Putting Yourself Out There: Designing Credible Assurances Through Endogenous Power Shifts *

Peter Schram† Brandon Yoder ‡

August 30, 2024

Abstract

Policymakers and scholars often identify low-level conflict as destabilizing and lament the existence and proliferation of tactics by which states can impose moderate costs on each other. We present an alternate perspective: limited conflict – both its availability and execution – can forestall high-level conflict and facilitate credible assurances and trust-building between potential adversaries. We examine a formal model of incomplete information under shifting power, in which a rising actor has private information about the degree to which its preferences align with a declining actor's. In round 1 the riser chooses a degree of revision against the decliner, which endogenously informs the size of the power shift. The decliner then responds with some level of preventive action that mitigates the power shift, ranging on a continuum from inaction, limited actions ("hassling"), or a full-scale war. We find that the availability of limited responses by the decliner both precludes war and allows the riser to credibly communicate its preferences by limiting or refraining from revisionist actions. Additionally, surprisingly, we find that by allowing the rising state to conduct costly power shifts can both increase overall welfare and grant the decliner greater overall payoffs. Our findings in this more general and realistic strategic setting contradict prominent arguments that mistaken preventive war frequently occurs against rising states with no hostile intent due to incomplete information.

*We thank Brenton Kenkel, Brad Smith, and other participants at the Vanderbilt Conflict Workshop.

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

‡Senior Lecturer, Australian National University. Brandon.Yoder@anu.edu.au

When and how do power shifts lead to war? Most literature on this topic finds that preventive war occurs when the size and speed of the power shift is sufficiently large. These dynamics are exacerbated, or sometimes activated altogether, by incomplete information about either capabilities or resolve. Moreover, many scholars argue that under uncertainty, preventive war can occur even between rising and declining states with mutually benign intentions – the decliner is sufficiently vulnerable to future exploitation by a potentially hostile riser that it initiates preventive conflict despite the possibility that the riser holds compatible preferences. Put another way, even if a rising state has compatible preferences, it may struggle to credibly assure an adversary that it is not an enemy, thus leading to an unnecessary and wasteful future conflict.

Existing studies have made at least one of two strong assumptions: 1) that the power shift is solely exogenous; i.e., outside the rising state's control and 2) that the declining state has no options for forestalling the power shift other than full-scale preventive war.[1] Yet each of these assumptions are well-recognized as being unrealistic. The majority of meaningful shifts in the military balance occur at least in part through the deliberate choices of the rising state (Chadefaux 2011; Debs and Monteiro 2014; Miller 2021) rather than simply through exogenous differential growth rates. And declining states clearly have many options for containing a rising state's power short of war (Schram 2021).

We relax these unrealistic assumptions in a series of models with shifting power under incomplete information, where the declining state does not know whether the rising state is benign or hostile.[2] Firstly, we model two ways in which a power shift can be at least partially endogenous to the rising state's choices: 1) deliberate investment in its military capabilities; 2) acquisition of disputed assets that are themselves sources of power. Secondly, we realistically model the declining state's menu of options to suppress the rising state's power gains on a

[1] Exceptions to one or the other of these assumptions are discussed in detail below.

[2] The terms "benign" and "hostile" are shorthand for rising states whose goals are relatively compatible and relatively incompatible with the decliner's, respectively. Two states' goals are compatible to the extent that the realization of one's goals advances (or at least does not impede) the other's.

continuum from complete acquiescence to preventive war, with a full range of containment measures short of war in between. We refer to these intermediate actions as "hassling."

Our models yield several novel results, which are robust across specifications. Most generally, we find that the availability of hassling options can forestall escalation to preventive war by allowing the declining state to reduce the power shift and thus mitigate the commitment problem. This result holds even with incomplete information about the rising state's intentions and about whether it has invested in its military capabilities. Moreover, we show that the combination of hassling options (by the decliner) and endogenous power shifts (by the riser) facilitates credible signals of the rising state's intentions. In brief, to the extent that the riser can enhance its power through military investment or revisionist behavior, refraining from such actions constitutes a credible assurance that its intentions are benign. The information from these signals then produces more efficient outcomes: it narrows the conditions under which preventive war occurs, and allows the decliner to reduce its degree of hassling in response. Indeed, under certain conditions the riser's restraint is credible enough to eliminate costly conflict *entirely* – the decliner takes no hassling measures at all. Thus, the endogenous component of a power shift, by engendering credible signals, can counteract the pernicious effects of an exogenous shift when both are present.

These findings indicate that power shifts are not as perilous as the conventional wisdom holds. When certain unrealistic assumptions are relaxed, so that states have more flexible policy options available to manage their rise/decline, we see peaceful outcomes occur even under worst-case conditions, i.e., when the power shift is pronounced and the rising state can conceal investments in its military power.

From this follows an even more surprising final result: Since the potential for an endogenous shift can yield information that forestalls costly conflict, declining states can actually *benefit* from giving a rising state the means to enhance its own military power. In practice, this can be done through asymmetric terms of exchange that boost the riser's underlying

economic power, but also by neglecting strategically valuable territory, transferring sensitive technology, and accepting various forms of dependence on the rising state. In other words, by "putting itself out there" – leaving itself vulnerable to a hostile rising state that would exploit such opportunities to enhance its own military power – the decliner can glean valuable information about the riser's intentions that can reduce or eliminate preventive conflict and enhance the decliner's welfare, on balance.

Our argument has implications for explaining the pattern of US foreign policy toward China since the end of the Cold War. The US policy of "engagement" toward China in the 1990s and early 2000s has been critiqued as a naive response that served to abet China's economic growth and allowed it to acquire formidable coercive capabilities. Yet our argument suggests that there is often a rational basis for such actions on pure security grounds, due to their value as a screening mechanism. Allowing China to empower itself yielded valuable information to the US about China's intentions, and potentially could have averted competition between the two powers entirely had China opted for restraint and credibly signaled benign intentions. However, China has instead invested heavily in its military and non-military coercive capabilities (e.g., the Belt-and-Road Initiative) in the past two decades, signaling revisionist aims and prompting the US to turn toward increasing containment of China. Our novel signaling mechanism demonstrates that even if China has now demonstrated its revisionist aims, there existed an underlying strategic rationale in facilitating China's rise initially.

1 Literature

The relationship between power shifts and war is one of the most heavily studied topics in international relations. The basic mechanism that makes power shifts dangerous is the commitment problem: as a rising state gains power, it cannot commit to maintain the status quo, and will instead have incentives to revise the international system in ways that may

harm the interests of others. Declining states therefore have incentives to initiate preventive war in the present, while they still hold a power advantage, rather than allow the rising state to effect unfavorable revisions in the future (Levy 1987; Kim and Morrow 1992; Fearon 1995; Copeland 2000; Powell 2006; Chadefaux 2011).

Moreover, a prominent realist literature argues that large power shifts can produce full-scale preventive wars between mutually-benign rising and declining states, due in part to the decliner's intractable uncertainty about the compatibility of a rising state's goals (Layne 1993; Copeland 2000; Mearsheimer 2001; Edelstein 2002; Montgomery 2006; Rosato 2015; Shifrinson 2018). These scholars argue that because hostile rising states have strong incentives to misrepresent their intentions, their reassurance signals are non-credible, leaving declining states highly uncertain.[3] The decliner's combination of uncertainty and future vulnerability gives it strong incentives to take preventive action even against risers that have exhibited cooperative behavior. Past rationalist work on power shifts under uncertainty has corroborated the basic finding that such "mistaken" preventive war against rising states with modest aims is possible under shifting power and incomplete information (Powell 1999, 1996; Wolford, Reiter and Carrubba 2011; Debs and Monteiro 2014).

This widespread conclusion that power shifts and uncertainty produce mistaken preventive conflict is predicated on two assumptions, at least one of which obtains in all existing models of the phenomenon.[4] The first is that declining states can only marshal all-or-nothing responses to shifting power, consisting of either full-scale preventive war or complete acquiescence. Yet declining states clearly have many options for containing a rising state's power short of war (Schram 2021), including low-level military operations but also economic sanctions (Martin 1992; Schwebach 2000; Lektzian and Sprecher 2007), counterbalancing

[3]This is contrary to a large rationalist literature on credible reassurance that shows how mutually-benign states can reliably signal their intentions to avoid unnecessary conflict, but which has largely assumed static power, e.g. Jervis (1978); Kydd (2005); Glaser (2010). Exceptions that examine reassurance under shifting power are discussed below.

[4]Other research like Schram (2022) and Joseph (2023) has hassling and endogenous power shifts, but credible assurance is not possible in those papers due to the nature (or lack thereof) of incomplete information.

alliances (Gowa 1995), exclusionary institutional structures (He 2008; Lascurettes 2020), cyberattacks (Lonergan and Lonergan 2022), disinformation campaigns (Spaniel and Poznansky 2018), etc. Moreover, recent work has shown the introduction of hassling options to be consequential for the outcomes of power shifts. Schram (2021) shows that under complete information, a declining state can use hassling to forestall preventive conflict by reducing the size of the power shift and thus mitigating the commitment problem. Yoder (2019) and Spaniel and Poznansky (2020) allows hassling under incomplete information and find that the availability of such strategies increase the credibility of a rising state's cooperative signals, allowing the decliner to identify benign risers with greater confidence avoid mistaken preventive war. Yet in contrast to our argument here, these models all assume a purely exogenous rise, and find that low-level preventive conflict short of war is unavoidable under large power shifts.

