks 1990, Lemma 3). Flexible responses like hassling introduce a new means for states to obtain more favorable terms from a settlement. Even with little or no threat of war, a state may use transgressions or hassling to shift the balance of bargaining power. But because hassling is costly, any increase in hassling must yield a commensurate benefit in the terms of settlement. Consequently, whenever we compare two types that both end up avoiding war in the equilibrium of a flexible-response crisis bargaining game, the one that hassles more must secure more favorable terms. If both hassle the same amount, then they should receive identical settlements—just like states that run the same risk of war in ordinary crisis bargaining games.

Proposition 5. *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* $V_D(\theta) < V_D(\theta')$*.*

We can say even more about the relationship between private type and the terms of settlement if we place additional structure on the model primitives. This exercise is closely related to the characterization theorems in Banks (1990) and Fey and Ramsay (2011). Specifically, if D's type space is an interval and the cost functions are Lipschitz continuous, then marginal changes in the settlement value can be derived from marginal differences in hassling costs and hassling effectiveness. See Appendix C.2 for details.

7 Amount of Hassling

We have extended the traditional crisis bargaining framework to include alternative policy options. Using the same mechanism design methodology that we have employed to study classic questions about the likelihood of conflict and the distribution of goods in this setting, we can also explore new questions about the amount of hassling we should expect to observe in a crisis. How does a state's private type affect its propensity to hassle? To what extent does the equilibrium level of hassling depend on the specifics of the bargaining protocol? We now turn to these questions.

We have shown that greater hassling effectiveness is associated with greater utility in case the crisis ends in a settlement (Proposition 3 and Proposition 4). We have also shown that the only way to improve one's payoff from a settlement is to choose greater levels of hassling (Proposition 5). Intuitively, then, it would appear to follow that more effective types hassle more in equilibrium. Yet this intuition only holds in general when the cost of hassling has decreasing differences—i.e., when more effective types not only have lower absolute costs of hassling, but also lower marginal costs.[14]

Proposition 6. *Assume D's hassling cost function has decreasing differences. If* $\pi(\theta) = \pi(\theta') = 1$ *and* θ' *has greater hassling effectiveness than* θ*, then* $h(\theta) \leq h(\theta')$.

This result leaves open two possibilities about why more effective types receive greater payoffs when war does not occur. One is that they hassle the same amount as lower types and receive the same settlement, so the higher payoff comes solely from the reduced cost of hassling. The other is that they hassle more and receive better terms at the bargaining table. As long as the decreasing differences condition holds, Proposition 6 does let us rule out a third possibility—that the more effective types hassle slightly less but at much lower cost, for a net increase in payoff despite the decrease in terms of settlement.

If θ always degrades or increases hassling effectiveness and the cost function has decreasing differences, Proposition 6 shows that the amount of hassling is a monotone function of D's private type. Can we establish more precise insights into the relationship between private type and the equilibrium selection of hassling within the class of flexible-response crisis bar-

[14] We define the condition formally in Appendix C.1.

gaining games? If θ degrades hassling effectiveness, the answer turns out to be no: virtually any weakly decreasing and continuous function $h(\theta)$ can be sustained as the equilibrium of some bargaining game.[15] For example, there is always a game form whose equilibrium entails every type of D either going to war or choosing the same amount of hassling. Consequently, different bargaining protocols can lead to radically different amounts of hassling in equilibrium, even when the underlying model primitives (war payoffs, costs of hassling, etc.) are the same.

Empirical research has found uncertain and variable effects of both conventional capabilities (Braithwaite and Lemke 2011) and lower-level capabilities (e.g., Healey and Jervis 2020) on crisis escalation. Our findings on the amount of hassling—which one might interpret as the extent of escalation short of war—may help explain these empirical inconsistencies. The equilibrium level of hassling in a given crisis situation depends not only on the underlying technological variables, but also on highly contingent features of the bargaining game. These "rules of the game" vary widely across international interactions as the product of endogenous choices by states (McKibben 2015). In the context of our findings, this variation makes it difficult to find consistent relationships between technological variables and escalation.

8 Additional Analysis

Our game-free approach allows us to characterize additional aspects of flexible-response crisis bargaining. Formal details are in Appendix C.

Possibility of peace. Is conflict an inevitable result of crisis negotiations between certain states, or can it be avoided with the right choice of bargaining protocol? Even if one particular game form ends inefficiently—whether through war or through high levels of transgressions and hassling—that does not mean such an outcome is inevitable, as states

[15] See Appendix C.3.

interacting under anarchy are not bound to follow any particular bargaining protocol.

We analyze when there is at least one game form that ends in a settlement with zero transgressions or hassling regardless of D's type. As long as the value of the prize is greater than the sum of C's expected war payoff and the strongest type of D's war payoff, there is a game form that ends with no war, transgressions, or hassling.[16] This is almost identical to the condition for a peaceful solution to the ordinary crisis bargaining problem (Fey and Ramsay 2009, 2011). Additionally, our condition for peaceful outcomes to flexible-response crisis bargaining is always satisfied when C's war payoff does not depend on D's type.[17]

Separate war payoffs and hassling costs. So far, we have treated D's war payoffs and hassling costs both as functions of the same private type, θ. We extend the framework to make these separate components of D's type, so that there is no longer necessarily a one-to-one correspondence between war payoffs and hassling effectiveness. The main results of our analysis continue to hold in this setting. An increase in D's war payoff combined with a decline in its hassling effectiveness can only increase the chance of war in equilibrium.[18] However, the effect of a simultaneous increase in D's war payoff and hassling effectiveness depends on the relative magnitudes of the differences.[19] Finally, D's equilibrium payoff increases with its war payoff and decreases with the cost of hassling, all else equal.[20]

9 Empirical Implications

9.1 When to Expect War

Our analysis offers a new perspective on what conditions lead to war. Past research suggests that if an actor possesses a strong private capability or willingness to go to war, then war is

[16] Proposition C.3 in the Appendix.
[17] Corollary C.2 in the Appendix.
[18] Proposition C.4 in the Appendix.
[19] Proposition C.5 in the Appendix.
[20] Proposition C.6 in the Appendix.

more likely to occur (Banks 1990; Fey and Ramsay 2011). In settings with multiple policy responses, we find that it is inadequate to only consider private wartime payoffs to identify when a crisis will end in war. We identify three primary cases of interest, which depend on how private hassling effectiveness and private war effectiveness are related.

In the first case, war and hassling effectiveness move together, with private type improving hassling effectiveness more than war effectiveness (formally, SURI holds). This case subverts the theoretical expectations presented in Banks (1990) and Fey and Ramsay (2011), as here being privately effective at war can lead to a lower likelihood of war. As an example, as discussed previously, in the lead-up to Operation Outside the Box, Israel possessed a technologically sophisticated private capacity for conducting an electronic warfare attack on Syria. If we only considered the lessons from traditional crisis bargaining models, this set of circumstance lends itself to a greater probability of war. Instead, the same capabilities that would have made Israel effective in a war also made Israel effective in the low-level attack they eventually carried out. To that end, this case offers an important caveat to when we should expect war to occur.

In the second case, war and hassling effectiveness move together, with private type improving war effectiveness more than hassling effectiveness (formally, WURI holds). Here, if a state is privately more effective at war, they will be more likely to go to war, despite them also being being more effective at hassling. As an example, consider the Soviet response to the Hungarian Revolution. In Hungary in 1956, anti-Soviet revolutionaries ousted Soviet-backed leadership and installed Imre Nagy as the Prime Minister, thus creating a crisis between the countries. In the crisis, at first, Nikita Khrushchev and Soviet leadership initially considered letting Nagy remain Prime Minister and engaging Hungary with some form of reconciliation (Göncz, Gati and Ash 2002, Doc. 39). However, following the outbreak of the Suez Crisis, Soviet leadership grew increasingly concerned about the international ramifications of letting the revolution in Hungary drag out without a resolution or a firm response (Boyle 2005).

In this setting, we can conceptualize the Soviet's internal desire to put forward a strong, visible response, as the Soviet leadership's private type. Note that this high private type—this strong desire to act—did not necessitate a war. Even without a war and the removal of Nagy, Soviet leadership believed they could continue exerting some influence in Hungary and could have used both overt and covert means to shape the new political reality between the two states (Göncz, Gati and Ash 2002, Doc. 39). However, Khrushchev was more willing to go to war than to engage in hassling (Göncz, Gati and Ash 2002, Doc. 53). On November 4, the Soviet Union initiated Operation Whirlwind, sending over 30,000 troops to Budapest, and suppressed the revolution in eight days. In this case, the Soviet Union possessed a private willingness to use both hassling and war, but ultimately more preferred war.

In the third case, private war and hassling effectiveness move in opposite directions: as the state becomes privately more capable or more willing to go to war, the state is less capable or less willing to hassle. In this case, when the state is privately more effective at war, war is more likely to occur. As an example of this, in the lead up to the 2003 invasion of Iraq. Here, private type here can be conceptualized as the perceived internal shortcomings of intermediate policy options. In this case, the United States possessed a strong private willingness to go to war with Iraq, and internally assessed that intermediate policy options or containment would not be effective (Coe 2018; Schram 2021).[21] The Bush Administration both internally and externally voiced concerns about Saddam Hussein's sponsorship of terror and desire to build weapons of mass destruction (McKinney 2005). While limited policy options could partially contain Iraq, the Bush Administration viewed the possibility of even low-level containment failures as potentially catastrophic and preferred the more decisive option of removing Hussein from power (Badie 2010). Additionally, while the Bush Administration did make public statements to this effect, all evidence suggests that Iraq underestimated US willingness to go to war and overestimated the likelihood of a low-level

[21] Note that this is different from the *external*, observed improvements in US hassling capabilities (as discussed in Schram (2022)).

attack (Woods et al. 2006, 15-16, 30, 96-97, 125). Through this conceptualization, because the United States possessed a high-private type—a high private willingness to go to war and a low private willingness to engage in low-level conflict—the US decision to invade and depose Saddam is consistent with what the theory expects.

