Uncertainty in Crisis Bargaining with Multiple Policy Levers

Brenton Kenkel and Peter Schram

December 1, 2020

Abstract

We examine the relationship between private capabilities and war in a new class of crisis bargaining games. While crisis bargaining games typically only considers outcomes of war or a peaceful bargain, we assume that actors can also engage in a wide range of low-level, costly policy options that shape final political outcomes in their favor (representing actions like sanctions, arming, cyberattacks, etc.). This modeling modification undermines seminal theoretical results that previously identify a positive relationship between a state's private capabilities for war and both that state's final expected utility and the overall likelihood of war in the game; we show that these relationships no longer hold generally when a state's private capabilities affect both war efficacy and low-level operation efficacy. Using the tools of Bayesian mechanism design, we establish the general conditions that define the relationships between private abilities, war, and expected utilities for what we call "flexible-response crisis bargaining games."

"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.

1 Introduction

The relationship between private information and war is central to contemporary international relations theory (Fearon, 1995; Powell, 1999; Schultz, 1999; Wagner, 2000; Slantchev, 2003; Meirowitz *et al.*, 2008; Chassang and Padró i Miquel, 2009; Acharya *et al.*, 2015; Spaniel and Bils, 2018). Theories of how private information affects war are typically developed using crisis bargaining games, in which states engage in a series of negotiations over a divisible asset and each state may go to war as an outside option. These models assume one or more states have a private "type," unknown to their adversaries, representing their hidden capabilities or willingness to go to war. Despite the variety of models of how this bargaining might take place, two results consistently emerge. First, when an actor possesses a greater private willingness to go to war, negotiations are more likely to end in war. Second, when an actor possesses a greater private willingness to go to war, that actor will attain a greater final expected utility. As research like Banks (1990) and Fey and Ramsay (2011) illustrate, these two results are not simply theoretical anecdotes being driven by modeling choices, but rather these results are borne out within game-form free analysis of crisis bargaining. What is powerful about the Banks and Fey and Ramsay results are that *any* game falling within the crisis bargaining framework with information asymmetry will share this relationship between private willingness and outcomes.

Recent trends in international relations research highlight an implausible assumption within the crisis bargaining framework: that actors either reach a peaceful and efficient bargain, or actors engage in some form of costly and destructive engagement, typically war. Consistent with a broad set of empirical and formal research, in practice, actors face a vast array of potential policy levers at their disposal that would not be considered a peaceful resolution or a decisive war. After all, actors can undertake any of the following actions: implement sanctions or tariffs (Coe, 2014; McCormack and Pascoe, 2017; Spaniel and Malone, 2019; Joseph, 2020); offer third-party support to rebels (Schultz, 2010); pursue a wide range of low-level operations sometimes classified as gray zone conflict (Mazarr, 2015), hybrid conflict (Lanoszka, 2016), or hassling (Schram, Forthcomming, 2020); engage in cyberwarfare (Gartzke and Lindsay, 2015; Baliga *et al.*, 2020); enter into brinkmanship (Powell, 2015); pursue arming, undertake an arms race, or pursue other strategic investments (Schultz, 2010; Gurantz and Hirsch, 2017; Coe and Vaynman, 2020); or select some combination of any of these. This wide range of possible policy levers not only suggests a more complex action space, but it also suggests that an actor's private willingness to engage in one type of policy could be linked to other policies as well. As one example, if an actor has a wide range of privately-known cyber-exploits, that actor may be both more willing to fight a cyberwar and more willing to fight a conventional war where they would use the cyberattacks at a grander scale.[1] To summarize, consistent with the wide range of observed international crisis behaviors, contemporary international relations research has begun embracing that actors face a wide range of possible policy choices within a crisis. But, previous attempts to draw general conclusions about crisis bargaining models have implicitly ruled out this diverse array of policy choices, meaning their results may not apply to many real-world crisis situations.

This paper re-examines the relationship between private information and war within what we call "flexible-response crisis bargaining games." The flexible-response crisis bargaining games share the key features of crisis bargaining games—one actor has a private type, and actors negotiate or can escalate to war—but the flexible-response games also allow multiple forms of costly conflict, and also allow the private type to influence the costs of these other forms of conflict. We find that previously established results for crisis bargaining models, like the relationship between a greater private willingness to go to war and a greater final likelihood of war, break down in the flexible-response model. Using a game-free analysis along the lines of previous mechanism design research (Banks, 1990; Fey and Ramsay, 2011), we show that these differences with conventional findings are not the result of any particular game form or functional form. Instead, they are rather general properties of all flexible-response crisis bargaining games.

Flexible-response crisis bargaining games consider two competing players, a challenger and a defender, in a crisis. In the games, the challenger begins by selecting some "transgression" from a continuum of possible transgressions, where the transgression is politically advantageous to the challenger and detrimental to the defender. In response, the defender either can enter into a decisive war over the transgression, can allow the transgression to come to fruition, or can "hassle" and undercut the transgression through some costly, non-decisive, low-level response. Critically, the defender possesses a private type, which represents their hidden ability or willingness to go to war and engage in hassling. Thus, the key strategic tension is that the challenger wants to undertake aggressive transgressions, but is uncertain over how aggressive a policy they can pursue without being met with war or other forms of conflict. To offer an

[1] As a second example, if an actor is privately concerned about losing popular domestic support, the leader may be more willing to fight a war to create a rally-around-the-flag, but may be less willing to implement tariffs.

example of this dynamic, throughout the 2000s and 2010s, Iran engaged in a series of transgressions, including supporting third-party militants and terrorist groups throughout the Middle East and building a nuclear bomb. In response, the U.S. and its allies have, at various points, negotiated with Iran, issued statements against Iran, sanctioned Iran, supported limited strikes against Iranian leadership, and supported cyberattacks against Iran's nuclear program. While generalized analysis of crisis bargaining models do consider Iran's transgression decisions, they cannot describe the broad set of U.S. actions or how Iran internalized the possibility of a multifaceted U.S. response beyond a dichotomous war-or-peace treatment.

Within the flexible response crisis bargaining framework, we find that a defender's increased private willingness to go to war—an increase in their private type—could result in more or less war and a greater or lower expected utility for the defender. While these results may sound indefinite, in fact, we have precisely identified the conditions under which these outcomes occur. The results are summarized in Table 1 that considers, following an improvement in private war capabilities, whether the hassling response is more expensive (middle column) or less expensive (right column). If an improvement in private wartime capabilities also makes hassling more expensive, then these increases in private type must make war more likely. This finding is consistent with Banks (1990) and Fey and Ramsay (2011). However, if improvements in private war capabilities also makes hassling less expensive, then increases in private type could increase, decrease, or non-monotonically alter the likelihood of war. In the latter case, this ambiguity exists because increases in type improve both war and hassling outcomes, but the defender can only select one policy lever, hassling or war. We can assign some structure to this interaction with further assumptions; if increases in type create larger marginal benefits on the costs of hassling relative to the effect on the war payoff, then we have the opposite of the usual result, with stronger types less likely to fight on the path of play.

	Hassling is less effective	Hassling is more effective
EU_D	U-shaped (Lemma 9)	Increasing (Lemma 4)
Probability of war	Increasing (Lemma 8)	Unclear (Lemma 6)

Table 1: Summarized results for when when D's private willingness to go to war increases. The column labeled "Hassling is less effective" implies that increases in D's private war capabilities makes low-level conflict more expensive. When "Hassling is more effective," increases in D's private war capabilities makes low-level conflict less expensive.