The second common assumption in the literature is that shifts in power are exclusively exogenous to the rising state's choices. Yet Debs and Monteiro (2014) have shown that the majority of large, rapid power shifts – the kind most likely to cause war – are underpinned at least in part by the endogenous policy decisions of the rising state. These can include either military investments or revision of the international system in ways that also confer power gains, such as territorial expansion, formation of new institutional structures, or the propagation of asymmetrically favorable economic arrangements. We lump these categories of power-boosting policies under the umbrella term "transgressions."

Several studies model endogenous power shifts arising from transgressions, generally finding them less war-prone than exogenous shifts. Fearon (1996) shows that war does not occur with endogenous shifts under complete information. Debs and Monteiro (2014) corroborate this result, showing that incomplete information over the rising state's arming decision is necessary for war to occur in the context of endogenously-shifting power. Others examine endogenous shifts with incomplete information over resolve rather than the riser's military

investment, and similarly find that potential rising states are generally deterred from enacting revisions that would enhance their power by the declining state's credible threat of a preventive response. Kydd and McManus (2017) and Gurantz and Hirsch (2017) show that war occurs only with "undeterrable" rising states that prefer war to the status quo under the *ex ante* distribution of power, while (Schwarz and Sonin 2008) find that transgression only occurs when it is possible for the rising state to credibly threaten a "brinkmanship" strategy that leaves war to chance, but preventive war by the decliner is always out of equilibrium.

Nevertheless, each of these works study endogenous power shifts to the exclusion of any exogenous component of shifting power, which is also unrealistic – differential growth rates are a near-constant feature of the international system (Gilpin 1981; Copeland 2000). Thus, exogenously rising states face decisions about whether to accelerate their power gains through transgressions. The exclusion of the exogenous power shift in prior models is consequential because, as just discussed, revision is disincentivized when the power gains from it would be wiped out by the decliner's preventive response. This stacks the deck against conflict because under most conditions no revision – and thus no power shift – occurs in equilibrium, yielding largely peaceful outcomes. In contrast, we examine the *interaction* of exogenous and endogenous power shifts, where preventive incentives exist for the declining state even in the absence of endogenous choices by the rising state to enhance its own power. This allows us to analyze how endogenous shifts facilitate credible signaling through the riser's restraint that can counteract the pernicious effects of a concurrent exogenous shift.[5]

This paper generalizes from past work by modeling, for the first time, both endogenous and exogenous sources of rising power. We model transgressions-as-investments following Debs and Monteiro (2014), wherein the rising state pays a cost to enhance its power, with no direct payoff consequences for the decliner. We also model transgressions-as-revisions, wherein the

[5]Like us, Gurantz and Hirsch (2017) find that restraint itself is a credible signal that allows the declining state to update its beliefs that the prospective riser is more likely to be benign. However, because their power shift is purely endogenous to the riser's revision, there is no incentive for preventive war in the absence of revision, making the information from the riser's restraint inconsequential for the decliner's behavior.

asset at stake is itself a source of power. Here, transgression, if realized, directly benefits the rising state and harms the declining state in terms of the distribution of the asset.

However, we also depart from Debs and Monteiro (2014) by recognizing the possibility of arms control. Whereas they assume that the rising state's investment decision is only ever probabilistically observable, others have noted that the rising state has the option of unambiguously revealing its investment decision by acceding to arms control agreements with rigorous monitoring provisions (Coe and Vaynman 2020). Coe and Vaynman (2020) argue, however, that arms control is often rejected because such monitoring would reveal damaging information to third parties that could undermine the government's political survival. We therefore model transgressions-as-investments in two steps: first, the rising state chooses whether to reveal or conceal its investment decision, where revealing is inherently costly. Next, the rising state makes an investment decision, where investment increases its power but also carries direct costs. If the rising state plays "reveal," the declining state observes whether the riser is arming. On the other hand, if the rising state plays "conceal" its investment decision is unobservable, but the decliner can still update its beliefs in response to the observation that the riser has concealed. This specification is more realistic than that of Debs and Monteiro (2014), who assume away the possibility of arms control, but remains consistent with Coe and Vaynman (2020)'s mechanism by which arms control might not happen even when available.

Our main results remain robust across different forms of endogenous power shifts. With both transgressions-as-investments and transgressions-as-revision we find that hassling can forestall war even with uncertainty and large power shifts. Moreover, endogenous power shifts generate opportunities for credible reassurance signals that otherwise would not exist. This reduces the degree of hassling that occurs in equilibrium toward benign rising states, and, under some conditions, can even result in no inefficient conflict at all. Finally, the information from these signals can benefit declining states to such a degree that they might

choose to abet a rising state's capacity to invest in its military capabilities. We now turn to our models and their results.

2 Theoretical Intuition

Suppose we are in a setting with an exogenously rising power, but the declining state does not know whether the rising state is revisionist or not. Via the standard logic of commitment problems, this setting can lead to war. If the rising state is not revisionist and war occurs, then this is bad for all sides: if the non-revisionist rising state could credibly assure their benign intent, then a costly war can be avoided. But this setting poses challenges for credible communication. For example, suppose the non-revisionist type sent a message that said they were non-revisionist to convince the declining state to not declare war. This signal cannot be informative because the revisionist type would want to mimic the non-revisionist type's message, thereby allowing the revisionist type to rise and demand more in the future. And, this problem does not only apply to cheap-talk messaging. Outside of the riser degrading their own capabilities, an option that is costly, inefficient, and may not always be possible,[6] a benign riser cannot credibly signal through costly means like "burning cash" or investments in capabilities.[7]

When hassling is introduced, the dynamic can change. The declining state can use hassling—taking efforts to slow the rising state's rise—as a substitute for war, as shown in Schram (2021).[8] However, hassling can also allow the rising state to signal their type with their transgressions—put another way, credible communication through transgressions becomes possible in a world with hassling. A rising state can signal that they are not a revisionist type by not undertaking a transgression that is available to them. And, the revisionist-type

[6] This mechanism and its caveats are discussed in Reich (2022).

[7] Burning cash is not possible because the revisionist type would always be willing to pay more than the non-revisionist type in order to mimic the non-revisionist type if it means avoiding war before the rise occurs.

[8] Schram demonstrated that hassling can prevent war when there is an exogenous rise. This paper demonstrates that hassling can prevent unnecessary wars when a rising state is non-revisionist.

will transgress, thus separating and signaling their type. The declining state then adopts a more efficient hassling scheme—implementing a higher degree of hassling in response to transgression, and a lower degree of hassling in response to a non-transgression.

Our primary result regards the value of transgressions. Consider two settings. In one setting the rising state can transgress, and in the other setting the rising state has no transgression option. We can show that the transgression option is valuable because it can credibly signal type, thus leading to more efficient hassling. Interestingly, while transgressions here are costly, inefficient ways of shifting the power distribution (and therefore shifting bargaining outcomes), granting the rising state the ability to conduct transgressions generates welfare gains (Result 1). Additionally, and more surprising still, giving the rising state the ability to transgress and weaken D can lead to utility gains for the declining state (Result 2). These results suggests that D can benefit from leaving itself vulnerable to actions that lower its future payoffs, while yielding more efficient international outcomes.

3 Model

3.1 Setup

Consider a two-period bargaining model with a rising state (R) and a declining state (D). The declining state most prefers policies set at 0. The rising state's policy preferences, which will be denoted $\theta \in [0, \bar{\theta}]$, are private. Essentially, the rising state could be moderate (lower values of θ, closer to D's ideal point), or revisionist (higher values of θ, further from D's ideal point). In this model, we will consider settings where revisionist types have different-enough policy preferences and rise fast enough to incentivize D to act, even if D does not know R's true preferences.[9]

[9]Note that this is a reassurance model with uncertainty over the distance between the actor's policy preferences (Bils and Spaniel 2017; Haynes and Yoder 2020), not a standard bargaining model in which preferences are known to be at odds and uncertainty pertains to resolve to fight over a disputed asset. Here, the question is whether there is a dispute at all. On the distinction between reassurance and bargaining, see Kydd (1997, 119).

In our model, the declining state can both hassle and choose whether to allow the rising state to transgress. Past work has shown how in settings like this the availability of hassling (or sanctions) can "resolve" commitment problems and prevent a preventative war (Schram 2021). We show here that adding transgressions to an environment with hassling can allow transgressions to signal the policy preferences, thereby leading to a new, more peaceful equilibrium.

We present the general model here. The game is as follows.

1. Nature sets R's type (or R's ideal point) θ with commonly known cumulative distribution function $F(\cdot)$ over $\Theta \subset \mathbb{R}_+$, with $\bar{\theta} = max\{\Theta\}$.

2. D either enables R, thereby allowing R to conduct transgressions in the future, or does not. Formally, when D enables R, D sets $E_D = 1$, and R's transgression action set is $\mathcal{T} = \{0, \bar{t}\}$ with $0 < \bar{t}$. If D does not enable R, D sets $E_D = 0$, and R's transgression action set is $\mathcal{T} = \{0\}$.