Together, these theoretical and historical cases demonstrate that knowing a state's private war capabilities or the state's private willingness to go to war is not enough to determine whether bargaining will end in war. Instead, knowing the relationship between private war and hassling effectiveness is critical to understand how states will behave.

9.2 How to Use Our Results

Our analysis also offers insight into how developing military capabilities shapes international affairs (i.e., when crises will end in war) and determines how states fare in international crisis bargaining. As an illustration, consider the ongoing debate over the usefulness of aerial bombing in conflict. Earlier studies have shown that aerial bombing capabilities can be useful in hassling operations (Kreps and Fuhrmann 2011) or in a conventional war (Horowitz and Reiter 2001; Allen and Martinez Machain 2019), but may be less effective in counterinsurgency (Lyall 2013; Dell and Querubin 2018). These findings, while important in their own right, do not address how airstrike technology shapes deterrence or bargaining.[22] Each of these papers evaluates the utility of aerial weapons conditional on bargaining having already failed. It would be valuable to know how these weapons influence decision making in the lead-up to crises, how they affect the chance of war, and whether they produce windfalls for states that develop them. These questions are difficult to answer empirically: we cannot observe the counterfactual where a given state does not develop airstrike capabilities, and it is difficult to identify how these capabilities affect crisis initiation or behavior. Our theory is well suited to address these questions. As shown in Figures 3 and 4, once we identify whether private wartime capabilities improve or degrade hassling capabilities, then we can identify

[22] One notable exception is Post (2019), which analyzes airpower events as signals in compellence.

how changes in wartime capabilities affect crisis outcomes. In other words, our results can leverage existing research on how weapons systems function in war and hassling to better speak to how international crises play out.

As an illustration, consider a crisis where:

(a) A defender state has private information about its ability to conduct airstrikes.

(b) The most feasible low-level conflict option is an airstrike (e.g., Operation Desert Fox).

(c) The war option would be a conventional war (e.g., Operation Desert Storm).

Because airstrike technology is dual-use for both hassling and war, then private capabilities would have a positive effect on both war and hassling outcomes (noted by + symbols in Table 1). Thus, if (a)–(c) held, then any model formalizing this case would find that the defender's privately known ability to conduct airstrikes would improve its overall payoffs (Figure 4), but would not necessarily increase the chance of conventional war (Figure 3).

In contrast, consider a crisis where (a) and (b) held, but instead of (c):

(d) The war option would be a protracted counterinsurgency.

Based on the research cited above, airstrike capabilities would now be relatively ineffective in war (the – symbol in Table 1). If (a), (b), and (d) held, then a model formalizing this case would find that a defender state with a better private ability or willingness to conduct airstrikes would be less likely to enter into a war (Figure 3), and may or may not end up with greater payoffs (Figure 4).

A number of other conflict capabilities can be dual-use for hassling and war. We discuss these further in Table 1. Consider electronic, cyber, or anti-satellite attack capabilities. These capabilities allow states to perform well in both low-level conflict and conventional war. During Operation Outside the Box (2007), Israel disabled Syrian air defenses with an electronic warfare attack. While the full details of the electronic warfare attack have not been disclosed, any attack that allowed multiple Israeli aircraft to enter Syria and conduct a raid without harassment plausibly could have also been used to conduct a more extensive conventional attack (Katz 2010). Additionally, cyberattacks have been used as part of a conventional war (Russia-Georgia War, 2008), in hassling as an independent act (Stuxnet and Estonian cyber attacks), and in hassling as part of a cluster of other operations (the NotPetya attacks targeting Ukraine) (Buchanan 2020; Gannon et al. 2020). Similarly, developments in anti-satellite technologies have opened the possibility for disruption of GPS signals; low-level disruptions could be used for hassling, and more serious disruptions could create problems for modern air and sea warfare (Harrison et al. 2020). The second row of Table 1 summarizes this. Electronic, cyber, and anti-satellite capabilities can be productive (+) both for targeted strikes (as the "Hassling Type") and conventional war (as the "War Type") because they are dual-use technologies. Using our theory, if a state possesses a robust private cyber-attack capability, we would anticipate this state doing better in the crisis bargaining, but we are not able (without additional structure) to say whether this will increase the likelihood of war.

Existing research suggests that a number of other capabilities positively affect some forms of war and hassling. Airlift capabilities can facilitate special operations for use in low-level conflict, conventional war, and irregular warfare (Bolkcom 2007). Third-party support for violent non-state actors has been used for both hassling and war (Schultz 2010; Schram 2021). Control of the information environment, intelligence operations, and influence operations have both been used in extensively in hearts-and-minds counterinsurgencies (Shapiro and Weidmann 2015).

Of course, some capabilities are only effective in certain forms of conflict. Airpower capabilities may be useful in conventional war or certain forms of low-level conflict, but less useful in winning protracted counterinsurgencies (Lyall 2013; Dell and Querubin 2018). Furthermore, if states invest in electronic warfare capabilities to disrupt air defense systems, cyberwarfare

Capability/Willingness Mechanism	Hassling Type	War Type	Rationale
Airstrikes/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
Information/Influence Operations	Undermining domestic authority/promoting discord (+)	Irregular War (+)	Dual-Use Technology
Information/Influence Operations	Undermining domestic authority/promoting discord (+)	Conventional War (–)	Specialized Technology/Budgetary Issues
Air, 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 mechanisms affect hassling and war capacities. The + symbol indicates that the mechanism improves the corresponding hassling operations or war type; the – symbol denotes a detriment.

tools to disrupt industrial or nuclear processes, or anti-satellite tools, they may invest less in the capabilities needed to win population-centric counterinsurgencies (Berman, Shapiro and Felter 2011). Similarly, influence or information operations may be critical to certain kinds of hassling or counterinsurgency contexts, but may be less important in conflicts that rely heavily on conventional firepower.[23] Finally, many conventional technologies are unsuited to combat hassling. For example, the U.S. Navy, Marine Force, and Coast Guard jointly issued a report suggesting that conventional naval vessels may not be as effective in deft handling of gray zone attacks in the Pacific theater (Berger, Gilday and Schultz 2020; Owen 2021).

Factors besides military capabilities could also alter leaders' willingness to opt for certain forms of conflict. 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). This effect could alter a leader's preferences, causing them to prefer war over hassling operations. Alternatively, domestic political economy considerations can also shape leaders' incentives. For example, in 1954, the United States provided arms, funds, and training to Guatemalan rebels who overthrew Jacobo Árbenz and installed right-wing dictator Castillo Armas—a move that benefited the politically connected United Fruit Company (Kinzer 2007, 125–147) in ways that sanctions would not. Similarly, while "blood for oil" may not fully explain the 2003 Iraq invasion (Stokes 2007), oil market considerations could have still increased leaders' private willingness to go to war rather than implement half-measures and hassle. Conversely, relevant domestic actors might prefer low-level action over war, such as those who would benefit from economic sanctions or tariffs.

10 Conclusion

Why does war occur? Using crisis bargaining models, a broad set of influential game theoretic research has identified a remarkably consistent answer to this foundational question: when

[23]For example, Operation Desert Storm (a conventional conflict) placed less emphasis on "winning hearts and minds" than did Operation Iraqi Freedom (a protracted counterinsurgency campaign) (Nagl et al. 2008).

one state possesses a greater private ability or willingness to go to war, the game is more likely to end in war. However, crisis bargaining models make the common simplifying assumption that the game ends in either war or in peace. In practice, we know that this is not true, as states face a broad array of policy options when in a crisis. The question, then, is whether this simplifying assumption is innocuous, or if this simplifying assumption shapes how these models answer the war puzzle.

To address this, we put forward a new class of models—flexible-response crisis bargaining models—and conducted a comprehensive analysis of them using the tools of mechanism design. These flexible-response crisis bargaining models represent a useful adaptation of the standard, dichotomous crisis bargaining framework, where war and peace are the only possible outcomes. In our framework, states can engage in a continuum of conflict operations; this better captures the conditions of actual international crises, in which states select from an array of options like sanctions or gray zone operations. Rather than solve a single game form, we have identified the properties shared by all equilibria in the full class of flexible-response crisis bargaining games. This general analysis allows us to be confident that our results are not driven by specifics of the game form or a specific equilibrium, but are generalizable to all flexible-response crisis bargaining models.

Our most surprising results are those that differ from the Banks (1990) and Fey and Ramsay (2011) monotonicity results. While existing research has shown that improved private war capabilities or a greater private willingness to go to war can never decrease the likelihood of war, we demonstrate that this relationship is more nuanced when war capabilities can also benefit low-level conflict capabilities. Similarly, while Banks (1990) and Fey and Ramsay (2011) have shown that an improved private ability to conduct war *always* produces a greater utility, we find that these results do not necessarily hold when a robust ability to go to war can reduce an actor's ability to effectively sanction or hassle.

One key implication of our research is that the concepts of "strength" and "power" in international relations are more complex than previous research has indicated. Within traditional crisis bargaining models, as discussed in Banks (1990) and Fey and Ramsay (2011), strength and power correspond to a state's ability to perform well in war. Within traditional crisis bargaining models, these concepts have theoretical importance, driving equilibrium outcomes and payoffs (Ramsay 2017). Furthermore, in conjunction with this theoretical work, empirical international relations scholars often use measures of wartime capacity as key explanatory variables for international interactions (see Singer 1988; Kadera and Sorokin 2004; Carroll and Kenkel 2019). But within our flexible-response crisis bargaining framework, the old paradigm of using wartime capabilities as measures of strength and power is less informative; after all, as Figure 2 shows, the "weaker" type (by traditional metrics) may outperform the "stronger" one. Instead, international interactions are shaped by a state's full range of capabilities, including their ability to conduct sanctions, their capacity to support third-party militants, their capability to execute effective cyberattacks, and their aptness for a wide range of other coercive policy options. In summary, our research advocates a broader approach to understanding strength and power: a state's coercive power is not a function of wartime strength alone, but rather is a holistic measure, taking into consideration the broad range of policies that states have at their disposal.