Also consistent with Banks (1990) and Fey and Ramsay (2011), when improvements in private war capabilities also makes hassling less expensive, then increases in private type will result in the defender attaining a greater final utility. But, if improvements in private war capabilities also makes hassling more expensive, then the relationship between type and utility can exhibit a U-shape, where the defender's utility is decreasing for the range of types where D hassles, and increasing for the range of types where D goes to war. This is a noteworthy contrast with ordinary crisis bargaining models, where types that expect to receive more from war also receive more through bargaining.

We present generalized results on the relationship between private information and conflict outcomes for a new class of crisis bargaining models. That actors have many possible policy choices outside of war and peace is not simply a technical flourish; rather, it is an adaptation of the crisis bargaining framework that is more consistent with contemporary research and a wide range of substantive topics. As a result, this paper makes several important contributions. First, while this is not the first paper to consider actors with multiple possible actions within a crisis (Schultz, 2010; Powell, 2015; McCormack and Pascoe, 2017; Spaniel and Malone, 2019; Schram, Forthcomming; Baliga *et al.*, 2020), this is the first paper to systematically examine how the spillover effects of improvements in one kind of conflict capability can affect other conflict capabilities as well. This allows us to offer novel insights into a wide-range of previously considered substantive topics. Second, this paper illustrates, for a broad class of games, the conditions under which a greater private willingness to go to war results in less war or lower utilities for the state with the greater willingness. These findings are, to the best of our knowledge, novel. Third, by using the tools of mechanism design, this paper generates results for any possible game that falls within our flexible-response crisis bargaining framework. Whereas typically a cluster of models illustrating the same comparative static can suggest a finding is not an artifact of a modeling choice, our results bypass the need for iterating over these kinds of models, and instead establishes general results outright.

Our paper is most similar to formal models where policymakers face multiple possible actions within a crisis (see citations above). Our paper is also similar to a class of work on political science topics that embraces the tools of mechanism design to establish relationships between private information and outcomes. Outside of sources listed above, mechanism design has been featured in work on crisis bargaining and arbitration (Fey and Ramsay, 2009; Hörner *et al.*, 2015; Fey and Kenkel, 2019), bureaucracies and delegation (Ashworth and Sasso, 2019), firm regulation (Baron and Besanko, 1987, 1992), legislation and policy-making (Meirowitz *et al.*, 2006; Meirowitz, 2007), voting (Aghion and Jackson, 2016), and many others.

2 What do Flexible Response Crisis Bargaining Models Describe?

Flexible Response Crisis Bargaining models formalize the following interaction. First, a challenger—State C—undertakes some opportunistic and costly action that we will refer to as a "transgression." Transgressions are politically beneficial to State C but are detrimental to State D. Moves like this have been proposed in the literatures considering enforcement problems in bargaining, deterrence, or endogenous power shifts (Schultz, 2010; Debs and Monteiro, 2014; Gurantz and Hirsch, 2017). In response, a defender—State D—has the choice between accepting the transgression, going to war to decisively resolve the political issues between the two states, or in engaging in some low-level actions that could undercut the future impact of the transgressions. The first two options (acceptance and war) are standard in the crisis bargaining framework; the last option is less well examined. We will refer to D's low-level response as "hassling." As originally defined in Schram (Forthcomming), hassling is the use of limited conflict to degrade a challenger's rise; our use of the term here is consistent with this definition, but also refers to any actions by the defender that undercuts the challenger's transgression. So long that D did not previously go to war, C and D then engage in a bargaining protocol that can end with a diplomatic resolution or war. The key strategic tension explored in the model is that State C wants to aggressively pursue transgressions that will help their future bargaining position, but because D has private capabilities, C does not know how far they can go in their transgressions before D responds with hassling or war.

The general terms above—"transgressions" and "hassling"—can apply to a broad range of observed international political behavior. Transgressions could be investing in conventional, nuclear, space, or cyber military technologies (Debs and Monteiro, 2014; Gartzke and Lindsay, 2017; Spaniel, 2019), forming alliances (Benson and Smith, 2020), securing geopolitically valuable territory (Fearon, 1996; Powell, 2006), or engaging in economic warfare to weaken a target. Essentially, transgressions are any kind of action that shifts future politics in State C's favor (Schultz, 2010; Gurantz and Hirsch, 2017). In response to the challenger's transgression, the defender can hassle to undermine the transgression, which could take the form of limited airstrikes, special operations, cyberattacks, supporting domestic (to State C) insurgent groups, or other military activities. Of course, while the Schram (Forthcomming) definition of hassling is limited to military activity, here it can apply to anything that can undermine the future political impact of C's transgressions, for example sanctions (McCormack and Pascoe, 2017).

Critical to our theory below, we claim that D's private type influences both war and hassling capabilities. In other words, if state D is (privately) very capable at conducting war or very willing to conduct war, we might also expect that state D is systematically better (or worse) at conducting hassling. To offer perhaps the simplest example, if D is a hawk, then D's hawkishness could make D more willing to conduct war or conduct limited military operations (i.e. hassle) in response to C's transgression. In the interest of providing more specifics, we offer a short list of substantive and theoretical vignettes where states are choosing from multiple policy levers, and where certain technologies or underlying conditions can influence both the efficacy of the war and hassling options. To be clear, by no means is this list comprehensive, nor does it definitively define the relationship between war and the various hassling forms.[2] Rather, the list here illustrates that there are natural linkages across war and hassling capabilities.

Electronic Warfare: In 2007, Israel discovered that Syria was building a nuclear reactor in Al Kibar. In response, in Operation Outside the Box, Israel used electronic warfare to disable Syrian air defenses and conducted an airstrike on the reactor (Trevithick, 2016). While the full details of the electronic warfare attack are not disclosed, an attack that allowed multiple Israeli aircraft to enter Syria and conduct a raid without harassment plausibly could have been used to conduct a more extensive attack as well. In this case, Israel's competence in electronic warfare likely would have made both a limited operation and a more expansive operation (or war) less costly.

Economic Warfare: An influential literature suggests that an integrated global economy can be a driver for peace because it raises the costs of war and other activities that disrupt trade (Gartzke *et al.*, 2001; Gartzke, 2007). Additionally, an integrated global economy also presents opportunities for more robust implementations of sanctions or tariffs for bad behavior. Together, a state that is highly economically integrated (or expecting future additional economic integration) may have high costs to conducting war, but lower costs of implementing sanctions or tariffs.

Precision Strike Capabilities: In 1998, in Operation Desert Fox, the United States used precision strike capabilities to attack Saddam Hussein in response to his failure to comply with international agreements designed to inhibit Iraq's development of weapons of mass destruction. Precision strike technologies were used again in 2003 in the invasion of Iraq. On one hand, the precision strike technologies were useful in both conducting a limited airstrike, and in facilitating the dismantling of Saddam's armed forces. On the other hand, investments in precision strike technologies (and similar weapons covered in the "revolution in military affairs") may have come at the expense of the investments needed to conduct effective counterterror, counterinsurgency, and major civil-military efforts (Cordesman, 2018). While precision strike capabilities certainly led to Operation Desert Fox being cheaper and more effective, their ultimate

[2]For example, below we describe how being economically integrated can make sanctions a more effective tool and war more costly. Historically, ships used for trade were retrofitted for conflict, suggesting, in other cases, being economically integrated may assist in the conduct of war.

impact on the total costs of war with Iraq could have been positive or negative.

