

Territorial Disputes in the Shadow of Law: Legal Claims and the Use of Force

Furkan Y Benliogullari
UC San Diego

Abstract

International law is thought to help pacify territorial disputes by giving disputants and third parties a focal point for peaceful resolution. If a disputant violates international law, third parties shall punish the violation and enforce the law. I propose a conceptual clarification that third-party enforcement is possible only when the violation is clear. A challenger using force with a strong legal claim presents a contestable case that is harder for third parties to agree on its categorization as a violation and mobilize to punish. Use of force with a strong legal claim does not excuse the illegal act, but the strong claim mitigates the anticipated response, and only the challengers with sufficient military capability can benefit from it. On an original panel of 4,539 dispute-years across 162 territorial disputes from 1945 to 2014, with territorial militarized disputes hand-matched to the disputes that produced them, I find that a challenger's legal claim emboldens a powerful challenger that can win a military confrontation easily and restrains a challenger that cannot. A strong claim on the target's side helps third parties mobilize and punish the challenger, which deters the likelihood of force. This paper shows that law does not only give states a reason to peacefully settle. It can also tell them what they can expect to get away with.

Keywords: territorial disputes, legal claims, international law, use of force, enforcement

1. Introduction

For more than eighty years, Iran pressed a claim to the Tunbs in the Strait of Hormuz, through protests, bilateral settlement attempts, and an attempt to put the question before the League of Nations. The claim over the islands was strong. Britain, which held them through the Trucial sheikhdoms, refused to submit the question to third-party settlement, and it was under no obligation to accept it. When Britain announced its withdrawal from the Gulf in 1968, the Shah stopped pressing for settlement. Britain, now unwilling to defend the islands, spent a year trying to broker terms. Ras al-Khaimah met the British negotiator repeatedly and would neither cede the Tunbs nor accept payment for them, as it was entitled to refuse (Mobley 2003). On 30 November 1971, the last day of Britain's obligation to defend the Tunbs, Iranian forces landed and took them, losing three marines and killing four Ras al-Khaimah policemen. If international law and strong legal claims facilitate peaceful resolution, why do challengers with a strong legal claim use force in territorial disputes?

It is clear why a disputant with a strong legal claim wants to use the tools international law provides: they help it avoid costly military conflict and resolve the dispute on terms its claim supports. It is not clear, however, why their opponent would participate in the process. States are obligated to settle their disputes peacefully, but they are nowhere obligated to settle them. The Charter leaves the choice of means to the parties (UN Charter Art. 2(3) and Art. 33(1)), and the jurisdiction of the International Court of Justice depends on their consent (ICJ Statute Art. 36). A target, in possession of the territory and shielded by the prohibition on the use of force, can refuse to engage in talks with the challenger. Many targets do participate, and as a result, I find that 62 percent of all territorial disputes since 1945 have been settled peacefully. However, nothing legally compels participation, and where it is withheld, the challenger is left with two options: hold to the status quo indefinitely and wait, or use force and face whatever response that provokes.

Third-party actors are not oblivious to this impasse. The challenger's revisionist actions that cause escalation invite third-party response for violating the prohibition (Zacher 2001; Fazal 2007). Still, the nature of the impasse and the disputants' legal claims are usually publicly known (Huth et al. 2012). Third-party actors may collectively enforce the prohibition if they can agree on the violation (Thompson 2009; Johns 2012; O'Mahoney 2018), but a challenger's stronger claim makes the violation contestable and the coalition needed to punish the use of force harder and costlier to assemble. This mitigates possible reaction to its violation.

Huth, Croco, and Appel (2011, 2012, 2013) established that strong legal claims can help a challenger reach a favorable settlement and, in doing so, make it less likely to use force, provided that the target agrees to settle. I argue that a challenger that uses force to take the disputed territory is not asking for support. It wants them to fail to punish by using its strong legal claim to make the case contestable and harder for third

parties to classify and act on together. In the absence of the target's consent, a strong legal claim therefore lowers the cost of acting alone and can embolden a militarily capable challenger to use force.