Theoretically, there is more work to be done to understand how flexible policy tools affect the outcomes of international crises. First, one could extend our flexible-response crisis bargaining framework to scenarios where observations of transgressions or hassling decisions are noisy. Such an extension could capture, for example, settings where a transgression is imperfectly observed or where attribution is difficult. These include cyberwarfare with attribution problems (Baliga, Bueno de Mesquita and Wolitzky 2020), as well as when the hidden development of technological capabilities is itself the transgression (e.g., Meirowitz et al. 2019). Second, our initial work here has treated the transgression and hassling options as choices along a single dimension. In practice, states may choose among many distinct instruments of flexible responses. Third, we treat hassling and war as distinct policy options.

Future game-form free analyses can do more to treat distinct policy options as related—for example, hassling choices in bargaining may shape war payoffs (as in Qiu 2022*a*), or forms of low-level conflict may increase or decrease the likelihood of an elevated response (as formalized in Powell 2015 and described in Schelling 1980 and Kahn 2017). By building out a more sophisticated framework with multidimensional flexible responses, future research could provide even more accurate descriptions of policy choices in international crises.

References

Akçay, Erol, Adam Meirowitz, Kristopher W. Ramsay and Simon A. Levin. 2012. "Evolution of Cooperation and Skew Under Imperfect Information." *Proceedings of the National Academy of Sciences* 109(37):14936–14941.

Allen, Susan Hannah and Carla Martinez Machain. 2019. "Understanding the Impact of Air Power." *Conflict Management and Peace Science* 36(5):545–558.

Ashworth, Scott and Ethan Bueno de Mesquita. 2006. "Monotone comparative statics for models of politics." *American Journal of Political Science* 50(1):214–231.

Badie, Dina. 2010. "Groupthink, Iraq, and the war on terror: Explaining US policy shift toward Iraq." *Foreign Policy Analysis* 6(4):277–296.

Baker, William D. and John R. Oneal. 2001. "Patriotism or Opinion Leadership? The Nature and Origins of the "Rally 'Round the Flag" Effect." *Journal of Conflict Resolution* 45(5):661–687.

Baliga, Sandeep, Ethan Bueno de Mesquita and Alexander Wolitzky. 2020. "Deterrence with Imperfect Attribution." *American Political Science Review* 114(4):1155–1178.

Baliga, Sandeep and Tomas Sjöström. 2020. "The Strategy and Technology of Conflict." *Journal of Political Economy* 128(8):3186–3219.

Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34(3):599–614.

Bapat, Navin A. 2006. "State Bargaining with Transnational Terrorist Groups." *International Studies Quarterly* 50(1):213–229.

Bapat, Navin A and Bo Ram Kwon. 2015. "When are sanctions effective? A bargaining and enforcement framework." *International Organization* 69(1):131–162.

Bas, Muhammet A and Robert Schub. 2017. The Theoretical and Empirical Approaches to Uncertainty and Conflict in International Relations. In *Oxford Research Encyclopedia of Politics*.

Benson, Brett V. and Bradley C. Smith. 2022. "Commitment Problems in Alliance Formation." *American Journal of Political Science* .

Berger, David H., Michael M. Gilday and Karl L. Schultz. 2020. *Advantage at Sea: Prevailing with Integrated All-Domain Naval Power*. United States Department of Defense.

Berman, Eli and David A. Lake. 2019. *Proxy Wars: Suppressing Violence through Local Agents*. Cornell University Press.

Berman, Eli, Jacob N Shapiro and Joseph H Felter. 2011. "Can hearts and minds be bought? The economics of counterinsurgency in Iraq." *Journal of Political Economy* 119(4):766–819.

Bolkcom, Christopher. 2007. Military Airlift: C-17 Aircraft Program. Library of Congress.

Boyle, Peter G. 2005. "The Hungarian Revolution and the Suez Crisis." *History* 90(300):550–565.

Braithwaite, Alex and Douglas Lemke. 2011. "Unpacking escalation." *Conflict Management and Peace Science* 28(2):111–123.

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

Carroll, Robert J and Brenton Kenkel. 2019. "Prediction, proxies, and power." *American Journal of Political Science* 63(3):577–593.

Carter, David B. 2012. "A blessing or a curse? State support for terrorist groups." *International Organization* 66(1):129–151.

Chassang, Sylvain and Gerard Padró I Miquel. 2010. "Conflict and Deterrence under Strategic Risk." *Quarterly Journal of Economics* 125(4):1821–1858.

Cobzaş, Ştefan, Radu Miculescu and Adriana Nicolae. 2019. *Lipschitz Functions*. Springer.

Coe, Andrew J. 2018. "Containing Rogues: A Theory of Asymmetric Arming." *Journal of Politics* 80(4):1197–1210.

Debs, Alexandre and Nuno P. Monteiro. 2014. "Known Unknowns: Power Shifts, Uncertainty, and War." *International Organization* 68(01):1–31.

Dell, Melissa and Pablo Querubin. 2018. "Nation Building through Foreign Intervention: Evidence from Discontinuities in Military Strategies." *Quarterly Journal of Economics* 133(2):701–764.

Di Lonardo, Livio and Scott Tyson. 2022*a*. "Deterrence and Preventive Sanctions." *Working Paper* .

Di Lonardo, Livio and Scott Tyson. 2022*b*. "Political Instability and the Failure of Deterrence." *Journal of Politics* 84(1):180–193.

Dunford, Joseph. 2016. "Gen. Dunford's Remarks and Q&A.". Center for Strategic and International Studies.

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

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

Fey, Mark and Brenton Kenkel. 2021. "Is an Ultimatum the Last Word on Crisis Bargaining?" *Journal of Politics* 83(1):87–102.

Fey, Mark and Kristopher W. Ramsay. 2009. "Mechanism Design Goes to War: Peaceful Outcomes with Interdependent and Correlated Types." *Review of Economic Design* 13(3):233.

Fey, Mark and Kristopher W Ramsay. 2011. "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict." *American Journal of Political Science* 55(1):149–169.

Fuhrmann, Matthew and Sarah E Kreps. 2010. "Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941–2000." *Journal of Conflict Resolution* 54(6):831–859.

Gannon, Andres J., Erik Gartzke, Jon R. Lindsay and Peter Schram. 2020. "The Shadow of Deterrence." Typescript, University of California, San Diego.
URL: https://jandresgannon.com/research

Gartzke, Erik and Jon R. Lindsay. 2015. "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace." *Security Studies* 24(2):316–348.

Gartzke, Erik and Jon R. Lindsay. 2017. "Thermonuclear Cyberwar." *Journal of Cybersecurity* 3(1):37–48.

Göncz, Árpád, Charles Gati and Timothy Garton Ash. 2002. *The 1956 Hungarian revolution: a history in documents.* Central European University Press.

Gurantz, Ron and Alexander V. Hirsch. 2017. "Fear, Appeasement, and the Effectiveness of Deterrence." *Journal of Politics* 79(3):1041–1056.

Harrison, Todd, Kaitlyn Johnson, Thomas G. Roberts, Tyler Way and Makena Young. 2020. *Space threat assessment 2020.* Center for Strategic and International Studies.

Healey, Jason and Robert Jervis. 2020. "The Escalation Inversion and Other Oddities of Situational Cyber Stability." *Texas National Security Review* .

Horowitz, Michael and Dan Reiter. 2001. "When Does Aerial Bombing Work? Quantitative Empirical Tests, 1917–1999." *Journal of Conflict Resolution* 45(2):147–173.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton University Press.

Kadera, Kelly and Gerald Sorokin. 2004. "Measuring national power." *International Interactions* 30(3):211–230.

Kahn, Herman. 2017. *On Escalation: Metaphors and Scenarios*. Routledge.

Katz, Yaakov. 2010. "And they struck them with blindness." *The Jerusalem Post* .
URL: https://www.jpost.com/Magazine/Features/And-they-struck-them-with-blindness

Kinzer, Stephen. 2007. *Overthrow: America's Century of Regime Change from Hawaii to Iraq*. Macmillan.

Kreps, Sarah E. and Matthew Fuhrmann. 2011. "Attacking the Atom: Does Bombing Nuclear Facilities Affect Proliferation?" *Journal of Strategic Studies* 34(2):161–187.

Kydd, Andrew H. and Roseanne W. McManus. 2017. "Threats and Assurances in Crisis Bargaining." *Journal of Conflict Resolution* 61(2):325–348.

Lanoszka, Alexander. 2016. "Russian Hybrid Warfare and Extended Deterrence in Eastern Europe." *International Affairs* 92(1):175–195.

Leventoğlu, Bahar and Ahmer Tarar. 2008. "Does Private Information Lead to Delay or War in Crisis Bargaining?" *International Studies Quarterly* 52(3):533–553.

Levy, Jack S. 2011. "Preventive War: Concept and Propositions." *International Interactions* 37(1):87–96.

Lin-Greenberg, Erik. 2022*a*. "Evaluating Escalation: Conceptualizing Escalation in an Era of Emerging Military Technologies." *Journal of Politics* .

Lin-Greenberg, Erik. 2022*b*. "Wargame of drones: remotely piloted aircraft and crisis escalation." *Journal of Conflict Resolution* 66(10):1737–1765.

Liu, Linqun. 2021. "Domestic Constraints in Crisis Bargaining." Typescript, University of Chicago.
URL: https://sites.google.com/site/liqunliu90/home

Lyall, Jason. 2013. "Airpower and Coercion in Counterinsurgency Wars: Evidence from Afghanistan." Typescript, Yale University.
URL: http://www.jasonlyall.com/wp-content/uploads/2013/10/AFGHAN_AIRSTRIKES.pdf

Mazarr, Michael J. 2015. "Struggle in the Gray Zone and World Order." *War on the Rocks* .
URL: https://warontherocks.com/2015/12/struggle-in-the-gray-zone-and-world-order/

McCormack, Daniel and Henry Pascoe. 2017. "Sanctions and Preventive War." *Journal of Conflict Resolution* 61(8):1711–1739.