Militants on Retainer: Since the 1990s, Russia maintained connections with and supported militants within GeorgiC's South Ossetia and Abkhazia regions (Asmus, 2010). In the 1990s and early 2000, these militants both conducted periodic limited operations against the Georgian government. Then, in 2008 these militants fought alongside Russia in the Russian-Georgian war, and were responsible for shaping conflict outcomes by conducting ethnic cleansing of Georgians living in the regions. Together, RussiC's access to militants in Georgia made both low-level conflict and war cheaper.

Conventional Forces and Brinkmanship: As Schelling (1966), Jervis (1979), and Powell (2015) describe, the nuclear revolution established a new role for conventional forces: to display resolve and to raise the risk of a nuclear confrontation. While this is distinct from the traditional role of conventional forces, we conjecture that maintaining a conventional force that is capable of working effectively within brinkmanship would undoubtedly be useful in other conflict settings where states do not enter into brinkmanship. For example, Russian conventional forces were useful in its conventional and low-level engagements in Georgia and Ukraine. And, these conventional forces would also be useful to generate risk should a new conflict arise between Russia and NATO that justified an escalation to nuclear brinkmanship.

3 Example Model

Here we include an example model demonstrating one instance of how better private wartime capabilities can lead to lower utilities and less war. The game form is loosely based on Fearon (1997) and Gurantz and Hirsch (2017), with a few key modifications. In the model below, both the challenger's transgression and the defender's response are selected from continua of possible choices. Additionally, we treat the defender as having a private type that influences their conflict payoffs. Overall, this model captures the case where a challenger wants to transgress, but does not know how far they can go without triggering hassling or war from the defender. This interaction is similar to that examined in Schram (2020), but here private type influences both hassling and war payoffs.

3.1 The Model and Assumptions

Two states, a challenger (C) and a defender (D) are in a crisis. The following model is one example of a flexible-response crisis bargaining game. The game form is as follows.

1. Define $\{\underline{\theta}, \bar{\theta}\} = \Theta$ with $\underline{\theta} < \bar{\theta}$. Nature sets $\theta = \underline{\theta}$ with probability $Pr(\underline{\theta})$ and sets $\theta = \bar{\theta}$ with probability $1 - Pr(\underline{\theta})$. D knows nature's selection θ, but C does not.
2. State C selects transgression $t \in \mathcal{T} = \mathbb{R}_{\geq 0}$.

3. State D can either go to war by setting $w_D = 1$ or not go to war by setting $w_D = 0$ and countering the transgression by selecting $h \in \mathcal{H} = \mathbb{R}_{\geq 0}$ (with $h = 0$ implying that D "accepts" the transgression). When D goes to war, States C and D recieve wartime payoffs $U_C = W_C$ and $U_D = W_D(\theta)$ (respectively). When D does not go to war, C and D recieve $U_C = X_C + t - h$ and $U_D = X_D - t + h - \frac{h^2}{F(\theta)}$.

We assume that $W_C < X_C$ and $W_D < X_D$, thus implying that an entirely peaceful outcome ($t = 0$ and $h = 0$) is preferred to war, but it is possible for C to select so great a t that D prefers going to war. For simplicity, we assume that t and h shift the political outcome linearly, that C faces no costs to transgressing, and that D incurs a simple strictly increasing cost term; the general analysis below makes none of these simplifying assumptions. Finally, while we assume that $W_D(\theta)$ is increasing in θ, we examine cases where $F(\theta)$ is increasing or decreasing in θ below. As described above, whether improved war capabilities (higher θ) also improves or is detrimental to hassling capabilities is an empirical question that is likely case-specific.

3.2 Equilibrium

We limit attention to pure strategy perfect Bayesian Nash equilibria. We use an asteriks to denote equilibrium actions.

In the third stage, D reacts to C's selected transgression level (t^*) by either going to war (setting $w_D^* = 1$) or not going to war (setting $w_D^* = 0$) and selecting an optimal response h^*. In the second stage, C selects a level of t^* based on how C expects D will respond, which is conditional on C's expectations over D's hidden type (θ). Because C faces no costs to the transgressions, C wants to be aggressive here, but knows that if C selects too great a t, then D will respond with war. Thus, C's utility from t is increasing, unless it provokes D to war. D's response is determined by D's hassling capabilities. For brevity, we will use the notation $D(\theta)$ to refer to a D that has type $\theta \in \{\underline{\theta}, \bar{\theta}\}$. Because C knows D is either type $\underline{\theta}$ or $\bar{\theta}$, C will select either a t that would make a $D(\underline{\theta})$ indifferent between going to war (or not), or select a t that would make a $D(\bar{\theta})$ indifferent between going to war or not. We let $t(\theta)$ denote the transgression level that would make a $D(\theta)$ indifferent between war and hassling.

The previous paragraph suggests that C will either select $t(\underline{\theta})$ or $t(\bar{\theta})$. C's decision is driven by two key factors. First, C must decide whether C wants to avoid war altogether, or if C is willing to risk war with some types in order to select a more aggressive transgression t^*. Second, it will matter which type of D will tolerate greater levels of transgressions, or if $t(\underline{\theta}) \leq t(\bar{\theta})$ or $t(\underline{\theta}) > t(\bar{\theta})$. This matters because if C selects $t^* = t(\underline{\theta})$ and $t(\underline{\theta}) \leq t(\bar{\theta})$, then D will never go to war, but if C selects $t^* = t(\underline{\theta})$ and $t(\underline{\theta}) > t(\bar{\theta})$, then $D(\bar{\theta})$ will go to war. We discuss this formally below, but here is some intuition. Because D's war payoffs are decreasing in θ, improvements in θ make war cheaper for higher types of D. However, θ could also influence the costs of D's hassling. For example, θ could represent D's hidden cyber-capabilities, which could be used both in a war or to counter C's transgression at a level of conflict that is short of war; here moving from $\underline{\theta}$ to $\bar{\theta}$ makes both war and D's hassling cheaper. This could mean that improvements in θ also make D more willing to challenge C's transgression through means outside of a conventional war rather than by going to war.

Thus, the equilibrium takes on one of four cases. When $t(\underline{\theta}) \leq t(\bar{\theta})$, C will sometimes optimally avoid war (Case 1), and other times optimally go to war (Case 2). When $t(\underline{\theta}) > t(\bar{\theta})$, C will sometimes optimally avoid war (Case 3), and other times optimally go to war (Case 4). The $\hat{Q}$ and $\tilde{Q}$ conditions below arbirate between C's desision to risk war or to avoid war altogether. The conditions differ across Cases 1 and 2 and Cases 3 and 4 due to whether $t(\underline{\theta}) \leq t(\bar{\theta})$ or $t(\underline{\theta}) > t(\bar{\theta})$. We offer an intuition for the equilibrium in the appendix.