3. Period 1 begins, and R selects transgression level $t \in \mathcal{T}$.

4. D can either go to war by setting $w_{D,1} = 1$, or not go to war by setting $w_D = 0$. When D sets $w_{D,1} = 0$, D also selects some period 1 offer $x_1 \in [0, \bar{\theta}]$, and some level of hassling $h \in \mathcal{H}$, where $\mathcal{H} \subset \mathbb{R}_+$ is a closed set with $max\{\mathcal{H}\} = \bar{h}$. When D does not go to war, the game moves to the next stage. Going to war terminates the game, and allows whoever wins the war to set the policy for periods 1 and 2.

5. R can go to war by setting $w_{R,1} = 1$ or accept the offer by setting $w_{R,1} = 0$. When R sets, $w_{R,1} = 1$, the game ends and whoever wins can set the policy for periods 1 and 2.

6. Period 2 begins. At the start of period 2, R's private type is revealed.

7. D can offers R some value $x_2 \in [0, \bar{\theta}]$ or can go to war by setting $w_{D,2} = 1$. When D sets $w_{D,2} = 1$, the game ends and whoever wins can set the policy for period 2.

8. R can go to war by setting $w_{R,2} = 1$ or can "accept" the offer by setting $w_{R,2} = 0$. When R sets $w_{R,2} = 1$, whoever wins can set the policy for period 2.

3.2 Payoffs

D's and R's payoffs are summarized in Table 1. We assume both actors have linear loss utilities in the distance between their ideal point and the implemented policy position, which is consistent with recent literature on signaling policy preferences (Bils and Spaniel 2017; Haynes and Yoder 2020). Additionally, in writing out these payoffs, we assume for now that if D and R go to war, the winner will implement their ideal point (we will prove this in the Appendix).[10]

Sometimes, the game ends in war. $P : \{1, 2\} \times \mathcal{T} \times \mathcal{H} \rightarrow [0, 1]$ is the likelihood that, should war occur, R wins. P is increasing in t, its second argument, and continuous and decreasing in h, its third argument. The first argument denotes the period number; to capture the exogenous power shift, $P(1, 0, 0) < P(2, 0, 0)$. Additionally, we assume $P(1, 0, 0) < P(2, 0, h)$ for all h, meaning D cannot hassle R into being a declining state. κ_R and κ_D are R's and D's costs from going to war (respectively). If D goes to war at step (4), then the game terminates, and players receive their first period war payoffs (based on likelihood $P(1, 0, 0)$) twice to account for both rounds. If R goes to war in step (7), then players receive their first period settlement payoff and second period war payoffs based on the transgressions and hassling that happen.

When war does not happen in step (4), the selected transgression and hassling levels shape future war payoffs. However, these are costly actions. We assume R incurs costs $K_R(t)$ (increasing, with $K_R(0) = 0$) from transgressions, and D incurs costs $K_D(h)$ (increasing, continuous, with $K_D(0) = 0$) from hassling. Consistent with the interpretation of transgressions and hassling as being low-level operations, we assume the costs of transgressions and

[10] As intuition, whoever wins has proposal power and can act without considering the opponent.

	R (type θ) payoff	D's payoff if R is θ
Round 1: War	$2\left(-(1-P(1,0,0))\theta-\kappa_R\right)-K_R(t)$	$2*(-P(1,0,0)\theta-\kappa_D)$
Round 1: Accept, Round 2: War	$-\lvert x_1-\theta\rvert-(1-P(2,t,h))\theta-\kappa_R-K_R(t)$	$-x_1-P(2,t,h)\theta-\kappa_D-K_D(h)$
D and R accept both rounds	$-\lvert x_1-\theta\rvert-\lvert x_2-\theta\rvert-K_R(t)$	$-x_1-x_2-K_D(h)$

Table 1: D's and R's payoffs. Note that these payoffs vary depending on R's true type θ, which is private information.

hassling are less than the costs of fighting a war ($K_R(\bar{t}) < \kappa_R$ and $K_D(\bar{h}) < \kappa_D$).

3.3 Equilibrium Preliminaries and Previewing the Analysis

Unless otherwise specified, we consider pure-strategy perfect Bayesian Nash equilibria (referred to as "equilibria" below).

Throughout the analysis, we will assume that the "**destabilizing rise assumption**" holds. This is the following.

Assumption 1. *The* ***destabilizing rise assumption*** *holds whenever*

$$-2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right] > -\mathbb{E}\left[max\left\{P(2,0,0)\theta-\kappa_R,0\right\}\right]$$

Informally, this assumption describes cases that are high stakes to D—cases where R is rising quickly and may be quite revisionist.[11] We limit our analysis to these cases because they describe instances where, if hassling or transgressions were off-the-table, the outcome would be a preventive war (Fearon 1995; Debs and Monteiro 2014). While it would be possible to

[11]Formally, the left-hand-side is D's expected payoff from going to war in the first round. The right-hand-side is D's expected payoff from not going to war when R does not transgress ($t = 0$), D does not hassle ($h = 0$), and D sets $x_1 = 0$ in the first round and $x_2 = max\left\{P(2,0,0)\theta-\kappa_R,0\right\}$ in the second round (all of which we assume will not result in war). Note that because x_2 must be weakly greater than zero, $x_2 = max\left\{P(2,0,0)\theta-\kappa_R,0\right\}$.

relax this assumption in the analysis below, doing so would complicate the equilibria statements (for example by allowing for fully peaceful bargaining to take place) while detracting from the case of interest where D is sufficiently concerned about R's rise.

Additionally, we assume that the "**ambiguity assumption**" holds. This is the following.

Assumption 2. *The **ambiguity assumption** holds whenever*

$$-2\bar{\theta}P(1,0,0) + max\left\{P(2,0,\bar{h}) - \kappa_R, 0\right\} \geq -2\kappa_R$$

Without this assumption, R may have some incentive to fight in the first round for some offers, which would result in R revealing their type without engaging in transgressions.[12] This condition similarly keeps the focus on the substantive case of interest: where revisionist types of R want to keep their type hidden until after they rise. Without this assumption, it is possible that some types of R would consider fighting in the first round, and that, in some cases, D may need to make policy concessions to R in the first round. Relaxing this assumption and attaining similar results is still possible, but the statement of equilibria conditions would be much more complicated.

As a final assumption, we must assign some structure to equilibrium selection. Below, when D enables R, multiple kinds of equilibria can exist as is common in the signaling setting. To address this in the simplest way possible, we assume that the **"selection into signaling"** assumption holds. This is the following:

Assumption 3. *The **selection into signaling assumption** implies that whenever a signaling equilibrium exists alongside other kinds of equilibria, the signaling equilibrium will be*

[12]Formally, the assumption holds whenever R's expected utility from accepting an offer of $x_1 = 0$ in the first round is greater than going to war in the first round, conditional on D selecting the greatest possible hassling level and the game playing out as characterized in the equilibrium below. This assumption is equivalent to $-2\bar{\theta} + max\left\{P(2,0,\bar{h}) - \kappa_R, 0\right\} \geq 2\left(-(1 - P(1,0,0))\bar{\theta} - \kappa_R\right)$.

selected into.

Because under some parameters multiple equilibria can exist, we assume that, whenever it exists, a signaling equilibrium will be selected into. This assumption is best explained in-depth later, after the equilibria are introduced.

4 Subgame Analysis and Equilibrium

The primary question of interest is when would D ever give a potentially adversarial rising state the means to rise more aggressively. To answer this question, we first analyze the two subgames determined by D's choice whether to enable R of not, then derive the conditions when D would prefer to enable R.

Throughout this analysis, it is useful to reference how a straightforward, specific example operates through the various game forms and equilibrium considerations. This will be a game with two types of C, $\theta \in \{\underline{\theta}, \bar{\theta}\}$ with $\underline{\theta} < \bar{\theta}$, and where probabilities of these two types are denoted as $Pr(\theta = \underline{\theta})$ and $1 - Pr(\theta = \underline{\theta})$ (respectively). Additionally, in the straightforward example, we assume $\mathcal{H} = \{0, \underline{h}, \bar{h}\}$, with $0 < \underline{h} < \bar{h}$, meaning D chooses between not hassling ($h = 0$), a low-level of hassling ($h = \underline{h}$) and a higher level of hassling $h = \bar{h}$). While stark, this example can illustrate the game's dynamics and intuitions simply, and will be referenced below as the "straightforward example."

4.1 Subgame When D Does Not Enable R

Here we analyze the subgame where D chooses not to enable R, meaning that R lacks the option to transgress. To preview behavior within this subgame, there are two types of equilibria: an equilibrium with hassling and offers being made and accepted in both periods, and an equilibrium where no hassling occurs and D goes to war in the first round.