McKibben, Heather Elko. 2015. *State Strategies in International Bargaining: Play by the Rules or Change Them?* Cambridge University Press.

McKinney, Bruce C. 2005. "Decision-Making: Case Studies including the Bush Administration's Decision to Invade Iraq." *Peace Research* 37(2):11–25.

Meirowitz, Adam, Massimo Morelli, Kristopher W. Ramsay and Francesco Squintani. 2019. "Dispute Resolution Institutions and Strategic Militarization." *Journal of Political Economy* 127(1):378–418.

Milgrom, Paul and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70(2):583–601.

Morrow, James D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33(4):941–972.

Morrow, James D. and Jessica S. Sun. 2020. "15 Models of Interstate Conflict." *The SAGE Handbook of Research Methods in Political Science and International Relations* .

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47(1):61–73.

Nagl, John A, James F Amos, Sarah Sewall, David H Petraeus et al. 2008. *The US Army/Marine Corps Counterinsurgency Field Manual*. Number 3-24 University of Chicago Press.

Opall-Rome, Barbara. 2018. "Declassified: How an Israeli operation derailed Syria's nuclear weapons drive." *Defense News* .
URL: https://www.defensenews.com/global/mideast-africa/2018/03/20/just-declassified-how-an-israeli-operation-derailed-syrias-nuclear-weapons-drive

Owen, Dan. 2021. "Time for a Fully Integrated Dual-Fleet Force Model." *War on the Rocks* .
URL: https://warontherocks.com/2021/03/time-for-a-fully-integrated-dual-fleet-force-m

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

Post, Abigail. 2019. "Flying to Fail: Costly Signals and Air Power in Crisis Bargaining." *Journal of Conflict Resolution* 63(4):869–895.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60(1):169–203.

Powell, Robert. 2015. "Nuclear Brinkmanship, Limited War, and Military Power." *International Organization* 69(3):589–626.

Qiu, Xiaoyan. 2022*a*. "Rebel strategies and the prospects for peace." *American journal of political science* 66(1):140–155.

Qiu, Xiaoyan. 2022*b*. "State Support for Rebels and Interstate Bargaining." *American Journal of Political Science* .

Ramsay, Kristopher W. 2017. "Information, Uncertainty, and War." *Annual Review of Political Science* 20:505–527.

Reiter, Dan. 2005. "Preventive Attacks Against Nuclear Programs and the "Success" at Osiraq." *Nonproliferation Review* 12(2):355–371.

Schelling, Thomas C. 1980. *The strategy of conflict*. Harvard university press.

Schram, Peter. 2021. "Hassling: How States Prevent a Preventive War." *American Journal of Political Science* 65(2):294–308.

Schram, Peter. 2022. "When Capabilities Backfire: How Improved Hassling Capabilities Produce Worse Outcomes." *The Journal of Politics* 84(4).

Schultz, Kenneth A. 2010. "The Enforcement Problem in Coercive Bargaining: Interstate Conflict over Rebel Support in Civil Wars." *International Organization* 64(2):281–312.

Shapiro, Jacob N and Nils B Weidmann. 2015. "Is the phone mightier than the sword? Cellphones and insurgent violence in Iraq." *International Organization* 69(2):247–274.

Singer, J David. 1988. "Reconstructing the correlates of war dataset on material capabilities of states, 1816–1985." *International Interactions* 14(2):115–132.

Slantchev, Branislav L. 2011. *Military threats: the costs of coercion and the price of peace*. Cambridge University Press.

Spaniel, William. 2019. *Bargaining over the Bomb: The Successes and Failures of Nuclear Negotiations*. Cambridge University Press.

Spaniel, William and Iris Malone. 2019. "The Uncertainty Trade-off: Reexamining Opportunity Costs and War." *International Studies Quarterly* 63(4):1025–1034.

Stokes, Doug. 2007. "Blood for Oil? Global Capital, Counter-Insurgency and the Dual Logic of American Energy Security." *Review of International Studies* 33(2):245–264.

Tarar, Ahmer. 2016. "A Strategic Logic of the Military Fait Accompli." *International Studies Quarterly* 60(4):742–752.

Tarar, Ahmer. 2021. "Risk Preferences, Uncertainty, and War." *International Interactions* 48(2):1–25.

Woods, Kevin M, Michael R Pease, Mark E Stout, Williamson Murray and James G Lacey. 2006. Iraqi Perspectives Project. Technical report United States Joint Forces Command, Norfolk, VA.

Online Appendix

Contents

A Equilibrium and Direct Mechanism

Consider a game form $G = (\mathcal{B}_C, \mathcal{B}_D, g)$, where $g = (\pi^g, V_C^g, V_D^g)$. A pure strategy perfect Bayesian equilibrium in G is an assessment consisting of:

- C's bargaining strategy, $b_C^* \in \mathcal{B}_C$.[1]
- C's transgression, $t^*(b_D, h)$. We write this as a mapping $t^* : \mathcal{B}_D \times \mathcal{H} \to \mathcal{T}$ because C's transgression may depend on D's bargaining strategy or hassling.
- Each type of D's bargaining action $b_D^*(\theta)$ and hassling $h^*(\theta)$, where $b_D^* : \Theta \to \mathcal{B}_D$ and $h^* : \Theta \to \mathcal{H}$.
- C's beliefs about θ at each information set.

[1] We allow b_C to represent a complete contingency plan in an extensive form game, so we do not explicitly write this as a function of D's bargaining strategy or hassling.

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

- Hassling level:
$$h(\theta) = h^*(\theta).$$
- Whether the interaction ends in a settlement:
$$\pi(\theta) = \pi^g \left(b_C^*, b_D^*(\theta)\right).$$
- D's spoils from bargaining in case bargaining prevails:
$$V_D(\theta) = V_D^g \left(t^*(b_D^*(\theta), h^*(\theta)), h^*(\theta), b_C^*, b_D^*(\theta)\right).$$

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

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

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

B.1 Proof of Lemma 1

This result, as well as Lemma 2 below, depends on the following auxiliary result on the expected utility from mimicking various types in the absence of hassling.

Lemma B.1. *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 $h = 0$:

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

□

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

Proof. Lemma B.1 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 $\pi(\theta') \geq \pi(\theta'')$. □

B.2 Proof of Proposition 1

Proposition 1. *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. □

B.3 Proof of Proposition 2

Proposition 2. *Assume θ improves hassling effectiveness, and let $\theta' < \theta''$. If (WURI) holds, then $\pi(\theta') \geq \pi(\theta'')$. If (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 $\pi(\theta'') = 0$). □

B.4 Proof of Lemma 2

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

Proof. Immediate from Lemma B.1. □

B.5 Proof of Proposition 3

Proposition 3. *Assume θ improves hassling effectiveness. If $\theta' < \theta''$, then $U_D(\theta') \leq U_D(\theta'')$. The inequality is strict if $\pi(\theta') = 0$ or $h(\theta') > 0$.*

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

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

□

If $\pi(\theta') = 0$, then the second inequality above is strict. The same is true if $\pi(\theta') = 1$ and $h(\theta') > 0$.

B.6 Proof of Proposition 4

The proof depends on a more general property of hassling effectiveness and equilibrium utility:

Lemma B.2. *If $\pi(\theta') = \pi(\theta'') = 1$ and θ' has greater hassling effectiveness than θ'', then $U_D(\theta') \geq U_D(\theta'')$ (strictly if $h(\theta'') > 0$).*

Proof. By (IC), we have

$$U_D(\theta') \geq \underbrace{V_D(\theta'') - K_D(h(\theta''), \theta')}_{\Phi_D(\theta'' \mid \theta')} \geq V_D(\theta'') - K_D(h(\theta''), \theta'') = U_D(\theta''). \tag{B.1}$$

The second inequality of Equation B.1 is strict if $h(\theta'') > 0$. □

Proposition 4. *Assume θ degrades hassling effectiveness. There exists $\hat{\theta}$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$. If $\theta' < \theta'' < \hat{\theta}$, then $U_D(\theta') \geq U_D(\theta'')$ (strictly if $h(\theta'') > 0$). If $\hat{\theta} < \theta' < \theta''$, then $U_D(\theta') < U_D(\theta'')$.*

Proof. The claim about $\hat{\theta}$ follows from Proposition 1. The next claim then follows from Lemma B.2. The final claim follows because W_D is strictly increasing in θ. □

B.7 Proof of Proposition 5

Proposition 5. *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.8 Proof of Proposition 6

Proposition 6. *Assume D's hassling cost function has decreasing differences. If $\pi(\theta) = \pi(\theta') = 1$ and θ' has greater hassling effectiveness than θ, then $h(\theta) \leq h(\theta')$.*

Proof. (IC) implies:

$$\begin{aligned} V_D(\theta) - K_D(h(\theta), \theta) &\geq V_D(\theta') - K_D(h(\theta'), \theta), \\ V_D(\theta') - K_D(h(\theta'), \theta') &\geq V_D(\theta) - K_D(h(\theta), \theta'). \end{aligned}$$

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 $h(\theta) \leq h(\theta')$. □

C Additional Analysis

C.1 Sufficient Condition for WURI

With additional conditions on the model primitives, we can ensure that the war utility is increasing relative to the hassling utility, meaning that the likelihood of conflict increases with D's type. In particular, we need the hassling cost function to have decreasing differences, which implies that types with lower absolute costs also have lower marginal costs. This condition ensures that D's settlement utility has the single-crossing property, allowing us to characterize monotone comparative statics without imposing specific functional forms (Ashworth and Bueno de Mesquita 2006).

Definition 4. The cost function K_D has decreasing differences in h and θ if

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

In addition to the cost function having decreasing differences, we also need 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 war regardless of the exact negotiating protocol employed.

Lemma C.3. *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). □

There is not an analogous sufficient condition for the settlement utility to be relatively increasing. The obstacle here is our assumption that $h = 0$ has the same cost (zero) for all types. This means the marginal effect of θ on hassling costs is zero for $h = 0$, ruling out any way 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 $\theta < \theta'$. In any equilibrium where this condition holds, we cannot have low types settle while high types go to war.