Three predictions follow. First, a challenger's legal claim should show little effect on the use of force on its own. Having a strong legal claim alone does not make a challenger capable of using force or compel a target to make concessions. Second, relative to a weak legal claim, the challenger's strong legal claim should strengthen its hand as its military capability increases. It should pacify the challenger when it cannot easily win a fight because the claim is worth more at the table than it is as mitigation after the fact, and it should embolden the challenger to militarize the dispute when it can. With a strong legal claim mitigating third-party reaction relative to a weak claim, an increase in military power can make use of force more attractive than peaceful means of resolution that come with certain costs attached. Third, the stronger the target's legal claim, the less likely the challenger is to use force. Relative to a weak legal claim, the target's strong legal claim makes the use of force more costly and less likely for the challenger, thereby serving as a deterrent. Although the target's claim is not independent of the challenger's claim, they are not perfectly symmetrical, which requires an independent prediction about its effect.¹

I test these predictions on a panel of territorial disputes between 1945 and 2014. I hand-collected all territorial MIDs from Gibler's (2018) MID narratives and MID v5 (Palmer et al. 2022) narratives and matched them with the disputes from Huth and Allee (2002) and the legal-claim variables of Huth, Croco, and Appel (2011), in an extended panel of 4,539 dispute-years across 162 disputes, with force used in 12.8 percent of dispute-years. I find support for all three predictions. First, the challenger's legal claim does not have any independent effect on the use of force on its own. Second, the effect of the interaction between the challenger's legal claim and the power scale is significant and substantial in the weighted models. It has two opposite effects at two points. Where the challenger is powerful enough to fight but not to win cheaply, a stronger claim restrains it, and the annual probability of force falls from 16.7 percent for a weak claim to 11.6 percent for a mixed one and 7.7 percent for a strong one. Where the challenger is decisively more powerful, the same claim emboldens it, and the probability rises from 7.1 percent to 12.5 percent and 20.6 percent. The effect is graded rather than a threshold: a mixed claim falls between a weak and a strong one at every level of capability. Third, the deterrent effect of the target's legal claim considerably lowers the annual probability of force. At the dispute-year level, the probability goes from 15 percent at the weak level down to 12.9 percent at the mixed and 11 percent at the strong level of the target's legal claim.

This paper makes three contributions. First, it shows that the mechanism by which law pacifies territorial disputes also has a permissive branch that can embolden a challenger and cause escalations. Huth, Croco, and Appel establish the pacifying effect of law and build their theory on the idea that third parties

¹ The challenger's and target's legal claims correlate at -0.56 at the dispute level. In seven disputes, neither side has a good legal claim. These descriptive statistics show that the coding is not symmetrical.

read the merits of a dispute and that their reaction varies with those merits. They trace that reaction to the support a legally advantaged state can expect when force is used against it. The same premise carries a second implication that they do not pursue and at one point reject. For a challenger that uses force to take territory it claims, what varies is not the support it receives but the punishment it does not. I show that this second branch is also in operation. A strong legal claim makes force more likely for a challenger able to act on it, and less likely for one that is not.

Second, it generalizes the results of two arguments on court decisions to the system of territorial disputes. Johns (2012) shows that the main function of international court decisions is to coordinate enforcement. Gent and Shannon (2014) show that the effect of a legal ruling on dispute behavior depends on the balance of power between the parties. My findings suggest that neither mechanism requires a court as long as the legal positions are clear, and the great majority of territorial disputes never reach one. This paper shows both operating on the underlying disputes across the full post-1945 population.

Third, in this paper, I expanded the dataset on territorial disputes collected by Huth and Allee (2002) and Huth et al. (2011) and present a unique dataset of territorial MIDs matched with specific territorial disputes. My hand-collected territorial MID variable (dependent variable) recovers 163 MID pairs that the MID project (Palmer et al. 2022) records but does not code as territorial. In addition, I added several new disputes to the dispute universe of Huth and Allee (2002), extended the temporal scope, and added dispute-ending events, including agreements, adjudication or arbitration decisions, and claim withdrawals. The dispute dataset now runs up until 2025, but the estimation panel stops at 2014 because the panel I used for this paper is limited by MID availability. Also, the findings are not dependent on my additions. Every model is reported across three MID groups, sourced from different sources. The findings are weakest on my own MIDs and stronger in the MIDs from MID v5 and corrected MIDs (Gibler et al. 2016).