We describe the intuition here by working backwards. If the game enters into period 2, the

game becomes a complete-information ultimatum bargaining game. At this point, D can always either set a policy position that gives R their expected utility from going to war, or, if R is not particularly revisionist, D can set $x_2 = 0$ and R will not fight. In response to this offer that is tailored to R's type, R will always accept. Because R is rising and is potentially revisionist, in expectation, this second period payoff could be quite bad for D. In response, in period 1, D may take one of two approaches. First, similar to other models with commitment problems (Fearon 1995; Debs and Monteiro 2014), D may simply want to go to a preventive war in period 1 and resolve the issue when they are strong rather than let the game proceed to the period 2 when they are weak. Second, similar to Schram, D may choose to hassle to slow R's rise, which can serve as an alternate to preventive war. Note that D will always want to do something, either hassling or going to war, rather than letting R rise peacefully—this is guaranteed by the destabilizing rise assumption. Also note that if D hassles in period 1, then D is able to make R an offer of $x_1 = 0$ and all types of R will accept—this is implied by the ambiguity assumption.

To summarize, we define the following types of equilibria as characterized by their on-the-path behavior.

Definition 1. In a **hassling equilibrium**, D will hassle and the game will proceed without war. In the first period, D sets $x_1 = 0$, $w_D = 0$, and $h = h^h$, where h^h is D's optimal level of hassling. In the second period, D sets $x_2 = max\{P(2, t, h)\theta - \kappa_R, 0\}$, and R accepts.[13]

Definition 2. In a **preventive war equilibrium**, D will go to war in the first period, and whoever wins the war with set the policy position equal to their ideal point.

With these equilibria defined, we can characterize when the subgame ends in one kind of behavior or another.

[13]Formally, this is $h^h \in argmax_{h \in \mathcal{H}}\{\mathbb{E}[-max\{P(2, 0, h)\theta - \kappa_R, 0\} - K_D(h)]\}$. Note that h^h could take on multiple values, but this would not change the results below.

Proposition 1. *Suppose the destabilizing rise and ambiguity assumptions hold, and that D chooses not to enable R.*

(a) A hassling equilibrium (with h^h as characterized above) exists if and only if the following holds:

$$\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta-\kappa_R,0\right\}\right]-K_D(h^h)\geq -2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right] \tag{1}$$

(b) A war equilibrium exists if and only if the following holds:

$$\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta-\kappa_R,0\right\}\right]-K_D(h^h)<-2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right] \tag{2}$$

Without R's option to transgress, D's choice is straightforward. Because R pools on setting $t=0$, there is no separation across the types and D must choose a general approach—either hassling or going to war. When D's payoff from the hassling equilibrium ($\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta-\kappa_R,0\right\}\right]-K_D(h^h)$) exceeds their payoff from the preventive war equilibrium ($-2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right]$), D will hassle. Otherwise, D will go to a preventive war.

Whenever the hassling equilibrium is selected into, it constitutes an ex-ante Pareto improvement over the preventive war equilibrium outcome. However, due to private information, the hassling equilibrium has its own inefficiencies. Consider the straightforward example where type $\underline{\theta}$ is quite moderate and type $\bar{\theta}$ is quite revisionist. In this game form, D responds to R's rise by implementing a fixed hassling level. In contrast, if R's type were revealed, D might prefer a greater hassling level in response to more-revisionist types of R ($\theta=\bar{\theta}$) and a lower hassling level in response to more moderate types of R ($\theta=\underline{\theta}$). Instead, because D does not know R's type, D must implement an ex-ante less-efficient one-size-fits-all approach to managing R's rise. This changes, however, when R can transgress.

4.2 Subgame when D Does Enable R

When D enables R, several types of equilibria can exist. We focus our attention here on a new kind of equilibrium that can be productive to D—an equilibrium where some types of R transgress and others do not, and where D tailors their hassling level to R's transgression.[14] We will refer to this as the signalling-and-hassling equilibrium.

Definition 3. In a **signalling with hassling equilibrium**, in the first period, moderate R's will not transgress, revisionist R's will transgress, and D will tailor their hassling response to R's actions.

The signalling-and-hassling equilibrium is characterized by a type-cutpoint θ', and two feasible hassling levels h^{s0} and $h^{s\bar{t}}$. In equilibrium, all types of R below (and including) θ' do not transgress ($t = 0$), and all types of R greater than θ' do transgress ($t = \bar{t}$), thus communicating their private preferences. For ease, we define $\Theta_{>\theta'}$ and $\Theta_{\leq\theta'}$ as all types of R that are strictly greater than cutpoint θ' or weakly less than cutpoint θ', respectively; broadly speaking, the set $\Theta_{\leq\theta'}$ includes moderate types of R, and the set $\Theta_{>\theta'}$ includes revisionist types of R. In response to R not transgressing, D selects hassling level h^{s0}, and, in response to R transgressing, D selects hassling level $h^{s\bar{t}}$, where $h^{s0} < h^{s\bar{t}}$. In both cases, D will set $x_1 = 0$, R will not go to war, and in the second period D will make a policy concession to R that will keep war from occurring.

Any equilibrium with signaling and hassling must be characterized by the values θ', h^{s0}, and $h^{s\bar{t}}$, which are defined in terms of one another. Put another way, this equilibrium must be characterized in terms of the triplet $(\theta', h^{s0}, h^{s\bar{t}})$, where the individual components relate to one another.

[14] We go into more detail on productivity below.

Definition 4. We characterize the triplet $(\theta', h^{s0}, h^{s\bar{t}})$ as having values satisfying

$$max\left\{P(2,0,h^{s0})\theta' - \kappa_R, 0\right\} = P(2,\bar{t},h^{s\bar{t}})\theta' - \kappa_R - K_R(\bar{t}), \tag{3}$$

$$h^{s0} \in argmax_{h\in\mathcal{H}}\left\{\mathbb{E}\left[-max\left\{P(2,0,h)\theta - \kappa_R, 0\right\}|\theta \in \Theta_{\leq\theta'}\right] - K_D(h)\right\} \tag{4}$$

and

$$h^{s\bar{t}} \in argmax_{h\in\mathcal{H}}\left\{\mathbb{E}\left[-P(2,\bar{t},h)\theta + \kappa_R|\theta \in \Theta_{>\theta'}\right] - K_D(h)\right\}, \tag{5}$$

where $\theta' \in (0, \bar{\theta})$.

Equation 3 implies that some cutpoint-type θ' exists where all types below θ optimally set $t = 0$ and all types above optimally set $t = \bar{t}$. Note that this condition can hold in our straightforward example (with $\theta \in \{\underline{\theta}, \bar{\theta}\}$), so long that $\underline{\theta} < \theta'$ and $\bar{\theta} > \theta'$. Equation 4 and Equation 5 and define D's optimal hassling response to observing no transgression (h^{s0}) and a transgression ($h^{s\bar{t}}$), conditional on R playing the cutpoint strategy and the game proceeding as described above.

With this triplet defined, we can introduce the full existence conditions for the signaling with hassling equilibrium.

Proposition 2. *Suppose the destabilizing rise and ambiguity assumptions hold, and that D chooses not to enable R. A signaling with hassling equilibrium with values* $(\theta', h^{s0}, h^{s\bar{t}})$ *exists if and only if the following holds:*

$$-P(2,\bar{t},h^{s\bar{t}})\mathbb{E}\left[\theta|\theta \in \Theta_{>\theta'}\right] + \kappa_R - K_D(h^{s\bar{t}}) \geq -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta|\theta \in \Theta_{>\theta'}\right], \tag{6}$$

and

$$\mathbb{E}\left[-max\left\{P(2,0,h^{s0})\theta-\kappa_R,0\right\}|\theta\in\Theta_{\leq\theta'}\right]-K_D(h^{s0})\geq-2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta|\theta\in\Theta_{\leq\theta'}\right], \tag{7}$$

Equation 6 and Equation 7 imply that, in response to $t=\bar{t}$ and $t=0$ (respectively), D prefers implementing their optimal hassling level rather than going to war in the first round. We will discuss this signalling with hassling equilibrium in more length below.

4.3 Equilibrium

We can now characterize D's decision.

Proposition 3. *Suppose the destabilizing rise, ambiguity, and selection into signaling assumptions all hold. If a signaling with hassling equilibrium with values* $(\theta', h^{s0}, h^{s\bar{t}})$ *exists and is selected into, and either (a)*

$$(1-F(\theta'))\left[-P(2,\bar{t},h^{s\bar{t}})\mathbb{E}\left[\theta|\theta\in\Theta_{>\theta'}\right]+\kappa_R-K_D(h^{s\bar{t}})\right]+$$
$$F(\theta')\left[\mathbb{E}\left[-max\left\{P(2,0,h^{s0})\theta-\kappa_R,0\right\}|\theta\in\Theta_{\leq\theta'}\right]-K_D(h^{s0})\right]\geq\mathbb{E}\left[-max\left\{P(2,0,h^{h})\theta-\kappa_R,0\right\}\right]-K_D(h^{h}).$$

holds when $\mathbb{E}\left[-max\left\{P(2,0,h^{h})\theta-\kappa_R,0\right\}\right]-K_D(h^{h})\geq-2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right]$,

or (b)

$$(1-F(\theta'))\left[-P(2,\bar{t},h^{s\bar{t}})\mathbb{E}\left[\theta|\theta\in\Theta_{>\theta'}\right]+\kappa_R-K_D(h^{s\bar{t}})\right]+$$
$$F(\theta')\left[\mathbb{E}\left[-max\left\{P(2,0,h^{s0})\theta-\kappa_R,0\right\}|\theta\in\Theta_{\leq\theta'}\right]-K_D(h^{s0})\right]\geq2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right].$$

holds when $\mathbb{E}\left[-max\left\{P(2,0,h^{h})\theta-\kappa_R,0\right\}\right]-K_D(h^{h})<-2\kappa_D-2P(1,0,0)\mathbb{E}\left[\theta\right]$, *then D will enable R and the gameplay will follow the signaling with hassling equilibrium. Otherwise, D will not enable R, and the gameplay will follow that described in Proposition 1.*

We describe the importance of Proposition 3 below in the Results. Here we highlight the importance of the selection into signaling assumption. By assuming that whenever a hassling with signaling equilibrium exists it would be selected into, then this clarifies D's choice whether to enable R (or not). Practically speaking, if R does not select into the signaling with hassling equilibrium after being enabled, then D cannot gain from allowing R to transgress.[15] In this way, this assumption describes settings where D's choice to enable R is rooted in D's belief that enabling will result in R relaying useful information.