C.2 Terms of Settlement

Placing additional structure on the model primitives allows us to be even more specific about the relationship between the extent of hassling and the value of settlement. First, we will assume D's type is drawn from an interval, $\theta \in [\underline{\theta}, \bar{\theta}]$. This requirement effectively allows us to strengthen the incentive compatibility conditions for equilibrium, as we can now say that every type of D must find it unprofitable to mimic the strategy of a marginally stronger or weaker type. Second, we will assume a degree of differentiability (and thus continuity) in the relationship between private type and war payoffs, as well as that between private type, hassling amount, and the cost of hassling.[2] These assumptions allow us to characterize local incentive compatibility conditions—the lack of incentive to mimic a slightly lower or higher type—in terms of derivatives of the war payoff and hassling cost functions. We refer to the collection of these assumptions as bounded variation conditions, or (BV).

Definition 5. The model has bounded variation if W_D and K_D are differentiable and

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

where $M_W \geq 0$ and $M_D \geq 0$ are real constants.

The bounded variation conditions allow us to apply the "envelope theorem" commonly employed in mechanism design analyses of crisis bargaining models (Banks 1990; Fey and Ramsay 2011).[3] The following lemma shows how these conditions imply differentiability of Φ_D with respect to D's true type.

Lemma C.4. *Assume (BV) holds. For all $\theta, \theta' \in \Theta$, Φ_D is differentiable with respect to θ, and*

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

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. □

[2] Specifically, we assume W_D and K_D are Lipschitz continuous, a weaker requirement than continuous differentiability.

[3] The M_W and M_D terms are simply real-valued constants.

Given just a few endogenous elements of the equilibrium, we can determine every type's equilibrium payoff, which in turn will allow us to back out the precise terms of settlement for each type that ends up avoiding war. All we need to know are the lowest type's equilibrium utility,[4] whether each type ends up at war, and the extent of hassling carried out by those types that end up avoiding war. The following proposition gives a precise statement of $U_D(\theta)$ as a function of these equilibrium quantities.

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

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

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),

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

at each point where U_D is differentiable. Furthermore, (BV) implies Φ_D is Lipschitz continuous in θ. The claim then follows from Lemma C.4 and Milgrom and Segal (2002, Corollary 1). □

The complex integral stated in Proposition C.1 boils down to two essential facts about the relationship between private type and equilibrium payoffs. First, for types that go to war, the marginal increase in utility as θ increases is, naturally, the same as the marginal increase in war payoff. Second, among those that settle in equilibrium, the marginal change in equilibrium payoff depends exactly on the marginal effect of private type on the cost of hassling. This second fact is what allows us to identify the value of settlement once we know which types settle and how much they spend on hassling. Suppose there is an interval of types $[\theta', \theta''] \subseteq \Theta$ which all choose to settle in equilibrium. We can use Proposition C.1 to characterize how the terms of the bargain differ between the poles of this interval:

$$V_D(\theta'') - V_D(\theta') = \underbrace{K_D(h(\theta''), \theta'') - K_D(h(\theta'), \theta')}_{\text{cost difference}} - \underbrace{\int_{\theta'}^{\theta''} \frac{\partial K_D(h, \theta)}{\partial \theta}\bigg|_{h=h(\theta)} d\theta}_{\text{effectiveness premium}}.$$

The first term here, the cost difference, is a baseline: incentive compatibility means the terms of settlement must adjust at least roughly in accordance with the cost paid for hassling. If not for the second term, the effectiveness premium, then each type $\theta \in [\theta', \theta'']$ would have the same equilibrium payoff. The effectiveness premium represents the additional benefit that states with greater hassling effectiveness can extract from bargaining. For example, suppose θ improves hassling capability, so K_D is decreasing in θ. Then the effectiveness term will be positive, more so if there is a steep relationship between private type and the marginal cost

[4] In fact, all that is necessary is to know the equilibrium payoff of a single type, not necessarily that of $\underline{\theta}$.

of hassling. Conversely, there will be no benefit from the effectiveness premium if hassling does not take place. If $h(\theta) = 0$ for all $\theta \in [\theta', \theta'']$, then the cost difference and effectiveness premium are both zero, and all types in this interval receive the same settlement. We are then back in the world of ordinary crisis bargaining games, where the only source of bargaining leverage is the threat of war.

In the special case where the model has bounded variation and θ degrades hassling capability, we can further identify the value of settlement. We know from Proposition 1 that any equilibrium in this case will be characterized by a cutpoint $\hat{\theta} \in \Theta$, with all types below $\hat{\theta}$ settling in equilibrium and all types above it going to war. We can then characterize the settlement value for all types below the cutpoint in terms of the cutpoint type's war payoff and the choice of hassling by each intermediate type.

Corollary C.1. *Assume θ degrades hassling effectiveness and (BV) holds. There exists $\hat{\theta} \in \Theta$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$. If $\underline{\theta} < \hat{\theta} < \bar{\theta}$, then for all $\theta_0 < \hat{\theta}$,*

$$V_D(\theta_0) = W_D(\hat{\theta}) + K_D(h(\theta_0), \theta_0) + \int_{\theta_0}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta. \tag{C.2}$$

Proof. Proposition 1 implies the existence of the cutpoint $\hat{\theta}$. We then have $U_D(\theta) = W_D(\theta)$ for all $\theta > \hat{\theta}$. (BV) implies that W_D is continuous and Proposition C.1 implies that U_D is continuous, so if $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ we have $U_D(\hat{\theta}) = W_D(\hat{\theta})$. For $\theta_0 < \hat{\theta}$, Proposition C.1 then gives

$$\begin{aligned} V_D(\theta_0) &= U_D(\theta_0) + K_D(h(\theta_0), \theta_0) \\ &= U_D(\hat{\theta}) + K_D(h(\theta_0), \theta_0) + \left[U_D(\theta_0) - U_D(\hat{\theta}) \right] \\ &= W_D(\hat{\theta}) + K_D(h(\theta_0), \theta_0) + \int_{\theta_0}^{\hat{\theta}} \left. \frac{\partial K_D(h, \theta)}{\partial \theta} \right|_{h=h(\theta)} d\theta. \end{aligned}$$

□

This result is useful for two reasons. First, it shows that we need relatively little information about the equilibrium to identify settlement values in flexible-response crisis bargaining games when θ degrades hassling capability. As long as we know the lowest type that goes to war and the amount of hassling exerted by each lower type, we can derive the exact settlement level in equilibrium. Importantly, if two bargaining games result in the same cutpoint type and the same amount of hassling below the cutpoint, they will also result in the exact same terms of settlement for each type of D, even if the bargaining processes themselves are quite dissimilar. Second, the necessary condition provided by Corollary C.1 turns out to be sufficient. Specifically, for any non-decreasing hassling plan $h(\theta)$, if we allocate settlement values according to the given formula for V_D, the resulting mechanism is incentive compatible and satisfies voluntary agreements.

C.3 Extent of Hassling

Here we formally state and prove the result mentioned in the text: If θ degrades hassling effectiveness, then any weakly decreasing and absolutely continuous hassling plan can be supported as the equilibrium of some game form.

Proposition C.2. *Assume θ degrades hassling effectiveness and (BV) and (DD) hold. Let h be any non-increasing and absolutely continuous function from $[\underline{\theta}, \bar{\theta}]$ into $\mathcal{H}$. Take any $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$, let $\pi(\theta) = \mathbf{1}\{\theta \leq \hat{\theta}\}$, and let $V_D(\theta)$ be defined by Equation C.2. The direct mechanism (h, π, V_D) 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 (Cobzaş, Miculescu and Nicolae 2019, 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

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

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

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

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

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

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

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

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. □

This result demonstrates that incentive compatibility and voluntary agreements alone place no restrictions on the pattern of hassling across types beyond the fact that more effective types cannot engage in less of it. Consequently, the specifics of the relationship between effectiveness and the degree of hassling are model-dependent. For example, we cannot determine from the primitives alone whether all types hassle to the same degree, or whether there is some separation in levels of hassling. This will depend on how bargaining takes place and the precise effects of hassling choices on non-war payoffs.

C.4 Possibility of Peace

In the context of flexible-response crisis bargaining, we must be explicit about what peace entails. At a minimum, as in ordinary crisis bargaining games, the game must end with a negotiated settlement for all types of D. Furthermore, because transgressions and hassling may be interpreted as forms of low-level conflict, our primary focus is 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 say an equilibrium is always peaceful if it meets these conditions.

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 C.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 $t \in \mathcal{T}$.
2. D chooses $h \in \mathcal{H}$.
3. C and D simultaneously choose $b_C \in \{0, 1\}$ and $b_D \in \{0, 1\}$.

War occurs if either player chooses $b_i = 1$:

$$\pi^g(b_C, b_D) = \mathbf{1}\{b_C + b_D > 0\}.$$

Baseline payoffs are divided according to the war payoff for the strongest type of D, with each player paying a penalty $Z > 0$ if the other chooses a non-zero flexible response:

$$V_C^g(t, h, b_C, b_D) = 1 - W_D(\bar{\theta}) - Z \cdot \mathbf{1}\{h > 0\},$$
$$V_D^g(t, h, b_C, b_D) = W_D(\bar{\theta}) - Z \cdot \mathbf{1}\{t > 0\}.$$

Assume Z is arbitrarily large—specifically, $Z > \max\left\{1 - W_D(\bar{\theta}) - \hat{W}_C, W_D(\bar{\theta}) - W_D(\underline{\theta})\right\}$. Note that this game has voluntary agreements, as each player can guarantee war by choosing $b_i = 1$.

We claim that the following strategy profile constitutes an equilibrium of this game:

1. C chooses $t = 0$.
2. Following all choices of t, D chooses $h = 0$.
   - C's beliefs about D's type remain at the prior following all (t, h).
3. C chooses $b_C = 1$ if and only if $h > 0$. D chooses $b_D = 1$ if and only if $t > 0$.