2. Why International Law Does Not Settle Territorial Disputes

As an instrument of the post-war world, international law and its constituent norms fundamentally changed interstate disputes. States do not declare wars and instead find other legal formulations to justify their use of force (Irajpanah and Schultz 2021). Major territorial aggrandizements are harder to achieve and sustain (Zacher 2001; Fazal 2007; Atzili 2012; Goertz et al. 2016). As a result, states adapt and find other ways to seize territory from others, such as smaller conquests by fait accompli (Tarar 2016; Altman 2017, 2020). Multiple other accounts (e.g. Schultz 2014; Goertz et al. 2016; Hathaway and Shapiro 2017) reinforce the position that the post-war norms that are a part of the current legal structure have been a pacifying force in interstate relations.

A closer reading of the norms that regulate state behavior in territorial disputes exposes multiple fault lines. First, the customary principles that regulate territorial ownership have two departments that pull in

opposite directions: the historical consolidation of title through prolonged and open possession, along with the acquiescence of others, and the prohibition on acquiring title by force (Jennings [1963] 2017). In the post-war legal order, it is no longer permissible to acquire territory by force, and force does not confer title. Yet when a state in possession takes administrative action over a territory and demonstrates its sovereignty, its title grows stronger. Jennings ([1963] 2017) finds these factors as forces pulling in opposite directions, and neither has priority over the other. A state that takes long-claimed territory can present the act in either register: as an acquisition by force, which the Charter forbids, or as the completion of a title already in the process of consolidation, which the law of title has always contemplated. International law supplies no rule for choosing between them.

The doctrine does limit what consolidation can achieve. The public manifestation of a dispute — crystallization — sets a critical date, after which actions to consolidate title confer no further legal advantage; only what came before it counts (Sharma 1997). Ironically, by the time states realize they are in a race, it is too late to compete, and their later actions can only preserve and press the title they have already built (Kohen and Hebie 2011). But this limitation constrains how title is acquired, not how a seizure is characterized. The ambiguity between the two departments remains and produces the genuine disagreement that is hard to manufacture. This is a property of the legal system rather than of anyone's rhetoric.

A good example of this problem is the Iran-UAE dispute over the Tunb (Tonb) islands. Iran's claims go back to the 18th century through local tribes under Persian authority (CIA Research Paper 1980). In the late 19th century, Iran's ruling dynasty, the Qajars, wanted to consolidate its hold over the ports on the Persian side and the Tunb islands managed by the port cities. The Iranian government arrested the Sheik of Lengeh and replaced him with an appointee. Britain objected to this appointment on the ground that the ports and islands were under the authority of not only the Sheik of Lengeh of Qasimi origin but also the Qasimi tribe that was located on the other side of the Gulf (Nazarahari 2005). Britain controlled Oman and the Trucial States, where the Qasimi tribe lived, claimed these islands through the Sheiks of Sharjah and Ras al-Khaimah, and later occupied them. Iran attempted to settle the dispute bilaterally or through the League of Nations, to no avail (Nazarahari 2005).

In 1971, when Britain was withdrawing from the region, Iran used force to occupy these islands. Iran characterized it as the resumption of sovereignty it had never surrendered. The newly formed UAE characterized it as an acquisition by force. Both characterizations reached the UN Security Council within days. Algeria, Iraq, Libya, and the People's Democratic Republic of Yemen requested an urgent meeting (S/10409); Iraq invoked Article 2(4) and Kuwait the inadmissibility of territorial acquisition by force, while Iran told the Council it "had no alternative but to establish the exercise of its sovereign rights." The Council chose between neither. It deferred consideration so that third-party efforts could proceed, adopted no resolution, and took no vote (S/PV.1610, 9 December 1971). The dispute had crystallized decades earlier,

in Iran's protests and its attempts at settlement, so the possession established in 1971 added nothing to Iran's legal position. What made the act hard to classify was a historical record that supported both readings.