5 Results

5.1 Allowing R to Transgress may Improve Welfare

Empirically, sometimes actors open themselves to arrangements that an adversarial opponent could take advantage of. In our model, these building out of trade arrangements, peace treaties, or international institutions grant openings for adversaries in the same way that moving from the subgame without transgressions to the subgame with transgressions represents. To that end, we might wonder if these kinds of arrangements are productive. To address this, we are taking the most challenging interpretation of this observation: we are considering the case where the new arrangement only gives potential adversaries the ability to undertake costly actions to disrupt the balance of power, without any outside gains.[16] Even in this unfavorable setting, we find that the ability to transgress can be beneficial in aggregate.

Result 1. *Allowing R to transgress can lead to welfare gains, especially when R is more likely*

[15] Of the other kinds of pure strategy equilibria that exist, two of these do not involve R transgressing and resemble the preventive war and hassling equilibrium described above, making them not any more productive for D. One new type of equilibria all types of R transgress, and D responds by always hassling; note that this kind of equilibria can never be productive to D, so, if this was the only equilibrium that existed, then D would never enable R. These are described more fully in the Appendix.

[16] Put another way, here the new international arrangement offer no inherent gains (like we might expect trade or security agreements to do), but rather only afford adversaries an ability to hurt the defender.

moderate and when D can significantly scale back their hassling in response to moderate R's.

Formally, suppose the destabilizing rise, ambiguity, and selection into signaling assumptions all hold. If a signaling with hassling equilibrium with values $(\theta', h^{s0}, h^{s\bar{t}})$ *exists and is selected into, and either (a)*

$$-F(\theta')\left(K_D(h^{s0})\right) - (1 - F(\theta'))\left(K_R(\bar{t}) + K_D(h^{s\bar{t}})\right) \geq - K_D(h^h) \tag{8}$$

holds when $\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta - \kappa_R, 0\right\}\right] - K_D(h^h) \geq -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right]$,

or (b) whenever $\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta - \kappa_R, 0\right\}\right] - K_D(h^h) < -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right]$*, then enabling R is ex-ante welfare improving.*

The expression in (a) 8 is derived by comparing the combined (for D and R) expected utility from the signaling with hassling equilibrium (when D enables R) with the combined expected utility from the hassling equilibrium (where D does not enable R). We will discuss this expression here, but we first pose two caveats to Result 1. First, Result 1 ignores D's strategic choice whether to enable R or not, but rather is a comparison of the joint utilities in the equilibria described in Propositions 1 and 2. Second, this discussion should be viewed as a local analysis where the existence conditions in Proposition 2 holds. With those caveats, the signaling with hassling equilibrium offers gains over the hassling equilibrium when transgression costs are low ($K_R(\bar{t})$ is low), when R is more likely a moderate type ($F(\theta')$ is large), and when D selects a lower level of hassling in response to R not transgressing (a low $K_D(h^{s0})$).[17] Together, these all imply that the signaling equilibrium offers welfare gains when R is unlikely to transgress, and when D can respond with very low levels of hassling when R is moderate.

[17]We are unable to make clear statements on $h^{s\bar{t}}$ relative to h^hbecause we did not characterize $P(2,t,h)$ as additively separable.

The expression in (b) points out that if a signaling with hassling equilibrium exists when D enables R but only a preventive war equilibrium exists when D does not enable R, then enabling R can always lead to efficiency gains.

This result is meaningful because giving R the ability to conduct a costly and potentially provocative action can generate aggregate gains. This is surprising in this competitive setting because, *ceteris paribus*, if R transgresses, it produces a welfare loss. Instead, Result 1 arises due to strategic dynamics. In the game with transgressions, D learns that R is a moderate type when R does not transgress, whereas such learning does not occur in the game without transgressions. As the most straightforward mechanism, upon learning that R is moderate, D scales back its selected level of hassling relative to what D would have implemented in the game without transgressions. This strategic behavior lowers D's hassling costs, potentially to an extent that offsets that added costs from hostile Rs transgressing and D's more aggressive hassling when R is a revisionist type.

Before moving on from this result, we offer two points of clarification on Result 1. First, this welfare result can arise if transgressions make hassling less marginally effective. If this is the case, then D may optimally select a lower level of hassling in response to R's transgression ($h^{s\bar{t}} < h^{h}$). We might find this result intuitively unappealing; in many empirical cases, we might expect D to respond more aggressively to transgressions rather than less. We demonstrate in the Appendix that the welfare result can still hold when $h^{s\bar{t}} \geq h^{h}$. Second, readers may be concerned that Result 1 implies that, while there are aggregate utility gains, by allowing R to transgress R will do better and D will do worse. If this is the case, then it is not clear why D would join such institutional arrangement where they are putting themselves at risk of R transgressing.[18] Result 2 speaks to this concern.

[18] As an alternate explanation for why Result 1 still matters, realistically, states may not be able to predict their rise or decline, meaning that states enter into these institutional arrangements first and then later are prepared to take on the role of R or D; if this is the case, then optimizing welfare can be reasonable. This would require a different model, but the results here would still predict when enabling was welfare enhancing.

5.2 Allowing R to Transgress may Improve D's *Ex-Ante* Expected Payoffs

Could giving R the ability to hurt D ever benefit D? Surprisingly, yes.

Result 2. *Allowing R to transgress can give D higher ex-ante payoff in the game form with transgressions, especially when R's transgressions are limited in their effect at shifting the future balance of power, when R is more likely to be benign, and when D's one-size-fits-all hassling in the game without transgressions is particularly inefficient.*

Formally, suppose the destabilizing rise, ambiguity, and selection into signaling assumptions all hold. If a signaling with hassling equilibrium with values $(\theta', h^{s0}, h^{s\bar{t}})$ *exists and is selected into, and either (a)*

$$(1 - F(\theta'))\left[-P(2, \bar{t}, h^{s\bar{t}})\mathbb{E}\left[\theta|\theta \in \Theta_{>\theta'}\right] + \kappa_R - K_D(h^{s\bar{t}})\right] + \tag{9}$$
$$F(\theta')\left[\mathbb{E}\left[-max\left\{P(2,0,h^{s0})\theta - \kappa_R, 0\right\}|\theta \in \Theta_{\leq\theta'}\right] - K_D(h^{s0})\right] \geq \mathbb{E}\left[-max\left\{P(2,0,h^h)\theta - \kappa_R, 0\right\}\right] - K_D(h^h).$$

holds when $\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta - \kappa_R, 0\right\}\right] - K_D(h^h) \geq -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right]$*, or (b)*

$$(1 - F(\theta'))\left[-P(2, \bar{t}, h^{s\bar{t}})\mathbb{E}\left[\theta|\theta \in \Theta_{>\theta'}\right] + \kappa_R - K_D(h^{s\bar{t}})\right] + \tag{10}$$
$$F(\theta')\left[\mathbb{E}\left[-max\left\{P(2,0,h^{s0})\theta - \kappa_R, 0\right\}|\theta \in \Theta_{\leq\theta'}\right] - K_D(h^{s0})\right] \geq -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right].$$

holds when $\mathbb{E}\left[-max\left\{P(2,0,h^h)\theta - \kappa_R, 0\right\}\right] - K_D(h^h) < -2\kappa_D - 2P(1,0,0)\mathbb{E}\left[\theta\right]$*, then D optimally enables R.*

Equation 9 compares the utility from the signaling with hassling equilibrium (the left-hand-side) to the utility from the hassling equilibrium (the right-hand-side). While this is a dense expression, some comparative statics can be parsed out; as it was in the discussion on Result 1, admittedly this discussion should be treated as a local analysis where the existence

conditions in Proposition 2 still hold (unless otherwise specified).