Because Z was chosen to be arbitrarily large, the choices of bargaining strategy when $h > 0$ or $t > 0$ are clearly best responses. Now consider the case when $h = t = 0$. For C, deviating to $b_C = 1$ would result in an expected payoff of $\hat{W}_C \leq 1 - W_D(\bar{\theta})$, which is unprofitable. For any type of D, deviating to $b_D = 1$ would result in a payoff of $W_D(\theta) \leq W_D(\bar{\theta})$, which is unprofitable. The bargaining strategies therefore comprise a Bayesian Nash equilibrium. Moving up the game tree, a deviation by C to $t > 0$ or by D to $h > 0$ would result in war, which we have just shown is worse than the payoffs from the proposed strategies. Finally, note that C's beliefs are updated in accordance with Bayes' rule whenever possible. Therefore, the proposed strategy profile is a perfect Bayesian equilibrium. □

The condition of Proposition C.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 strongest type of D while giving C at least its expected war payoff. Because strong types of D are rare in this setting, the equilibrium chance of conflict will be low, but not zero.

If C's war payoff is independent of D's type (i.e., D's type only affects its cost of war, not its probability of victory), then the condition of Proposition C.3 always holds. 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 C.2. *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 C.3, as $\hat{W}_C + W_D(\bar{\theta}) = 1 - c_C - c_D(\bar{\theta}) < 1$. □

Because we have assumed transgressions and hassling only affect payoffs from negotiations, our conditions for always peaceful equilibria do not depend on the costs of these options. If the war payoffs were also functions of t and h, then the players' reservation values for conflict would depend on their marginal effects and costs. Therefore, in order for flexible responses to materially affect the prospects for peace, responses must shape payoffs in war as well as peace.[5]

C.5 Two-Dimensional Type

In the main framework, we treat D's war payoff and hassling costs as both being functions of a single type parameter, θ. We now relax this assumption, allowing separate types to control D's war payoff and hassling costs. In contrast with the main framework, this allows for situations where two types of D might have identical war payoffs but different hassling costs. Our goal is to show that our qualitative findings on the relationship between private types and the probability of war still hold up in this more realistic environment.

In the extended framework, let D's type be a pair (θ, ξ). As in the main framework, $\theta \in \Theta \subseteq \mathbb{R}$ determines D's war costs. However, D's hassling costs are now a function of $\xi \in \Xi \subseteq \mathbb{R}$. Formally, the cost term for D in case of settlement is now $K_D(h, \xi)$, which strictly increases in both parameters.[6] In terms of our main framework analysis, lower values of ξ correspond to greater hassling effectiveness, and so on. We now write D's utility in a given game form g as

$$u_D^g(t, h, b_C, b_D \,|\, \theta, \xi) = (1 - \pi^g(b_C, b_D))W_D(\theta) + \pi^g(b_C, b_D)\left[V_D^g(t, h, b_C, b_D) - K_D(h, \xi)\right].$$

As a regularity condition to simplify the statement of Proposition C.5, we assume K_D is continuous in ξ.

We assume the type pair (θ, ξ) is drawn from a joint probability distribution with support on $\mathcal{S} \subseteq \Theta \times \Xi$. We do not impose any statistical relationship between them: θ and ξ may be positively correlated, negatively correlated, or uncorrelated.

The components of a direct mechanism—h, π, and V_D—have the same substantive interpretation as before, but are now functions of the type pair (θ, ξ). D's utility from reporting (θ', ξ') when its true type is (θ, ξ) is now

$$\Phi_D(\theta', \xi' \,|\, \theta, \xi) = (1 - \pi(\theta', \xi'))W_D(\theta) + \pi(\theta', \xi')\left[V_D(\theta', \xi') - K_D(h(\theta', \xi'), \xi)\right]. \qquad \text{(C.3)}$$

Mirroring the main analysis, we define $U_D(\theta, \xi) = \Phi_D(\theta, \xi \,|\, \theta, \xi)$. The incentive compatibility and voluntary agreements conditions for a direct mechanism have the same substantive interpretation as before, but we restate them here for the two-dimensional setting. Incentive compatibility is

$$U_D(\theta, \xi) \geq \Phi_D(\theta', \xi' \,|\, \theta, \xi) \qquad \text{for all } (\theta, \xi), (\theta', \xi') \in \mathcal{S}. \qquad \text{(IC}'\text{)}$$

[5] A similar mechanism drives results in Liu (2021).

[6] The exception, as in the main framework, being when $h = 0$: we assume $K_D(0, \xi) = 0$ for all $\xi \in \Xi$.

Voluntary agreements is

$$\pi(\theta,\xi)\left[V_D(\theta,\xi) - K_D(h(\theta,\xi),\xi)\right] \geq \pi(\theta,\xi)W_D(\theta) \qquad \text{for all } (\theta,\xi) \in \mathcal{S}. \tag{VA$'$}$$

We continue to restrict attention to direct mechanisms in which $\pi \in \{0,1\}$.

C.5.1 Increased War Payoff, Degraded Hassling Effectiveness

Proposition 1 in the main analysis shows that if hassling effectiveness decreases with war payoffs, then higher types are more likely to go to war rather than settling. We establish an analogous result for the two-dimensional setting, showing that the chance of war increases with D's war payoff as long as D's hassling effectiveness remains the same or decreases.

Proposition C.4. *Assume $(\theta',\xi'),(\theta'',\xi'') \in \mathcal{S}$. If $\theta'' > \theta'$ and $\xi'' \geq \xi'$, then $\pi(\theta'',\xi'') \leq \pi(\theta',\xi')$.*

Proof. For a proof by contradiction, suppose $\theta'' \geq \theta'$, $\xi'' \geq \xi$, $\pi(\theta',\xi') = 0$, and $\pi(\theta'',\xi'') = 1$. (IC$'$) implies

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

as well as

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

Combining these, along with the assumption that $\xi'' \geq \xi'$, gives us

$$\begin{aligned} W_D(\theta') &\geq V_D(\theta'',\xi'') - K_D(h(\theta'',\xi''),\xi') \\ &\geq V_D(\theta'',\xi'') - K_D(h(\theta'',\xi''),\xi'') \\ &\geq W_D(\theta''). \end{aligned}$$

But this in turn implies $\theta' \geq \theta''$, a contradiction. □

C.5.2 Increased War Payoff and Hassling Effectiveness

Proposition 2 in the main analysis shows that if hassling effectiveness increases with war payoffs, then the relationship between type and the probability of war depends on auxiliary conditions. Specifically, if the war payoff increases quickly enough relative to the decrease in the cost of hassling—a condition we call WURI—then war is more likely for higher types. In the opposite case, when the SURI condition holds because the hassling cost declines more quickly than the war payoff increases, we have the opposite relationship.

We now prove an analogous result for the two-dimensional setting. We consider the effects of increasing D's war payoff *and* hassling effectiveness on the equilibrium probability of war. If the change in hassling effectiveness is small enough, then the change cannot shift the equilibrium outcome from war to settlement.

Proposition C.5. *Assume $\pi(\theta', \xi') = 0$, and consider any $(\theta'', \xi'') \in \mathcal{S}$ such that $\theta'' > \theta'$. There exists $\hat{\xi} > \xi'$ such that if $\xi'' < \hat{\xi}$, then $\pi(\theta'', \xi'') = 0$.*

Proof. For a proof by contraposition, suppose $\pi(\theta'', \xi'') = 1$. Then (IC′) implies

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

as well as

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

Combining these yields

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

The RHS of the above expression is strictly positive. Meanwhile, the LHS approaches 0 as $\xi'' \to \xi'$. Therefore, there is a cutpoint $\hat{\xi} > \xi'$ such that $\xi'' > \hat{\xi}$. □

C.5.3 Equilibrium Payoffs

Our next result generalizes our results on the relationship between private type and equilibrium payoffs (Proposition 3 and Proposition 4) for the setting in which war payoffs and hassling costs are separate type components. We find that D's equilibrium utility is weakly increasing in the private war payoff and weakly decreasing in the cost of hassling. Additionally, the marginal decrease in payoff with the hassling cost is strict for any type that settles and hassles along the path of play.

Proposition C.6.

(a) For all ξ, if $\theta'' > \theta'$, then $U_D(\theta'', \xi) \geq U_D(\theta', \xi)$. The inequality is strict if $\pi(\theta', \xi) = 0$.

(b) For all θ, if $\xi'' > \xi'$, then $U_D(\theta, \xi'') \leq U_D(\theta, \xi')$. The inequality is strict if $\pi(\theta, \xi'') = 1$ and $h(\theta, \xi'') > 0$.

Proof. Claim (a). Fix any ξ, and assume $\theta'' > \theta'$. If $\pi(\theta', \xi) = 1$, then (IC′) implies

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

Otherwise, if $\pi(\theta', \xi) = 0$, then (IC′) implies

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

Claim (b). Fix any θ, and assume $\xi'' > \xi'$. If $\pi(\theta, \xi'') = 0$, then (IC′) implies

$$U_D(\theta, \xi') \geq W_D(\theta) = U_D(\theta, \xi'').$$

Otherwise, if $\pi(\theta, \xi'') = 1$, then (IC′) implies

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

If $h(\theta, \xi'') > 0$, then the second inequality in the above expression is strict. □

C.5.4 Binary Hassling Decision

We can obtain further characterization by assuming the choice of whether to hassle is dichotomous. Formally, assume the space of feasible hassling decisions is $\mathcal{H} = \{0, 1\}$. Without loss of generality, we may now allow ξ to represent the cost of $h = 1$, so that $K_D(h, \xi) = \xi h$. We also let $W_D(\theta) = \theta$ without loss of generality. Equation C.3 now reduces to

$$\Phi_D(\theta', \xi' \mid \theta, \xi) = (1 - \pi(\theta', \xi'))\theta + \pi(\theta', \xi') \left[V_D(\theta', \xi') - \xi h(\theta', \xi') \right].$$

As a helpful preliminary, we establish that an important implication of Proposition 5 still holds in the two-dimensional environment: if two types both settle in equilibrium and both have the same hassling choice, then they yield the same settlement value.