Proposition 1. *The following is the Perfect Pure-Strategy Bayesian Nash Equilibrium. Letting* $\hat{Q} = Pr(\underline{\theta})\left(X_C + X_D - W_C - \frac{F(\underline{\theta})}{4}\right) - W_D(\underline{\theta}) - (1 - Pr\underline{\theta}))\left(\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} - W_D(\bar{\theta})\right)$ *and* $\tilde{Q} = (1 - Pr(\underline{\theta}))\left(X_A + X_D - W_A - \frac{F(\bar{\theta})}{4}\right) - W_D(\bar{\theta}) + Pr(\underline{\theta})\left(\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} + W_D(\underline{\theta})\right)$*, there are four cases:*

- ***Case 1.*** *When* $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} \geq W_D(\bar{\theta}) - W_D(\underline{\theta})$ *and* $\hat{Q} \geq 0$ *holds, C selects transgression level* $t^* = t(\underline{\theta})$*, which results both types of D choosing to not go to war, setting* $h^* = \frac{F(\theta)}{2}$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*.* $D(\underline{\theta})$ *attains utility* $U_D(\sigma^*(\underline{\theta})) = W_D(\underline{\theta})$*, and* $D(\bar{\theta})$ *attains utility* $U_D(\sigma^*(\bar{\theta})) = W_D(\underline{\theta}) - \frac{F(\theta)}{4} + \frac{F(\theta)}{4}$*.*
- ***Case 2.*** *When* $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} \geq W_D(\bar{\theta}) - W_D(\underline{\theta})$ *and* $\hat{Q} < 0$ *holds, C selects transgression level* $t^* = t(\bar{\theta})$*, which results in* $D(\underline{\theta})$ *declaring war and* $D(\bar{\theta})$ *not going to war and setting* $h^* = \frac{F(\bar{\theta})}{2}$*. Both* $D(\underline{\theta})$ *and* $D(\bar{\theta})$ *attain their wartime utility, or* $U_D(\sigma^*(\theta)) = W_D(\theta)$ *for* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*.*
- ***Case 3.*** *When* $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} < W_D(\bar{\theta}) - W_D(\underline{\theta})$ *and when* $\tilde{Q} \geq 0$ *holds, C selects transgression level* $t^* = t(\bar{\theta})$*, which results both types of D choosing to not go to war, setting* $h^* = \frac{F(\theta)}{2}$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*.* $D(\bar{\theta})$ *attains utility* $U_D(\sigma^*(\bar{\theta})) = W_D(\bar{\theta})$*, and* $D(\bar{\theta})$ *attains utility* $U_D(\sigma^*(\underline{\theta})) = W_D(\bar{\theta}) - \frac{F(\bar{\theta})}{4} + \frac{F(\underline{\theta})}{4}$*.*
- ***Case 4.*** *When* $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} < W_D(\bar{\theta}) - W_D(\underline{\theta})$ *and when* $\tilde{Q} < 0$ *holds, C selects transgression level* $t^* = t(\underline{\theta})$*, which results in* $D(\bar{\theta})$ *declaring war and* $D(\underline{\theta})$ *not going to war and setting* $h^* = \frac{F(\underline{\theta})}{2}$*. Both* $D(\underline{\theta})$ *and* $D(\bar{\theta})$ *attain their wartime utility, or* $U_D(\sigma^*(\theta)) = W_D(\theta)$ *for all* $\theta \in \{\underline{\theta}, \bar{\theta}\}$*.*

Across all cases, $D(\theta)$*'s best response function is to hassle if* $t^* \leq t(\theta)$ *and to go to war otherwise. C's beliefs follow from the initial probability distribution of types.*

While Proposition 1 is dense, we summarize the key findings in the following Corollaries.

Corollary 1. *If D's private wartime payoffs increase (shifting $\underline{\theta}$ to $\bar{\theta}$), then the likelihood of war can be decreasing (Case 2), increasing (Case 4), or unchanging (Case 1 or 3).*

Corollary 2. *If D's private wartime payoffs increase (shifting $\underline{\theta}$ to $\bar{\theta}$), then D's final utility can be decreasing (Case 3), increasing (Case 1) or unchanging (Cases 2 and 4).*

Most notably, we have identified cases where private increases in D's wartime payoffs result in less war (in Case 2) and lower utility for D (in Case 3). Within the standard crisis bargaining framework explored in Banks (1990) and Fey and Ramsay (2011), these results would never arise. We speak to each of these results below, but in short, we arrive at the results we do because now D's private type θ influences both the payoffs from war and from hassling.

How can private increases in D's wartime payoffs make war occur less? Within Case 2, the shift in D's private type (from $\underline{\theta}$ to $\bar{\theta}$) makes both war and hassling better outcomes for D. More specifically, while $D(\bar{\theta})$ is better at war than $D(\underline{\theta})$, the condition $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} \geq W_D(\bar{\theta}) - W_D(\underline{\theta})$ implies that $D(\bar{\theta})$ is much better at hassling than $D(\underline{\theta})$, to an extent that $D(\bar{\theta})$'s improvements in hassling overshadows $D(\bar{\theta})$'s gains from war. This means that C can select more aggressive transgressions against a type $\bar{\theta}$ relative to a type $\underline{\theta}$ without risking war (formally, $t(\bar{\theta}) > t(\underline{\theta})$). And, because C is willing to risk war sometimes to pursue a more aggressive transgressions (by the condition on $\hat{Q}$), C is selecting a transgression level that will make a type $D(\bar{\theta})$ indifferent between war and peace, and will make a type $D(\underline{\theta})$ go to war.

How can private increases in D's wartime payoffs make D worse off? Within Case 3, the shift in D's private type (from $\underline{\theta}$ to $\bar{\theta}$) makes war a better outcome and hassling a worse outcome.[3] These imply that $D(\bar{\theta})$ is more willing to go to war when facing aggressive transgressions relative to $D(\underline{\theta})$, or that $t(\bar{\theta}) < t(\underline{\theta})$. Together, when C wants to avoid going to war, C will select $t(\bar{\theta})$. While this transgression level will make $D(\bar{\theta})$ indifferent between their wartime payoff and their optimal hassling payoff, C selecting $t(\bar{\theta})$ affords $D(\underline{\theta})$ some surplus because $D(\underline{\theta})$ is more efficent at hassling than $D(\bar{\theta})$. Under the conditions of the case, it is $D(\underline{\theta})$ that attains information rents.

This is one example model of how introducing multiple conflict options where private type influences these various forms of conflict can undermine the Banks (1990) and Fey and Ramsay (2011) results. We believe that reasonable readers may take issue with idiosyncrasies within the game form—for example, that there is no bargaining or signalling to speak of. This is why the next set of results are so critical: by establishing game-form free findings, we are essentially able to punt on the particulars of how the game plays out. In an environment without clearly defined institutions that inform

[3]For an empirical grounding, see the discussion above on precision strike capabilities and economic warfare.

how states bargain—which is a defining feature of the anarchic international order (Waltz, 2010)—the mechanism design approach insures that our results are robust to the broadest possible set of actualized bargaining protocols.

4 The Flexible-Response Crisis Bargaining Model

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

4.1 Structure of the Interaction

At the outset of the interaction, Nature assigns D's type, $\theta \in \mathbb{R}$. Without loss of generality, we assume D's war payoff increases with D's type,[4] while D's cost of hassling may increase, decrease, or neither. The realized value of θ is known only to D, but the prior distribution from which it is drawn is common knowledge. Let F denote the CDF of this prior distribution, and let Θ denote its support.