First, this inequality is strengthened when R is more likely to be benign (larger $F(\theta')$) for two reasons. First, directly, if R is more likely to be moderate, then R will transgress with lower probability. Because R's transgressions directly lowers D's second round payoffs, D prefers when R transgresses less. Second, indirectly, when R is more likely to be moderate, D's hassling becomes more efficient. D incurs inefficiencies when R cannot transgress by hassling moderate Rs. This expression is strengthened when D will set a high level of hassling in response to the possibility of a very-revisionist R, even if this very-revisionist R is only realized with low probability. If this is the case, then D is over-hassling moderate Rs with high probability, and incurring costs from doing so. In contrast, in the game with transgressions where moderate Rs signal their type, D can do better by selecting a lower, less costly level of hassling.[19]

Second, this inequality is strengthened when R's transgressions do not significantly alter the future balance of power (formally, when $P(2, \bar{t}, h^{s\bar{t}} - P(2, 0, h^{s0}$ is small). When transgressions generate larger power shifts, this leads to a direct effect, where D simply does worse now that R can transgress, and an indirect effect, where D (potentially) incurs greater costs implementing higher levels of hassling to moderate the transgression. Note that, while the inequality is strengthened when $\bar{t}$ only moderately affects P, we also consider the alternate constraints here as well. For Equation 3 to hold, it must be that revisionist types of R benefit enough from transgressing; without this feature then revisionist Rs cannot gain by transgressing, and the signaling equilibrium cannot exist. Together, for D to do better by enabling R, transgressions must be productive enough for R to consider transgressing, but not so productive that it is highly detrimental to D.

Third, this inequality is strengthened whenever D can hassle more efficiently when information is revealed. Most simply, if D does not change their hassling behavior in response

[19]Note that this logic of D hassling more efficiently would also hold if R was more likely to be revisionist, but D would do worse in this setting because R would transgress with greater likelihood.

to the new information (that R is revisionist or moderate), then the inequality can never hold.[20] Similar to the discussion in Result 1, D can lower their hassling costs in the game with transgressions whenever R does not transgress, and D can hassle revisionist types of R more whenever R does transgress. Essentially, the more D can tailor their hassling response as they gain information about R's type, the better D will do when R is able to transgress.

(Draft Note: pending discussion on second inequality)

This result is substantively meaningful. When these conditions hold, D is willing to grant adversaries the ability to harm them in the future, as, ex-ante, this can produce better results. While ex-post the decision to give R the ability to hurt D may look problematic—sometimes R does transgress, D engages in higher levels of hassling, and D does worse—D may have undertaken these actions because D knew there was a chance R was moderate, which could have lead to greater efficiency.

5.3 For Enabling R to be Beneficial, D Must Be Able to (Broadly) Hassle

Pending, here's the logic.

If D cannot hassle, then D will always want to go to war with R. There's no way that R's transgressions can ever convey useful information, as revisionist types would want to mimic non-revisionist types if it somehow resulted in D not going to war with them. Essentially, no separating equilibrium can be supported.

If D can only select one hassling level, then D would never want to enable R. This point is more difficult, so it's useful to walk through this as an example. Suppose D can only select one positive level of hassling (hassling level $h = h'$ or no hassling $h = 0$). Suppose D enables R, and R behaves based on some cutpoint where types below don't transgress and types

[20] If $h^{s\bar{t}} = h^{s0} = h^h$, then R's transgression insures the right-hand-side is larger.

above do. If this occurs, then by transgressing, R would incur hassling level h', which would be sufficient to keep D from wanting to go to war with this rising, revisionist type. However, by not transgressing, this rising revisionist type could incur no hassling, which ultimately would be more beneficial (they would rise to their un-hassled level, which D would have preferred to go to war with). Essentially, if D can only select one level of hassling, the only equilibrium that can exist is one where D hassles all the time, and R must always transgress (i.e. it must be that R always gains from it). Because this is strictly worse for D than an equilibrium where R cannot transgress, D optimally does not enable R in this situation.

This is interesting because it suggests, specifically, multiple hassling options are needed to sustain this kind of equilibrium (more tools are useful/needed for transgressions to work).

5.4 Allowing R to Transgress Could Generate Fully Peaceful Outcomes

Until this point, we have discussed pure strategy equilibria. One subtlety of the equilibrium in Proposition 2 is that even if R reveals themselves to be fairly moderate by not transgressing, D must still implement a non-zero level of hassling. This begs the question: if D can hassle and D enables R—the subgame where actors have the most tools for engaging in different levels of conflict (low-level or war)—is it possible that no forms of conflict will occur and the game will be fully peaceful? In the appendix, we define a new kind of mixed strategy equilibrium with this kind of behavior; we summarize the key feature of this equilibrium here in the text.

Result 3. *Consider a setting with a rising revisionist state. In settings where R can transgress and D can hassle, a fully-peaceful equilibrium can exist where, with positive probability, R does not transgress and D does not hassle.*

Formally, Suppose the destabilizing rise, ambiguity, and selection into signaling assumptions

all hold. If D enables R, a mixed-strategy perfect Bayesian Nash equilibrium can exist with the following on-the-path behavior. First, moderate R's do not transgress and revisionist R's mix between not transgressing and transgressing. In response to R's not transgressing, D mixes between setting some positive hassling level and setting some zero hassling level. In response to R's transgression, D always hassles R.

What makes this result substantively meaningful is that, in the pure strategy equilibria discussed above (Proposition 3), either D went to war or D hassled R—in other words, no equilibrium was ever fully peaceful.[21] To this end, allowing for both transgressions and hassling together is meaningful: the availability of hassling and transgressions opened the possibility for some communication to occur, for D to learn about R's type, and, sometimes, for the actors to avoid undertaking costly actions altogether.

6 Discussion

The overarching finding of our models is that when rising and declining states have a realistic range of policy options that affect the size of a power shift, credible signals of the rising state's intentions become possible that mitigate, or even eliminate, incentives for preventive conflict. This novel result stems from the interaction of the declining state's hassling and the rising state's transgressions, which facilitates credible reassurance. Transgressions capture the two mechanisms identified in the literature by which power shifts are endogenous to the rising state's decisions: acquisition of assets that confer power and internal investment in coercive capabilities. The possibility of such endogenous power shifts facilitates credible signals of the rising state's intentions because only hostile risers that seek revision at the decliner's expense seek to expand their power.[22] Benign risers that share the decliner's preferences

[21] As intuition, if R is not transgressing, then by not hassling D does strictly better going to war (as the destabilizing rise condition specifies).

[22] In reality, of course, even benign risers might want to increase their power for other reasons, e.g., to deal with third-party threats or to achieve goals on issues of low salience to the decliner. Nevertheless, to the extent that the rising state has aims at odds with the decliner's, the more the riser should want to enhance

have no need to pay the costs of doing so, because they are relatively satisfied with the status quo. This allows the declining state to update its beliefs in response to the riser's transgression decision and adopt a more optimal response, increasing its degree of hassling toward risers that have transgressed and moderating its degree of hassling toward those that have not. Indeed, under some conditions, the declining state may not impose any hassling *at all* in response to the rising state's restraint. This occurs when low-level hassling is inefficient relative to high-level hassling as a means of curtailing the riser's power gains, which prompts the decliner to probabilistically mix between no hassling and an efficient higher degree of hassling in response to restraint. Thus, the endogenous sources of a power shift help to counteract the commitment problems engendered by an exogenous power shift and facilitate peaceful power transitions between states with compatible goals.

Importantly, the possibility of hassling by the declining state is necessary for the rising state to use transgressions to signal its type and avert war. This is because if the declining state's only recourse in response to transgressing were preventive war, even hostile risers would not transgress. The decliner's preventive response would wipe away the power gains from doing so, eliminating the benefits of transgression to risers of any type.[23] Thus, without the possibility of hassling, no transgression ever occurs, eliminating it as an effective signaling mechanism: most hostile types would be deterred by the artificially credible threat of preventive war into pooling with benign risers.[24] But again, although the *possibility* of hassling is necessary for our signaling mechanism, hassling may not occur in equilibrium: as shown above, the declining state sometimes responds to restraint with no hassling at all. Thus, contrary to the conventional wisdom in IR that the proliferation and employment of

its own power, making its decision of whether to do so an informative signal of its intentions.

[23]This dynamic is consistent across the two forms of transgressions we examine. With transgressions-as-revisions, the direct benefit of successful transgression to hostile risers in terms of acquiring the asset at stake would likewise be vitiated by preventive war, along with the accompanying power gains.

[24]The only risers that would attempt transgression are those so dissatisfied with the status quo that they prefer full-scale war in the present to waiting for the exogenous power shift to progress so that they can achieve their goals under a more favorable distribution of power in the future. But these highly-revisionist types would prefer to reveal themselves and fight under any circumstances, so such signals have nothing to do with the availability of transgressions.

low-level coercive tactics are destabilizing, we show that they can play an important role in facilitating credible reassurance, helping to resolve security dilemmas and forestall preventive war.