Lemma C.5. *For each $h \in \{0, 1\}$, there exists $V_h \in \mathbb{R}$ such that if $\pi(\theta, \xi) = 1$ and $h(\theta, \xi) = h$, then $V_D(\theta, \xi) = V_h$.*

Proof. Suppose $\pi(\theta', \xi') = \pi(\theta'', \xi'') = 1$ and $h(\theta', \xi') = h(\theta'', \xi'') = h$. (IC′) for (θ', ξ') implies

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

and thus $V_D(\theta', \xi') \geq V_D(\theta'', \xi'')$. By the same token, (IC′) for (θ'', ξ'') implies $V_D(\theta'', \xi'') \geq V_D(\theta', \xi')$. Therefore, $V_D(\theta', \xi') = V_D(\theta'', \xi'')$. □

To avoid trivialities in the remaining results, we will restrict attention to direct mechanisms in which all three outcomes—war, settlement with no hassling, and settlement with hassling—are realized by at least one type.

Among the types that settle in equilibrium, the choice of whether to hassle is determined entirely by one's cost of hassling. Naturally, the cost of hassling must be no greater than the marginal increase in the settlement due to hassling.

Lemma C.6. *Let V_0 and V_1 be defined as in Lemma C.5.*

- *If $\pi(\theta, \xi) = 1$ and $h(\theta, \xi) = 0$, then $\xi \geq V_1 - V_0$.*
- *If $\pi(\theta, \xi) = 1$ and $h(\theta, \xi) = 1$, then $\xi \leq V_1 - V_0$.*

Proof. To prove the first claim, suppose $\pi(\theta', \xi') = 1$ and $h(\theta', \xi') = 0$. (IC′), combined with Lemma C.5, implies $V_0 \geq V_1 - \xi'$, which proves the result. The proof of the second claim is analogous. □

Each type's expected utility from settling is thus the upper envelope of the constant V_0 and the linear function $V_1 - \xi$. Whether a type settles or fights depends on the comparison between its war payoff and this upper envelope.

Lemma C.7. *Let V_0 and V_1 be defined as in Lemma C.5.*

- *If $\pi(\theta, \xi) = 0$, then $\theta \geq \max\{V_0, V_1 - \xi\}$.*
- *If $\pi(\theta, \xi) = 1$, then $\theta \leq \max\{V_0, V_1 - \xi\}$.*

Proof. Immediate from Lemma C.5 and the incentive compatibility condition. □

These results allow us to fully characterize the set of incentive compatible direct mechanisms in the binary hassling case.

Proposition C.7. *Assume $\mathcal{S} \subseteq [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$. A direct mechanism is incentive compatible if and only if there exist $V_0 \in [\underline{\theta}, \overline{\theta}]$ and $V_1 \in [V_0 + \underline{\xi}, V_0 + \overline{\xi}]$ such that:*

(a) If $\pi(\theta, \xi) = 1$, then $V_D(\theta, \xi) = V_{h(\theta,\xi)}$.

(b) If $\pi(\theta, \xi) = 1$ and $\xi < V_1 - V_0$, then $h(\theta, \xi) = 1$.

(c) If $\pi(\theta, \xi) = 1$ and $\xi > V_1 - V_0$, then $h(\theta, \xi) = 0$.

(d) If $\theta > \max\{V_0, V_1 - \xi\}$, then $\pi(\theta, \xi) = 0$.

(e) If $\theta < \max\{V_0, V_1 - \xi\}$, then $\pi(\theta, \xi) = 1$.

Proof. Necessity. Suppose the direct mechanism satisfies (IC′). Lemma C.5 implies the existence of V_0 and V_1, and (IC′) implies $V_0 \in [\underline{\theta}, \overline{\theta}]$ and $V_1 \in [V_0 + \underline{\xi}, V_0 + \overline{\xi}]$.[7] Claim (a) then follows from Lemma C.5, claims (b) and (c) follow from Lemma C.6; and claims (d) and (e) follow from Lemma C.7.

Sufficiency. Suppose the direct mechanism satisfies the conditions of the proposition. Property (a) implies that each type's payoff from settlement without hassling is V_0, and each type's payoff from settlement with hassling is $V_1 - \xi$.

[7] A violation of these boundary conditions would imply that one or more outcomes are not reached by any type along the path of play, violating our restriction to mechanisms where all three possibilities occur. For example, $V_0 < \underline{\theta}$ would imply that every type strictly prefers war over settlement with no hassling, so there would be no type for which $\pi(\theta, \xi) = 1$ and $h(\theta, \xi) = 0$.

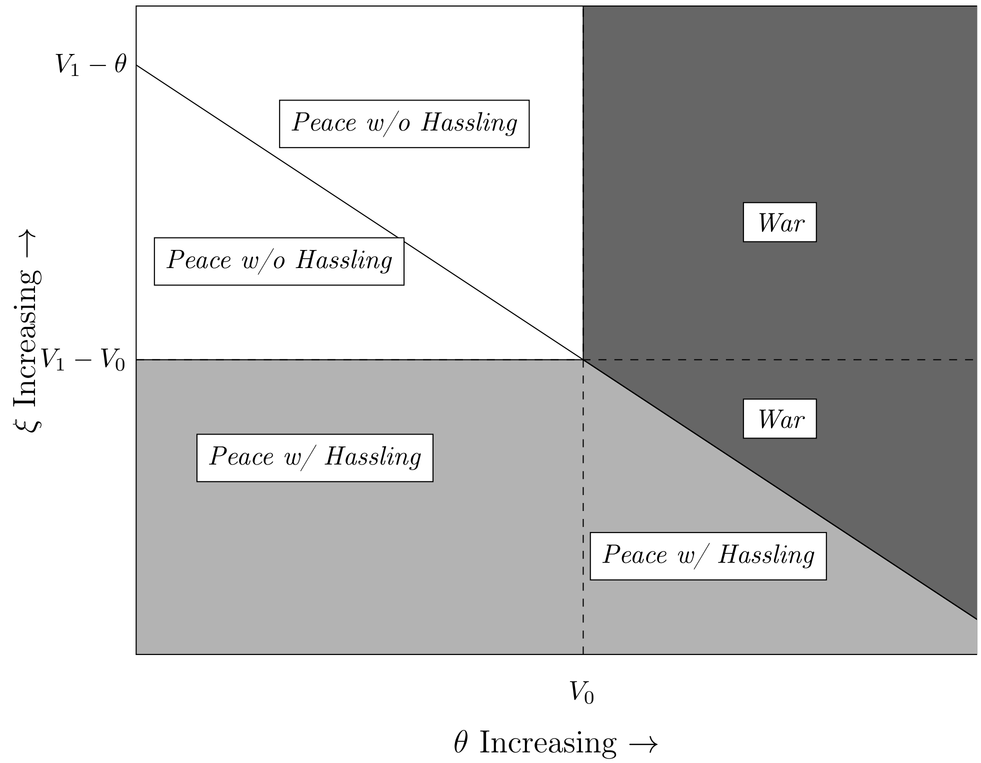

Figure 5: Equilibrium outcomes in the framework with multidimensional types and a binary hassling decision.

Now take any type (θ', ξ') that goes to war: $\pi(\theta', \xi') = 0$. Property (e) implies $\theta' \geq \max\{V_0, V_1 - \xi'\}$, so there is no incentive for (θ', ξ') to deviate to settlement with or without hassling.

Next take any type (θ', ξ') that settles and does not hassling: $\pi(\theta', \xi') = 1$ and $h(\theta', \xi') = 0$. Property (b) implies $V_0 \geq V_1 - \xi'$, so there is no incentive for this type to deviate to settlement with hassling. Property (d) in turn implies $V_0 \geq \theta'$, so there is also no incentive to deviate to war.

Finally, take any type (θ', ξ') that settles and hassles: $\pi(\theta', \xi') = 1$ and $h(\theta', \xi') = 1$. Properties (c) and (d) then imply $V_1 - \xi' \geq \max\{V_0, \theta'\}$, so there is no incentive for this type to deviate to settlement without hassling or to war. □

Figure 5 illustrates the shape of an incentive compatible direct mechanism in the binary hassling case. Naturally, war occurs when D's war payoff and hassling costs are both high (top-right of the figure). If D's war payoff is low (left side), then the equilibrium ends in settlement, with D hassling if its costs of doing so are low (bottom-left) and not hassling otherwise (top-left). The most interesting scenario is when D's war payoff is high and its hassling costs are low (bottom-right). In this case there may be war or settlement with hassling, depending on the exact magnitude of war payoffs compared to hassling costs.

We also see our general results on the two-dimensional framework, Proposition C.4 and

Proposition C.5, reflected in the figure. Consider any point in the figure where the equilibrium outcome is war. Per Proposition C.4, any increase in D's war payoff accompanied by no change or an increase in D's hassling costs (moving right or upper right) will still have war as the outcome. Meanwhile, per Proposition C.5, the effect of a simultaneous improvement in D's war payoff and reduction in its hassling costs (moving down and to the right) is conditional. If the reduction in hassling costs is greater than the increase in war payoff, then the type with the higher war payoff may settle instead of going to war.

C.6 Example Model with Signaling

As an additional example to illustrate how types with greater war payoffs may be less likely to fight in equilibrium, consider the following flexible response crisis bargaining game. The players are C and D, and the timing is as follows:

1. Nature selects D's type, $\theta \in \{\theta', \theta''\}$ from a nondegenerate prior distribution ($\theta' < \theta''$).
2. D observes θ and chooses Hassle or ¬Hassle.
3. C observes D's choice of hassling and chooses Challenge or ¬Challenge.
4. If C chooses to Challenge, D chooses Settle or Fight.