The interaction between the states takes the familiar form of a crisis bargaining game, except each state may engage in activity that affects outcomes short of war. First, C selects transgression $t \in \mathcal{T} \subseteq \mathbb{R}_+$. Following these choices, C and D partake in a bargaining process that may end in war. Like Banks (1990) and Fey and Ramsay (2011), we place no particular structure on the bargaining process. We simply assume that each player chooses from a set of available bargaining actions; these choices determine whether the game ends in war, and, if not, how the prize is divided. In the negotiation stage, we let $b_C \in \mathcal{B}_C$ denote C's bargaining strategy (offers, counteroffers, accept-reject plans, etc.). D's strategy consists of an analogous bargaining strategy $b_D \in \mathcal{B}_D$, as well as a level of hassling, $h \in \mathcal{H} \subseteq \mathbb{R}_+$.[5] A *game form* G consists of the bargaining action spaces, $\mathcal{B}_A$ and $\mathcal{B}_D$, along with an *outcome function* g mapping from the choices (t, h, b_C, b_D) into the set of possible crisis bargaining outcomes.[6] We decompose the

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

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

[6]The game form represents the elements of the model that are specific to a particular bargaining protocol. Implicitly, then, we take the type space, prior distribution, transgression action sets, hassling action set, cost

outcome function g into three components: whether war occurs, what C receives from bargaining, and what D receives from bargaining.[7] Whether war occurs or not depends solely on actions taken in bargaining. Let $\pi^g(b_C, b_D) \in \{0, 1\}$ be an indicator for whether the interaction ends without war.[8] Conditional on war not occurring, each player's payoff depends on the bargaining behavior, C's choice of transgression, and D's selection of hassling. Let $V_C^g(t, h, b_C, b_D)$ and $V_D^g(t, h, b_C, b_D)$ denote the benefits that C and D receive, respectively, in case when war is avoided.

Payoffs depend on the outcome of bargaining, including the costs of the hassling or transgression when war does not occur.[9] War payoffs may depend on D's private information, but do not depend on any of the endogenous choices in the game, including transgression and hassling. We therefore write war payoffs as $W_A(\theta)$ and $W_D(\theta)$. We assume W_D is strictly increasing, so higher types of D can be interpreted as stronger in wartime. If war is avoided, each player receives their division of the spoils but must pay the cost of their transgression or hassling. Let $K_C(t, h)$ denote the cost to A, and let $K_D(h, \theta)$ denote the cost to D. We assume that K_C is strictly increasing in t and weakly decreasing in h, and we assume K_D is strictly increasing in h. We let $h = 0$ denote no hassling, which entails assuming $0 \in \mathcal{H}$ and $K_D(0, \theta) = 0$ for all θ. Putting these together, the players' utility functions in a given game form are as follows:

$$u_A^g(t, h, b_C, b_D \mid \theta) = (1 - \pi^g(b_C, b_D))W_A(\theta) + \pi^g(b_C, b_D)[V_C^g(t, h, b_C, b_D) - K_C(t, h)], \tag{1}$$

$$u_D^g(t, h, b_C, b_D \mid \theta) = (1 - \pi^g(b_C, b_D))W_D(\theta) + \pi^g(b_C, b_D)[V_D^g(t, h, b_C, b_D) - K_D(h, \theta)]. \tag{2}$$

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

functions, and war payoff functions as primitives of the model rather than features of a game form G.

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

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

[9] Here we restrict attention to models where final bargaining payoffs captured in V_D do not depend on D's true type.

4.2 Solution Concept and Direct Mechanisms

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

We will focus on the incentives of D, the player with private information. Given an equilibrium of a flexible-response crisis bargaining game, we can summarize the outcome of the game for each type of D with three functions:[10]

- Their hassling level, $h(\theta)$.
- Whether a bargained outcome prevails, $\pi(\theta)$.
- Their settlement value in case 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 the equilibrium bargaining strategy of type θ', D's expected utility from doing so would be:

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

While mimicking another type's strategy may change the hassling level, the occurrence of conflict, and the settlement value for θ, it does not change D's war payoff, nor the cost D pays for any given hassling level.[11] In equilibrium, no type may have a strict incentive to mimic another type's bargaining strategy. We can phrase this requirement as an *incentive compatibility* condition on the direct mechanism. Let $U_D(\theta)$ denote each type's expected utility along the path of play, so that $U_D(\theta) = \Phi_D(\theta \mid \theta)$.

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

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

To identify regularities in the outcomes of flexible-response crisis bargaining games,

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

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

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

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

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

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

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

5 Results

5.1 General Patterns

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

Our first general result is that if two types both reach a settlement in equilibrium, then the one with greater hassling effectiveness must end up no worse off than the other. All of our formal results, including the following lemma, apply only to incentive compatible direct mechanisms with voluntary agreements—those that satisfy (IC) and (VA). By the revelation principle, this means that the properties identified in our results hold for *all* pure strategy equilibria of *all* flexible-response crisis bargaining games.[12]

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

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

Corollary 3. *If $\pi(\theta) = \pi(\theta') = 1$, $h(\theta) > 0$, and θ' has greater hassling effectiveness than θ, then $U_D(\theta) < U_D(\theta')$.*

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

Lemma 2. *If $\pi(\theta) = \pi(\theta') = 1$ and $h(\theta) \leq h(\theta')$, then $V_D(\theta) \leq V_D(\theta')$. Furthermore, if $h(\theta) < h(\theta')$, then $V_D(\theta) < V_D(\theta')$.*

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

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

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

[12] All proofs are in the Appendix.

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

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

If more effective types choose higher levels of hassling, then they must also receive more favorable settlements in equilibrium. The next result is an immediate corollary of Lemma 2 and Lemma 3.

Corollary 4. *Assume* (DD) *holds. If* $\pi(\theta) = \pi(\theta') = 1$ *and* θ' *has greater hassling effectiveness than* θ*, then* $V_D(\theta) \leq V_D(\theta')$.

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

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

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

The bounded variation condition essentially reduces the number of degrees of freedom to consider in our analysis of incentive compatible direct mechanisms. It allows us to back out the settlement value for each type that ends up at when war does not occur, $V_D(\theta)$, from three components of the model: (1) the indicator for which types end up at war, $1 - \pi(\cdot)$; (2) the hassling of each type, $h(\cdot)$; and (3) the lowest type's equilibrium utility, $U_D(\underline{\theta})$. It becomes even simpler when the lowest type's equilibrium outcome is war, as then $U_D(\underline{\theta}) = W_D(\underline{\theta})$, an exogenous constant. The following proposition summarizes the relationship between these factors and each type's equilibrium utility.[13]

Proposition 2. *Assume (BV) holds. For all $\theta_0 \in \Theta$,*

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

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

5.2 When Stronger Types Are Better at Hassling

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

When higher types are more effective at both war and hassling, then higher types of D must be better off in equilibrium regardless of which outcome prevails. The following result extends Lemma 1 to apply even when a bargained outcome is not guaranteed.