The informational benefits that arise from endogenous power shifts carry the surprising implication that it may be in the declining state's interest to enable the rising state to conduct transgressions. By creating opportunities for the rising state to augment its power, the decliner simultaneously generates opportunities for the riser *not* to do so, showing restraint that signals its benign intentions and yields more efficient outcomes. This could manifest in a wide range of concrete policies. Most obviously, asymmetric economic arrangements that enhance the rising state's economic growth enables future investment in its military capability and the cultivation of economic dependence of other states that the riser can wield for coercive leverage. But declining states may also derive informational benefits from, *inter alia*, leaving strategically valuable territory vulnerable to conquest, allowing the diffusion of militarily sensitive technology, and the acquiescence to "vulnerability dependence" on the rising state for crucial economic and military inputs. Such strategies are, of course, risky for the decliner: if the rising state actually harbors hostile intentions, the decliner is empowering its future rival. In equilibrium, we find that this situation leads to more intense and costly hassling to contain a hostile riser later on. But under certain conditions – namely, when the decliner holds relatively confident priors that the riser is benign, the efficacy of hassling is relatively high, and the decliner's default hassling strategy is relatively inefficient – the upside of potentially identifying a benign riser and reducing its degree of hassling makes enabling an endogenous power shift a worthy gamble for the decliner. In other words, by "putting itself out there" – leaving itself vulnerable to a hostile rising state that would exploit such opportunities to enhance its own military power – the decliner can glean valuable information about the riser's intentions that can forestall preventive conflict and enhance the decliner's welfare, on balance.

The positive informational effects of endogenous power shifts are a novel contribution to the literatures on credible reassurance and on shifting power and preventive conflict. Most of the literature concerns exogenous power shifts exclusively, and nearly all find that preventive war occurs either under complete (Gilpin 1981; Kim and Morrow 1992; Powell 2006; Tammen et al. 2000) or incomplete information (Powell 1996, 1999; Copeland 2000; Wolford, Reiter and Carrubba 2011; Shifrinson 2018). Likewise, previous works that have identified hassling as a key means by which states can forestall preventive war (Schram 2021) and/or elicit credible signals of rising states' intentions (Yoder 2019; Spaniel and Poznansky 2020; Haynes and Yoder 2024) assume purely exogenous power shifts. These signaling mechanisms are distinct from ours.[25] Hassling previously has been identified as a means by which declining states can disincentivize misrepresentation by hostile rising states: by inflicting a degree of unconditional punishment even in response to cooperation, the decliner incentivizes hostile rising states to pursue their true goals instead of cooperating dishonestly. This generates a semi-separating equilibrium that makes continued cooperation a more credible signal of benign intent, and forestalls further escalation to preventive war. In contrast, our signaling mechanism is not brought about by hassling directly, but rather by the riser's costly restraint in foregoing the opportunity to enhance its power. Hassling merely facilitates this mechanism by giving the declining state an intermediate response to rising power short of preventive war that does not eliminate incentives for hostile risers to transgress, thereby allowing restraint to distinguish the riser's benign intentions.

The literature on endogenous power shifts has, conversely, omitted exogenous sources of shifting power, and thus has not examined how endogenous shifts facilitate signals that can forestall preventive war that otherwise would have occurred. The only model we know of that has identified the rising state's restraint in the context of an endogenous power shift as a credible signal in itself is Gurantz and Hirsch (2017), which departs from ours in

[25] Our models reproduce the finding of Schram (2021) that hassling, by reducing the size of the power shift, ameliorates the commitment problem and makes it rational for the declining state to refrain from further escalation to full-scale war. This is captured in proposition 1.

several ways. Like other existing models, their power shift is purely endogenous to the riser's revision. This exclusion of the exogenous power shift means that there is no incentive for preventive war in the absence of revision, making the information from the riser's restraint inconsequential for the decliner's behavior. Furthermore, under the standard bargaining assumption that war is inefficient, their framework would leave no incentive for even hostile senders to ever attempt revision, because (as discussed above) it would simply be wiped out by the receiver's preventive action. Gurantz and Hirsch (2017) achieve a separating equilibrium only by assuming the existence of a war-loving type of sender that is willing to transgress even though doing so leads to war under the *ex ante* distribution of power. In short, when coupled with the declining state's hassling options and an exogenous shift, the signaling mechanisms that arise from endogenous power shifts become both more potent – applying under a more realistic range of conditions – and more consequential for forestalling war.

Our findings also offer a more nuanced picture than prior models of the conditions under which endogenous power shifts cause war. Most models of endogenous power shifts find that they rarely result in war, primarily because the declining state can deter the rising state from increasing its power and thereby eliminate the source of the commitment problem (Fearon 1996; Schwarz and Sonin 2008; Chadefaux 2011).[26] To the contrary, however, Debs and Monteiro (2014) show that this is not the case when the rising state's transgression is only probabilistically observable to the declining state. Under such conditions, a decliner may initiate mistaken preventive war with benign rising states that have not actually invested in their coercive capabilities.

We corroborate Debs and Monteiro (2014)'s finding that incomplete information over the

[26]In these models, endogenous power shifts yield peaceful bargains in an infinite-horizon setting, leading to a gradual stream of peaceful concessions over time. However, they do not identify endogenous power as a signaling mechanism, and generally assume complete information. Kydd and McManus (2017) show that even in an incomplete-information setting endogenously shifting power does not lead to war when rising states can signal limited aims through "tied-hands" reassurances, but these signals are unrelated to the rising state's choices regarding its acquisition of power.

riser's investment decision can produce war – indeed, our mechanism by which endogenous power shifts facilitate credible signals operates only when the riser's transgressions are sufficiently observable. However, we depart from Debs and Monteiro in that we find more restrictive conditions under which mistaken preventive war occurs than they do, despite our inclusion of a concurrent exogenous power shift. Whereas Debs and Monteiro assume that investment decisions are observable with some exogenous probability, we model the observability of such decisions as itself an endogenous choice of the rising state. As Coe and Vaynman (2020) point out, the observability of an arms buildup is itself an asset over which states can bargain, by alternately demanding and then accepting or rejecting monitoring provisions of an arms control regime. Thus, declining states are in a systematically more information-rich environment than Debs and Monteiro assume. Either the rising state accepts rigorous monitoring, such that its investment decision is completely observable, or it rejects monitoring, which then itself serves as a signal that the rising state is likely arming. The probative value of that signal depends, as Coe and Vaynman (2020) argue, on how costly the verification regime would be to the rising state's government in terms of its political survival: the more damaging the information would be, the less likely even a benign rising state would be to accept arms control, and thus the less informative rejection is. Thus, we concur with Debs and Monteiro that endogenous power shifts can cause mistaken preventive war under incomplete information, but specify a more demanding set of conditions necessary for this equilibrium to hold.

Finally, our finding that the capacity for the rising state to endogenously augment its own power benefits the declining state runs counter to a large "offensive realist" literature that contends that states must constantly compete for power in order to maintain their security (Mearsheimer 2001; Layne 2006; Rosato 2015). On the contrary, we show that promoting a rising rival's ability to enhance its power can function as a screening mechanism that increases the declining state's security. This argument differs from previous critiques of offensive realism, which have emphasized that competition for power is often suboptimal

due to the availability of information that allows benign states to identify each other with reasonable confidence (Glaser 2010; Kydd 1997, 2005). Whereas these arguments hold that this information arises from exogenous signaling mechanisms, we show that it can actually arise endogenously from the signal sent by the rising state's decision of whether or not to attempt to bolster its power.

This last result helps make sense of the pattern of US responses to China's rise since the end of the Cold War. In the 1990s and early 2000s the US adopted an "engagement" strategy that facilitated China's capacity for subsequent revision and military investment, which was criticised as naive by some at the time, and by many more in retrospect (Layne 2009; Mearsheimer 2010; Friedberg 2011). Conversely, the US has since 2009 increasingly moved toward greater containment of China, beginning with Obama's "rebalance," escalating to an outright trade war under Trump, and continuing with further export and investment restrictions under Biden. This competitive shift has been widely decried as unwarranted, counterproductive and irrational, particularly as China has attempted remarkably few concrete revisions to the international order in that time (Johnston 2019; Li 2021; Christensen 2015; Yoder 2020).

Our argument helps to reconcile these seemingly incompatible strategies. Although the US certainly took a risk by facilitating China's economic rise and embracing a high degree of US economic dependence on China, it also opened up space for China to signal its intentions through its degree of investment in its coercive capabilities. Our argument shows that, even setting aside the economic benefits of engagement, this strategy could be rational *ex ante* because of the upside that it might allow the US to confidently identify a potentially-benign China and avoid inefficient competition altogether. But in the event, China signalled the opposite. Even granting that it has not yet attempted extensive revision, China has clearly increased its capacity to do so by its massive military investments, economic statecraft toward its neighbors (e.g., through BRI), acquisition of overseas bases, and territorial

expansion in the South China Sea, among other activities. Our mechanism shows why these investments in coercive power are *themselves* credible signals of China's hostile intent (from a US perspective), and thus why increasing bipartisan pessimism and the recent trend in favor of greater US hassling toward China may be warranted.

The contrast with the US response to other rising powers in the 1980s and 1990s – namely, Japan and Germany – is instructive, and further supports our argument. As it did toward China, the US empowered Japanese and German economic recovery and rapid ascendance after WWII, and even encouraged German reunification and the formation of an EU with Germany at its center. During this time Germany and Japan studiously avoided any hint of remilitarization, even in the high-threat environment both countries faced at the height of the Cold War. These policies persisted even after the end of the Cold War, eliminating the possibility that common communist threats induced them to align with the US. This sustained restraint by Germany and Japan from investing in coercive capabilities was a highly credible signal that they had no revisionist ambitions against the US-led order, and allowed the US to confidently maintain largely unfettered cooperation with both countries indefinitely. Thus, in facilitating Japanese and German economic growth, the US gave both states the capacity to increase their military power, and by extension the means to credibly signal their compatible goals by exhibiting restraint. In contrast to China, toward whom this gamble has apparently not paid off and resulted in a higher degree of hassling *ex post*, US engagement with Germany and Japan was rewarded with a much lower degree of costly hassling than would have been warranted toward rising states whose intentions were highly uncertain.