Hassling shifts the outcome in case of settlement by a fraction $\gamma \in (0, 1)$ in favor of D. Letting $h \in \{0, 1\}$ denote D's choice of whether to Hassle, payoffs from each possible outcome are given in Table 2.

Outcome	C's utility	D's utility
¬Challenge	0	$1 - K_D(h, \theta)$
Settle	$1 - \gamma h$	$\gamma h - K_D(h, \theta)$
Fight	W_C	$W_D(\theta)$

Table 2: Payoffs from the example model with signaling.

To illustrate the case of most substantive interest, where the probability of war might decrease with D's privately known war payoff, we assume θ improves hassling effectiveness. Specifically, we assume

$$\kappa' \equiv K_D(1, \theta') > K_D(1, \theta'') \equiv \kappa''.$$

In case of no hassling, neither type pays a cost: $K_D(0, \theta') = K_D(0, \theta'') = 0$.

Under appropriate parameter restrictions, this game has a perfect Bayesian equilibrium in which the low type of D fights and the high type settles.

Proposition C.8. *In the example model with signaling, assume $W_C > 0$, $W_D(\theta') > 0$, and*

$$\gamma - \kappa' < W_D(\theta') < W_D(\theta'') < \gamma - \kappa''.$$

The following assessment is a fully separating perfect Bayesian equilibrium:

- *D's initial choice: θ' chooses ¬Hassle, θ'' chooses Hassle.*
- *C's response: C chooses Challenge regardless of D's initial choice.*
- *D's final choice after C chooses Challenge:*
  - *If D previously chose ¬Hassle, both types of D choose Fight.*
  - *If D previously chose Hassle, θ' chooses Fight and θ'' chooses Settle.*
- *C's beliefs: C believes $\theta = \theta'$ with probability 1 after observing ¬Hassle and believes $\theta = \theta''$ with probability 1 after observing Hassle.*

Proof. To confirm sequential rationality, we work backward. The proposed strategies for D's final choice are best responses by construction. Moving up, it is optimal for C to Challenge after D chooses Hassle, as $1 - \gamma > 0$. It is also optimal for C to Challenge after D chooses ¬Hassle, as $W_C > 0$. Now consider D's initial hassling decision. Deviating to ¬Hassle would yield a payoff of $W_D(\theta'')$ for θ'', strictly less than the payoff of $\gamma - \kappa''$ from Hassle. Meanwhile, for θ', deviating to Hassle would yield a payoff of $W_D(\theta')$, the same as this type's equilibrium payoff. Finally, to confirm that this is an equilibrium, we have that C's beliefs are consistent with Bayes' rule whenever possible. □

C.7 Application to Endogenous Power Shift Models

Our general framework encompasses at least some games in which states may take costly actions to increase their future power, and where states may start a preemptive war to halt such shifts. To make the connection between these games and our mechanism design framework clearer, here we illustrate one such model and how its equilibrium can be translated into a direct mechanism in our terms.

Consider a game between C and D, bargaining over a prize whose value is normalized to 1, with the following timing:

1. Nature assigns D's type, $\theta \in \Theta$.
2. C decides whether to start a preemptive war. Doing so ends the game; otherwise it continues to the next step.
3. D decides whether to start a preemptive war. Doing so ends the game; otherwise it continues to the next step.
4. C updates its beliefs about D's type and selects transgressions $t \geq 0$. Simultaneously, D selects hassling $h \geq 0$.
5. D makes an offer $x \in \mathbb{R}$.

6. C accepts or rejects the offer.[8] If C accepts, then the prize is divided on the basis of the offer. If C rejects, war occurs. Either way, the game ends.

Transgressions and hassling shift the balance of power in case of war. Specifically, let $p : \mathbb{R}^2_+ \to [0,1]$ be a function that maps from transgression and hassling choices into C's probability of victory. We assume D's type affects only its cost of fighting (as well as hassling, if the game reaches that stage), but not the balance of power. In case of a preemptive war, payoffs are

$$W^0_C = p(0,0) - c_C,$$
$$W^0_D(\theta) = 1 - p(0,0) - c_D(\theta),$$

where $c_D : \Theta \to \mathbb{R}_{++}$ is a strictly decreasing function denoting D's baseline war cost. If instead war occurs in the final period, payoffs are

$$W^1_C(t,h) = p(t,h) - c_C - K_C(t),$$
$$W^1_D(t,h,\theta) = 1 - p(t,h) - c_D(\theta) - K_D(h,\theta).$$

Finally, if a settlement is reached in the final period, C's payoff is $x - K_C(t)$ and D's is $1 - x - K_D(h,\theta)$.

A pure strategy equilibrium of this game is an assessment consisting of:

- C's initial decision of whether to start a preemptive war: $i^*_C \in \{0,1\}$
- Each type of D's decision of whether to start a preemptive war: $i^*_D : \Theta \to \{0,1\}$
- C's updated belief about D's type conditional on D not starting a preemptive war, a probability measure μ^* whose support is a subset of Θ
- C's choice of transgressions in case of no preemptive war, $t^* \geq 0$
- Each type of D's hassling in case of no preemptive war, $h^* : \Theta \to \mathbb{R}_+$
- Each type of D's offer, depending on the choices of transgressions and hassling, $x^* : \mathbb{R}^2_+ \times \Theta \to \mathbb{R}$
- C's decision whether to accept an offer, depending on the choices of transgressions and hassling, $a^* : \mathbb{R}^2_+ \times \mathbb{R} \to \{0,1\}$

Because war payoffs in the final stage depend on transgression and hassling, it is not immediately obvious that our framework subsumes this model. However, in any equilibrium, the final stage will result in a negotiated settlement if reached. We prove this now, and then show below that it allows us to translate any equilibrium of the game into an incentive compatible direct mechanism in our framework.

[8] Technically, C would update its beliefs again here. However, because D's type only affects its cost of fighting, C's beliefs are payoff-irrelevant at this point. Hence we omit them for simplicity.

Lemma C.8. *In any equilibrium, following any choices of transgressions and hassling, all types of D offer $x = p(t,h) - c_C$, and the offer is accepted.*

Proof. The proof follows the standard logic of the ultimatum game. It is a best response for C to accept the offer if and only if $x - K_C(t) \geq p(t,h) - c_C - K_C(t)$, so C's equilibrium acceptance rule is

$$a^*(t,h,x) = \begin{cases} 1 & \text{if } x \geq p(t,h) - c_C, \\ 0 & \text{otherwise.} \end{cases}$$

As usual, C must accept when indifferent to avoid an open set problem for D. Because $c_D(\theta) > 0$ for all $\theta \in \Theta$, it follows immediately that $x^*(t,h,\theta) = p(t,h) - c_C$. □

This result allows us to characterize the equilibrium payoffs in case of no preemptive war as a function of the equilibrium transgression and hassling decisions. This characterization in turn allows us to express the relevant equilibrium quantities in terms of a direct mechanism in our framework. The only war that is possible in equilibrium is a preemptive war, so we may simply define the function $W_D \equiv W_D^0$. For the three functions defining a direct mechanism, we have:

- The hassling level is simply the extent of hassling a type employs if there is no preemptive war:
$$h(\theta) = h^*(\theta).$$
- The indicator for whether war occurs is a function of the preemption decisions:
$$\pi(\theta) = (1 - i_C^*)(1 - i_D^*(\theta)).$$
- The settlement value is the equilibrium bargaining outcome in case of no preemptive war:
$$\begin{aligned} V_D(\theta) &= 1 - x^*(t^*, h^*(\theta)) \\ &= 1 - p(t^*, h^*(\theta)) + c_C. \end{aligned}$$

Suppose type θ were to deviate to the equilibrium strategy of type θ'. By Lemma C.8, the deviation would yield a utility of

$$\begin{aligned} &\underbrace{[i_C^* + (1 - i_C^*) i_D^*(\theta')] W_D^0(\theta)}_{\text{war}} + \underbrace{(1 - i_C^*)(1 - i_D^*(\theta'))[1 - x^*(t^*, h^*(\theta')) - K_D(h^*(\theta'), \theta)]}_{\text{settlement}} \\ &\quad = [1 - \pi(\theta')] W_D(\theta) + \pi(\theta')[V_D(\theta') - K_D(h(\theta'), \theta)] \\ &\quad = \Phi_D(\theta' \mid \theta). \end{aligned}$$

Consequently, the direct mechanism that corresponds to an equilibrium must satisfy our incentive compatibility condition, (IC). Additionally, sequential rationality implies that type θ must choose preemptive war ($i_D^*(\theta) = 1$) if the equilibrium bargaining outcome is worse than its preemptive war payoff:

$$\begin{aligned} &V_D(\theta) - K_D(h(\theta), \theta) < W_D(\theta) \\ &\qquad \Rightarrow 1 - x^*(t^*, h^*(\theta)) - K_D(h^*(\theta), \theta) < W_D^0(\theta) \\ &\qquad \Rightarrow i_D^*(\theta) = 1 \\ &\qquad \Rightarrow \pi(\theta) = 0. \end{aligned}$$

In other words, the corresponding direct mechanism must satisfy our voluntary agreements condition, (VA). Because our results apply to all IC and VA direct mechanisms taking the specified form, our analysis covers any equilibrium of the model given here.

The model we have laid here has some special features. Some of these are important for our framework to apply, while others are not. For example, there is no bargaining in the first stage of the model, before the opportunity for preemptive war. Allowing for such bargaining would complicate the form of the equilibrium and therefore make it harder to see how the equilibrium maps into a direct mechanism, but the resulting direct mechanism would still meet our conditions. Our assumptions about the second stage are more consequential. As is clear from the analysis above, the straightforward application of our framework depends on the second stage always ending with a settlement. That would not necessarily be the case if, for example, C made the offer in the second period or if θ affected the distribution of power rather than D's cost of war. A more complex framework would be required to analyze games where hassling shifts the long-run distribution of power *and* war occurs on the equilibrium path even after the power shift. That said, war typically only occurs in the first period in shifting-power models, whether the shift occurs exogenously (e.g., Powell 2006) or endogenously (e.g., Spaniel 2019).