Lemma 4. *Assume θ improves hassling effectiveness. If $\theta < \theta'$, then $U_D(\theta) \leq U_D(\theta')$.*

In fact, in this case higher types of D should usually be strictly better off than lower types. If two types both go to war in equilibrium, then of course the higher type is better off. If the outcome is a bargained outcome with nonzero hassling for both types, then the higher type must also be better off, per Corollary 3. Finally, if the lower type fights and the higher type does not, the voluntary agreements condition ensures that the higher type must still receive at least its war payoff, making it better off. Therefore, two types can receive the same payoff only if the lower type avoids war and invests nothing in hassling.

Lemma 5. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and $U_D(\theta) = U_D(\theta')$, then $\pi(\theta) = 1$ and $h(\theta) = 0$.*

[13] The condition on K_D in (BV) is slightly stronger than necessary for Proposition 2. For this result, we only need $K_D(h, \cdot)$ to be Lipschitz in θ for each fixed $h \in \mathcal{H}$.

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

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

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

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

The *settlement utility is relatively increasing* if

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

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

Lemma 6. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and (WURI) holds, then $\pi(\theta) \geq \pi(\theta')$. If $\theta < \theta'$ and (SURI) holds, then $\pi(\theta) \leq \pi(\theta')$.*

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

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

While Lemma 6 is useful for understanding how private information affects the occurrence of war in flexible response crisis bargaining games, its practical applicability is somewhat limited. Ideally, we would be able to say on the basis of the model primitives—the war payoff and hassling cost functions—whether stronger types will be associated with a greater likelihood of conflict in any given strategic environment. However, the WURI and SURI conditions do not exclusively concern model primitives, as they depend on the levels of hassling chosen on the path of play. This raises the possibility that the relationship between D's private type and the likelihood of conflict may vary depending on the exact bargaining protocol.

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

Lemma 7. *Assume θ improves hassling effectiveness,* (DD) *holds, and* $\max \mathcal{H} = \bar{h} < \infty$*. If $W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta')$ for all $\theta, \theta' \in \Theta$ such that $\theta < \theta'$, then* (WURI) *holds.*

There is not an analogous sufficient condition for the settlement utility to be relatively increasing. The obstacle to such a condition is our assumption that $h = 0$ is always feasible at zero cost.[14] This means the marginal effect of θ on the cost of the hassling is zero for $h = 0$, ruling out any kind of sufficient condition for the marginal effect of θ on the hassling cost to always exceed its effect on the war payoff. At most, if we assume decreasing differences in the cost of hassling, we can make the (SURI) condition slightly less onerous to check. In this case, letting $\underline{h}$ denote the minimal hassling among types that end up in a bargained resolution, a sufficient condition is that $W_D(\theta') - W_D(\theta) < K_D(\underline{h}, \theta) - K_D(\underline{h}, \theta')$ for all $\theta < \theta'$. If this condition holds, then the equilibrium must entail low types of D fighting and high types of D settling.

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

5.3 When Stronger Types Are Worse at Hassling

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

While this situation may seem counter-intuitive, this can occur whenever hassling technologies are not well designed for warfare (or vice versa). For example, while in the early 2000s the U.S. possessed an ability to conduct sophisticated precision strike attacks, it lacked a robust ability to conduct a prolonged counterinsurgency; thus, the U.S. could capably hassle Iraq, but could not as easily go to war with Iraq. Section 2 includes additional examples.

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

Lemma 8. *Assume θ degrades hassling effectiveness. If $\theta < \theta'$, then $\pi(\theta) \geq \pi(\theta')$.*

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

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

Lemma 9. *Assume θ degrades hassling effectiveness.*

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

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

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

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

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

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

Lemma 9 does not pin down anything about $h(\theta)$, the relationship between D's type and D's final hassling level. We know from Lemma 3 that higher types will choose lower hassling levels if the decreasing differences property is satisfied. Besides that, our results so far place no restrictions on how types affect hassling. This is no accident. If higher types have higher marginal costs of hassling (i.e., decreasing differences is satisfied) and there is bounded variation in the model primitives, then we can rationalize a wide variety of relationships between types and the equilibrium level of hassling. More formally, for any mapping from types into hassling that is non-increasing and absolutely continuous, we can design a direct mechanism that is incentive compatible and has voluntary agreements in which the choices of revisionism follow this mapping. Under these conditions, the *necessary* conditions identified in Lemma 9 are *sufficient* for a bargaining equilibrium.

Proposition 3. *Assume* θ *degrades hassling effectiveness and* (BV) *and* (DD) *hold. If the direct mechanism* (h, π, V_D) *satisfies conditions (a)–(c) of Lemma 9 with* $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ *and* h *is absolutely continuous and non-increasing, then this direct mechanism satisfies* (IC) *and* (VA)*.*

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

6 Conclusion

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

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

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

References

Acharya, A., Grillo, E. *et al.* (2015) War with crazy types, *Political Science Research and Methods*, **3**, 281–307. 1

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

Ashworth, S. and Bueno de Mesquita, E. (2006) Monotone comparative statics for models of politics, *American Journal of Political Science*, **50**, 214–231. 5.1

Ashworth, S. and Sasso, G. (2019) Delegation to an overconfident expert, *The Journal of Politics*, **81**, 692–696. 1

Asmus, R. (2010) *A little war that shook the world: Georgia, Russia, and the future of the West*, St. Martin's Press. 2

Baliga, S., de Mesquita, E. B. and Wolitzky, A. (2020) Deterrence with imperfect attribution, *American Political Science Review*, **114**, 1155–1178. 1, 1

Banks, J. S. (1990) Equilibrium behavior in crisis bargaining games, *American Journal of Political Science*, pp. 599–614. 1, 1, 3.2, 4.1, 7, 5.2, 5.3, 5.3, 6

Baron, D. P. and Besanko, D. (1987) Commitment and fairness in a dynamic regulatory relationship, *The Review of Economic Studies*, **54**, 413–436. 1

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

Benson, B. and Smith, B. C. (2020) Do alliances cause commitment problems and war?, *Working Paper*. 2

Chassang, S. and Padró i Miquel, G. (2009) Defensive weapons and defensive alliances, *American Economic Review*, **99**, 282–86. 1

Cobzaş, c., Miculescu, R. and Nicolae, A. (2019) *Lipschitz Functions*, Springer. C.12

Coe, A. J. (2014) Sanctions as instruments of regime change: A case study of civil conflict in iraq during containment, Tech. rep., Working Paper. 1

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

Cordesman, A. H. (2018) 21st century conflict: From "revolution in military affairs" (rma) to a "revolution in civil-military affairs" (rcma), *Center for Strategic and International Studies*. 2

Debs, A. and Monteiro, N. P. (2014) Known unknowns: power shifts, uncertainty, and war, *International Organization*, **68**, 1–31. 2

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

Fearon, J. D. (1995) Rationalist explanations for war, *International organization*, **49**, 379–414. 1

Fearon, J. D. (1996) Bargaining over objects that influence future bargaining power. 2

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

Fey, M. and Kenkel, B. (2019) Is an ultimatum the last word on crisis bargaining? 1, 8

Fey, M. and Ramsay, K. W. (2009) Mechanism design goes to war: peaceful outcomes with interdependent and correlated types, *Review of Economic Design*, **13**, 233. 1, 11

Fey, M. and Ramsay, K. W. (2011) Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict, *American Journal of Political Science*, **55**, 149–169. 1, 1, 3.2, 4.1, 7, 10, 4.2, 4.2, 5.3, 5.3, 6