7 Extensions

7.1 When R's Transgressions are Probabalistically Observed

Under construction.

7.2 When R's Transgressions Alter the Immediate Political Settlement

Under construction.

8 Conclusion

We have investigated shifting power under incomplete information where both the rising and declining state can affect the size of the shift through their actions. This specification is both novel and straightforwardly realistic. Rising states choose whether or not to invest in their coercive capabilities or increase their power through external expansion or revision, while declining states have a range of hassling options short of war to mitigate the rise of a potential competitor. Yet no prior models have given both actors this range of options.

Under these broadly realistic conditions, we find that power shifts are substantially less conflict-prone than previous scholarship has implied. Most fundamentally, we show that the possibility of an endogenous power shift generates a signaling mechanism that can overcome the pernicious effects of an exogenous power shift. When the rising state has the capacity to enhance its own power, it can credibly signal non-revisionist intentions by refraining from doing so. This signaling mechanism is facilitated by the declining state's hassling options. Since the decliner can respond to the riser's transgressions with containment measures short of preventive war, hostile rising states retain incentives to transgress in an effort to boost their power gains. In equilibrium, war does not occur, and the rising state's signals of

restraint mitigate, or in some cases eliminate altogether, costly competition between rising and declining states with compatible preferences. We show that these results can hold even when the rising state's transgressions (military investments) are not directly observable, because the possibility of arms control monitoring arrangements makes the riser's choice to conceal its behavior informative in itself.

Even more strikingly, we showed that the informational benefits that arise from the rising state's control over its own power trajectory can make it rational for a declining state to abet the riser's ability to transgress. Giving the riser the means to increase its coercive capabilities serves as a screening mechanism that allows the decliner to distinguish benign types from hostile ones. Although this strategy is risky – it creates the possibility that a hostile riser will choose to enhance its own power, forcing the decliner into more costly hassling and larger future concessions – the upside of identifying and avoiding inefficient competition with benign risers sometimes makes it a worthy tradeoff.

References

Bils, Peter and William Spaniel. 2017. "Policy Bargaining and Militarized Conflict." *Journal of Theoretical Politics* 29(4):647–678.

Chadefaux, Thomas. 2011. "Bargaining Over Power: When do Shifts in Power Lead to War?" *International Theory* 3(2):228–253.

Christensen, Thomas. 2015. *The China Challenge: Shaping the Choices of a Rising Power*. New York: W.W. Norton & Co.

Coe, Andrew and Jane Vaynman. 2020. "Why Arms Control Is so Rare." *American Political Science Review* 114:342–355.

Copeland, Dale C. 2000. *The Origins of Major War*. Ithaca, NY: Cornell University Press.

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

Edelstein, David. 2002. "Managing Uncertainty: Beliefs about Intentions and the Rise of Great Powers." *Security Studies* 12(1):1–40.

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

Fearon, James D. 1996. "Bargaining over Objects that Influence Future Bargaining Power." Annual Meeting of the American Political Science Association.

Friedberg, Aaron L. 2011. *A Contest for Supremacy: China, America, and the Stuggle for Mastery in Asia*. New York: W.W. Norton & Company.

Gilpin, Robert. 1981. *War and Change in World Politics*. Cambridge, UK: Cambridge University Press.

Glaser, Charles L. 2010. *Rational Theory of International Politics: The Logic of Competition and Cooperation*. Princeton, NJ: Princeton University Press.

Gowa, Joanne. 1995. *Allies, Adversaries, and International Trade*. Princeton, NJ: Princeton University Press.

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

Haynes, Kyle and Brandon K. Yoder. 2020. "Offsetting Uncertainty: Reassurance with Two-Sided Incomplete Information." *American Journal of Political Science* 64(1):38–51.

Haynes, Kyle and Brandon K. Yoder. 2024. "Trust, cooperation, and the tradeoffs of reciprocity." *Conflict Management and Peace Science* 41(1):26–46.

He, Kai. 2008. "Institutional Balancing and International Relations Theory: Economic Interdependence and Balance of Power Strategies in Southeast Asia." *European Journal of International Relations* 14:790–808.

Jervis, Robert. 1978. "Cooperation Under the Security Dilemma." *World Politics* 30(2):167–214.

Johnston, Alastair Iain. 2019. "The Failures of the 'Failure of Engagement' with China." *The Washington Quarterly* 42(2):99–114.

Kim, Woosang and James Morrow. 1992. "When do Power Shifts Lead to War?" *American Journal of Political Science* 36(4):896–922.

Kydd, Andrew. 1997. "Sheep in Sheep's Clothing: Why Security Seekers Do Not Fight Each Other." *Security Studies* 7(1):114–155.

Kydd, Andrew H. 2005. *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.

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

Lascurettes, Kyle M. 2020. *Orders of Exclusion*. New York: Oxford University Press.

Layne, Christopher. 1993. "The Unipolar Illusion: Why New Great Powers Will Rise." *International Security* 17(4):5–51.

Layne, Christopher. 2006. "The Unipolar Illusion Revisited: The Coming End of the United States' Unipolar Moment." *International Security* 31(2):7–41.
URL: http://dx.doi.org/10.1162/isec.2006.31.2.7

Layne, Christopher. 2009. "US Hegemony in a Unipolar World: Here to Stay or Sic Transit Gloria?" *International Studies Review* 11(4):784–787.
URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2486.2009.00902.x/abstract

Lektzian, David J. and Christopher M. Sprecher. 2007. "Sanctions, Signals, and Militarized Conflict." *American Journal of Political Science* 51(2):415–431.

Levy, Jack S. 1987. "Declining Power and the Preventive Motivation for War." *World Politics* 40(1):82–107.

Li, Cheng. 2021. *Middle Class Shanghai: Reshaping U.S.-China Engagement*. Washington, D.C.: Brookings.

Lonergan, Erica D. and Shawn W. Lonergan. 2022. "Cyber Operations, Accommodative Signaling, and the De-escalation of International Crises." *Security Studies* 31(1):32–64.

Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.

Mearsheimer, John. 2001. *The Tragedy of Great Power Politics*. New York: W.W. Norton and Co.

Mearsheimer, John. 2010. "The Gathering Storm: China's Challenge to U.S. Power in Asia." *Chinese Journal of International Politics* 3(4):381–396.

Miller, Manjari Chatterjee. 2021. *Why nations rise: Narratives and the path to great power*. Oxford.

Montgomery, Evan Braden. 2006. "Breaking Out of the Security Dilemma: Realism, Reassurance, and the Problem of Uncertainty." *International Security* 31(2):151–185.

Powell, Robert. 1996. "Uncertainty, Shifting Power, and Appeasement." *American Political Science Review* 90(4):749–764.

Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics*. Princeton, NJ: Princeton University Press.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60(01):169–203.

Reich, Noam. 2022. "Signaling strength with handicaps." *Journal of Conflict Resolution* 66(7-8):1481–1513.

Rosato, Sebastian. 2015. "The Inscrutable Intentions of Great Powers." *International Security* 39(3):48–88.

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

Schram, Peter. 2022. "When capabilities backfire: How improved hassling capabilities produce worse outcomes." *The Journal of Politics* 84(4):2246–2260.

Schwarz, Michael and Konstantin Sonin. 2008. "A Theory of Brinkmanship, Conflicts, and Commitments." *The Journal of Law, Economics, and Organization* 24(1):163–183.

Schwebach, Valerie. 2000. Sanctions as Signals: A Line in the Sand or a Lack of Resolve. In *Sanctions as Economic Statecraft: Theory and Practice*, ed. Steve Chan and Cooper Drury. New York: Palgrave pp. 869–895.

Shifrinson, Joshua. 2018. *Rising Titans, Falling Giants: How Great Powers Exploit Power Shifts*. Ithaca, NY: Cornell University Press.

Spaniel, William and Michael Poznansky. 2018. "Credible Commitment in Covert Affairs." *European Journal of International Relations* 62:668–681.

Spaniel, William and Michael Poznansky. 2020. "Bad Faith Cooperation." *International Interactions* 46(4):579–605.

Tammen, Ronald L., Jacek Kugler, Douglas Lemke, Carole Alsharabati, Brian Efird and A.F.K. Organski. 2000. *Power Transitions: Strategies For the 21st Century*. Seven Bridges Press, LLC / Chatham House.

Wolford, Scott, Dan Reiter and Clifford Carrubba. 2011. "Information, Commitment, and War." *The Journal of Conflict Resolution* 55(4):556–579.

Yoder, Brandon. 2020. "How Credible are China's Foreign Policy Signals? IR Theory and the Debate about China's Intentions." *Chinese Journal of International Politics* 13:577–605.

Yoder, Brandon K. 2019. "Hedging for Better Bets: Power Shifts, Credible Signals, and Preventive Conflict." *Journal of Conflict Resolution* 63(4):923–949.