Gartzke, E. (2007) The capitalist peace, *American journal of political science*, **51**, 166–191. 2

Gartzke, E., Li, Q. and Boehmer, C. (2001) Investing in the peace: Economic interdependence and international conflict, *International organization*, **55**, 391–438. 2

Gartzke, E. and Lindsay, J. R. (2015) Weaving tangled webs: offense, defense, and deception in cyberspace, *Security Studies*, **24**, 316–348. 1

Gartzke, E. and Lindsay, J. R. (2017) Thermonuclear cyberwar, *Journal of cybersecurity*, **3**, 37–48. 2

Gurantz, R. and Hirsch, A. V. (2017) Fear, appeasement, and the effectiveness of deterrence, *The Journal of Politics*, **79**, 1041–1056. 1, 2, 3

Hörner, J., Morelli, M. and Squintani, F. (2015) Mediation and peace, *The Review of Economic Studies*, **82**, 1483–1501. 1

Jervis, R. (1979) Why nuclear superiority doesn't matter, *Political Science Quarterly*, **94**, 617–633. 2

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

Lanoszka, A. (2016) Russian hybrid warfare and extended deterrence in eastern europe, *International Affairs*, **92**, 175–195. 1

Mazarr, M. J. (2015) Struggle in the gray zone, *War on the Rocks*. 1

McCormack, D. and Pascoe, H. (2017) Sanctions and preventive war, *Journal of Conflict Resolution*, **61**, 1711–1739. 1, 1, 2

Meirowitz, A. (2007) In defense of exclusionary deliberation: communication and voting with private beliefs and values, *Journal of Theoretical Politics*, **19**, 301–327. 1

Meirowitz, A., Sartori, A. E. *et al.* (2008) Strategic uncertainty as a cause of war, *Quarterly Journal of Political Science*, **3**, 327–352. 1

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

Milgrom, P. and Segal, I. (2002) Envelope theorems for arbitrary choice sets, *Econometrica*, **70**, 583–601. C.5

Myerson, R. B. (1979) Incentive compatibility and the bargaining problem, *Econometrica: journal of the Econometric Society*, pp. 61–73. 4.2

Powell, R. (1999) *In the shadow of power: States and strategies in international politics*, Princeton University Press. 1

Powell, R. (2006) War as a commitment problem, *International organization*, **60**, 169–203. 2

Powell, R. (2015) Nuclear brinkmanship, limited war, and military power, *International Organization*, **69**, 589–626. 1, 1, 2

Schelling, T. C. (1966) *Arms and influence*, Yale University Press. 2

Schram, P. (2020) When deterrence fails: How improved hassling capabilities produce worse outcomes, *Working Paper*. 1, 3

Schram, P. (Forthcomming) Hassling: How states prevent preventive wars, *American Journal of Political Science*. 1, 1, 2

Schultz, K. A. (1999) Do democratic institutions constrain or inform? contrasting two institutional perspectives on democracy and war, *International Organization*, **53**, 233–266. 1

Schultz, K. A. (2010) The enforcement problem in coercive bargaining: Interstate conflict over rebel support in civil wars, *International Organization*, **64**, 281–312. 1, 1, 2

Slantchev, B. L. (2003) The principle of convergence in wartime negotiations, *American Political Science Review*, pp. 621–632. 1

Spaniel, W. (2019) *Bargaining over the bomb: the successes and failures of nuclear negotiations*, Cambridge University Press. 2

Spaniel, W. and Bils, P. (2018) Slow to learn: Bargaining, uncertainty, and the calculus of conquest, *Journal of Conflict Resolution*, **62**, 774–796. 1

Spaniel, W. and Malone, I. (2019) The uncertainty trade-off: Reexamining opportunity costs and war, *International Studies Quarterly*, **63**, 1025–1034. 1, 1

Trevithick, J. (2016) Israel details long secret raid on syrian nuclear reactor, *The Warzone*. 2

Wagner, R. H. (2000) Bargaining and war, *American Journal of Political Science*, pp. 469–484. 1

Waltz, K. N. (2010) *Theory of international politics*, Waveland Press. 3.2

A Example Model Intuition

Assume that for a selected t, a type θ D optimally does not go to war. This type θ D selects an optimal h^* following

$$h^*(t) \in argmax_{h \in \mathbb{R}_+} \left\{ X_D - t + h - \frac{(h)^2}{F(\theta)} \right\}.$$

I take first-order conditions with respect to h of the expression above to identify the optimal level of hassling h^*. This unique value is $h^* = \frac{F(\theta)}{2}$. Using h^*, we re-write D's utility in terms of the selected t and parameters α and θ. This is

$$U_D(\theta) = \begin{cases} X_D - t + \frac{F(\theta)}{4} & if\, X_D - t + \frac{F(\theta)}{4} \geq W_D(\theta), \\ W_D(\theta) & otherwise. \end{cases}$$

Because C faces no costs from increases in t, C wants to select greater values of t, unless it results in too great a disutility from the possibility of war. Thus, C will select values of t that will make either type $\underline{\theta}$ or type $\bar{\theta}$ indifferent between war and not going to war. Using D's utility above, the transgression level that makes $D(\theta)$ indifferent between war and hassling is

$$t(\theta) = X_D - W_D(\theta) + \frac{F(\theta)}{4}.$$

While we assumed $W_D(\theta)$ is increasing in θ, $F(\theta)$ could be increasing or decreasing, implying that $t(\theta)$ could be increasing or decreasing. So, we must consider cases.

Cases 1 and 2: When $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} \geq W_D(\bar{\theta}) - W_D(\underline{\theta})$ holds, then $t(\bar{\theta}) \geq t(\underline{\theta})$. C's utility from selecting $t(\underline{\theta})$ (which never provokes D to escalate to war) is

$$EU_C(t(\underline{\theta})) = X_C + t(\underline{\theta}) - Pr(\underline{\theta})\frac{F(\underline{\theta})}{2} - (1 - Pr\underline{\theta}))\frac{F(\bar{\theta})}{2},$$

or

$$EU_C(t(\underline{\theta})) = X_C + X_D - W_D(\underline{\theta}) - Pr(\underline{\theta})\left(\frac{F(\underline{\theta})}{4}\right) + (1 - Pr\underline{\theta}))\left(\frac{F(\underline{\theta})}{4} - \frac{F(\bar{\theta})}{2}\right).$$

Similarly, we can say

$$EU_C(t(\bar{\theta})) = Pr(\underline{\theta})W_C + (1 - Pr\underline{\theta}))\left(X_C + X_D - W_D(\bar{\theta}) - \frac{F(\bar{\theta})}{4}\right).$$

Using these expressions, C will select $t(\underline{\theta})$ when

$$Pr(\underline{\theta})\left(X_C + X_D - W_C - \frac{F(\underline{\theta})}{4}\right) - W_D(\underline{\theta}) - (1 - Pr\underline{\theta}))\left(\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} - W_D(\bar{\theta})\right) \geq 0,$$

and $t(\bar{\theta})$ otherwise. Note that the left hand side is $\hat{Q}$.

We can then define D's utility. D will attain their wartime utility in all cases other than when $\hat{Q} \geq 0$ and D is type $\bar{\theta}$. This D attains

$$U_D(\bar{\theta}) = X_D - X_D + W_D(\underline{\theta}) - \frac{F(\underline{\theta})}{4} + \frac{F(\bar{\theta})}{4},$$

or

$$U_D(\bar{\theta}) = W_D(\underline{\theta}) - \frac{F(\underline{\theta})}{4} + \frac{F(\bar{\theta})}{4}.$$

Cases 3 and 4: When $\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} < W_D(\bar{\theta}) - W_D(\underline{\theta})$ holds, then $t(\underline{\theta}) > t(\bar{\theta})$. C's utility from selecting $t(\bar{\theta})$ (which never provokes D to escalate to war) is

$$EU_A(t(\bar{\theta})) = X_A + t(\bar{\theta}) - Pr(\underline{\theta})\frac{F(\underline{\theta})}{2} - (1 - Pr(\underline{\theta}))\frac{F(\bar{\theta})}{2},$$

or

$$EU_A(t(\bar{\theta})) = X_A + X_D - W_D(\bar{\theta}) + Pr(\underline{\theta})\left(\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{2}\right) + (1 - Pr(\underline{\theta}))\left(-\frac{F(\bar{\theta})}{4}\right).$$

We can also say

$$EU_A(t(\underline{\theta})) = (1 - Pr(\underline{\theta}))W_A + Pr(\underline{\theta})\left(X_A + X_D - W_D(\underline{\theta}) - \frac{F(\underline{\theta})}{4}\right).$$

Using these expressions, C will select $t(\bar{\theta})$ when

$$(1 - Pr(\underline{\theta}))\left(X_A + X_D - W_A - \frac{F(\bar{\theta})}{4}\right) - W_D(\bar{\theta}) + Pr(\underline{\theta})\left(\frac{F(\bar{\theta})}{4} - \frac{F(\underline{\theta})}{4} + W_D(\underline{\theta})\right) \geq 0$$

and $t(\underline{\theta})$ otherwise. Note that the left hand side is $\tilde{Q}$.

We can then define D's utility. D will attain their wartime utility in all cases other than when $\tilde{Q} \geq 0$ and D is type $\underline{\theta}$. This D attains

$$U_D(\underline{\theta}) = X_D - t(\bar{\theta}) + \frac{F(\underline{\theta})}{4},$$

or

$$U_D(\underline{\theta}) = W_D(\bar{\theta}) - \frac{F(\bar{\theta})}{4} + \frac{F(\underline{\theta})}{4}.$$

This can give the key features of the equilibrium. Across all cases, $D(\theta)$ hassles if $t^* \leq t(\theta)$ and goes to war otherwise.

B Equilibrium and Direct Mechanism

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

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

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

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

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

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

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

Consequently, incentive compatibility is equivalent to

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

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

C Proofs of Results in Text

C.1 Proof of Lemma 1

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

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

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

□

C.2 Proof of Corollary 3

Corollary 3. *If $\pi(\theta) = \pi(\theta') = 1$, $h(\theta) > 0$, and θ' has greater hassling effectiveness than θ, then $U_D(\theta) < U_D(\theta')$.*

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

C.3 Proof of Lemma 2

Lemma 2. *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. □

C.4 Proof of Lemma 3

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

Proof. (IC) implies:

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

A rearrangement of terms gives

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

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

C.5 Proof of Proposition 2

We first state a helpful lemma.

Lemma 10. *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. □

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

Proposition 2. *Assume* (BV) *holds. For all* $\theta_0 \in \Theta$,

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

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

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

at each point where U_D is differentiable. Furthermore, (BV) implies Φ_D is Lipschitz continuous in θ. The claim then follows from Lemma 10 and Milgrom and Segal (2002, Corollary 1). □

C.6 Proof of Lemma 4

Lemma 4. *Assume* θ *improves hassling effectiveness. If* $\theta < \theta'$, *then* $U_D(\theta) \leq U_D(\theta')$.

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

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

□

C.7 Proof of Lemma 5

Lemma 5. *Assume* θ *improves hassling effectiveness. If* $\theta < \theta'$ *and* $U_D(\theta) = U_D(\theta')$, *then* $\pi(\theta) = 1$ *and* $h(\theta) = 0$.

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

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

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

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

C.8 Proof of Lemma 6

Lemma 6. *Assume θ improves hassling effectiveness. If $\theta < \theta'$ and (WURI) holds, then $\pi(\theta) \geq \pi(\theta')$. If $\theta < \theta'$ and (SURI) holds, then $\pi(\theta) \leq \pi(\theta')$.*

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

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

while (IC) implies

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

Combining these gives

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

which in turn implies

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

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

C.9 Proof of Lemma 7

Lemma 7. *Assume θ improves hassling effectiveness, (DD) holds, and $\max \mathcal{H} = \bar{h} < \infty$. If $W_D(\theta') - W_D(\theta) > K_D(\bar{h}, \theta) - K_D(\bar{h}, \theta')$ for all $\theta, \theta' \in \Theta$ such that $\theta < \theta'$, then (WURI) holds.*

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

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

which is equivalent to

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

Therefore, under the hypothesis of the lemma, we have

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

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

C.10 Proof of Lemma 8

Lemma 8. *Assume θ degrades hassling effectiveness. If $\theta < \theta'$, then $\pi(\theta) \geq \pi(\theta')$.*

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

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

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

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

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

C.11 Proof of Lemma 9

We first prove a lemma demonstrating equivalence between the characterization of V_D in Lemma 9 and that of U_D in Proposition 2.

Lemma 11. *Assume (BV) holds and consider any direct mechanism (h, π, V_D). Suppose there exists $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ such that $\pi(\theta) = 1$ for all $\theta < \hat{\theta}$ and $\pi(\theta) = 0$ for all $\theta > \hat{\theta}$. Then Equation 3 holds for all $\theta_0 \in \Theta$ if and only if $U_D(\hat{\theta}) = W_D(\hat{\theta})$ and Equation 4 holds for all $\theta_0 < \hat{\theta}$.*

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

$W_D(\hat\theta)$. Then, for all $\theta_0 < \hat\theta$, Equation 3 implies

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

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

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

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

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

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

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

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

Lemma 9. *Assume θ degrades hassling effectiveness.*

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

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

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

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

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

Proof. (a) follows from Lemma 8. (b) and (c) follow from Lemma 11. (d) follows from Lemma 1. □

C.12 Proof of Proposition 3

Proposition 3. *Assume θ degrades hassling effectiveness and* (BV) *and* (DD) *hold. If the direct mechanism (h, π, V_D) satisfies conditions (a)–(c) of Lemma 9 with $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ and h is absolutely continuous and non-increasing, then this direct mechanism satisfies* (IC) *and* (VA).

Proof. As a preliminary, note that because K_D is Lipschitz and h is absolutely continuous, $K_D(h(\theta), \theta)$ is absolutely continuous when viewed as a function of θ (Cobzaş *et al.*, 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

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

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

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

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

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

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

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

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

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

Finally, we have $\Phi_D(\theta' \,|\, \theta) = W_D(\theta)$ for all $\theta' > \hat{\theta}$. Altogether, these findings imply $\Phi_D(\theta' \,|\, \theta)$ 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 $U_D(\theta) \ge \lim_{\theta \to \hat{\theta}-} \Phi_D(\theta' \,|\, \theta) = W_D(\theta)$ for all $\theta \le \hat{\theta}$, so (VA) also holds. □