The second fault line is the impasse in dispute resolution created by a target that is not under any international legal obligation to resolve it. In the post-war legal order, states are no longer permitted to use force to resolve their disputes (Orakhelashvili 2022). Since the adoption of the UN Charter, a treaty law that binds all members of the United Nations, the principle of territorial integrity and the prohibition on the use of force to resolve interstate disputes have become established parts of international law. Article 2(4) of the UN Charter prohibits the use of force and solidifies the territorial integrity norm. Article 2(3) obligates states to solve their disputes peacefully.

The prohibition on the use of force is not an unclear rule. Franck (1990) treats Article 2(4) as a case of determinacy: it says what it means, and states know when they have breached it. But determinacy is not the same as legitimacy. A rule can be perfectly clear and still lose its claim on those it binds when mechanical application produces results the rule's own purpose cannot justify. Franck's remedy is process-determinacy: a credible interpreter with authority to apply a general rule to particular circumstances, so that clarity does not become rigidity. This interpreter is the International Court of Justice (ICJ) or other legal dispute-resolution mechanisms.

This remedy may be structurally unavailable in territorial disputes. The ICJ has jurisdiction over a dispute if the disputing parties consent to it (ICJ Statute Art. 36). A target in possession of the disputed territory is asked to consent to a process that may take the territory away from it, and it is under no obligation to agree. Article 2(3) only obligates peaceful resolution; it does not establish an obligation to resolve (Pellet 2012). The Charter does not empower the Security Council or the General Assembly to compel resolution either. The UNSC may recommend terms (Art. 37(2)) but cannot impose them. Overall, the duty of peaceful settlement does not compel the target, whereas the prohibition on the use of force binds the challenger. Consequently, international law favors the targets in possession of the disputed territory.

Lack of resolution alone does not affect the strength of a challenger's claim, but waiting indefinitely is certainly not costless. A challenger can preserve its claims through protests, but the factual situation on the ground may change over time. A target can take actions to consolidate its position (Carter 2010). Generational change and the consolidation of the existing boundary reduce the political and economic salience of a historical claim (Carter and Goemans 2011; Abramson and Carter 2016). Over time, evidence ages, and the target's increasing control of the territory may make it harder to find guiding facts. The target can effectively strengthen the shield provided by law on other fronts.

This leaves the challenger with a bitter choice: force, which the law forbids; settlement, which requires the consent of the party benefiting from the delay; and maintaining the status quo by waiting, which is lawful but is slowly becoming costly without any promise of return. A rule that produces this result in

disputes where the claimant has the better title is precisely the case in which Franck expects legitimacy — and with it, the willingness of others to enforce — to weaken.

3. Theory: Logic and Function of Legal Claims in Territorial Disputes Revisited

The institutionalist focus has primarily been on the settlement route: which disputes reach negotiation, which reach a court, and what makes agreements hold (e.g. Simmons 2002, 2005; Mitchell and Hensel 2007; Powell and Wiegand 2010; Huth et al. 2011, 2013; Mattes 2018), but not all challengers' stories end in that route. Some challengers accept the cost and use force. Although the use of force violates international law, their legal claims to ownership do not disappear. The claims continue to cast a shadow over what the challenger expects from third parties, and it is that expectation, formed before the decision is taken, that this section develops.

a. Legal Claims and Enforcement of International Law

In deciding whether to use force, a challenger anticipates how others will respond. The legal order has no central enforcer, and international law is enforced, when it is enforced at all, by states acting on their own account (Guzman 2008). Johns (2012) shows how enforcement can work even where no central mechanism can compel anyone: a legal signal coordinates enforcement by states that have no stake in the dispute itself, and they act because a rule they may need later is worth maintaining now. Some of these enforcers may be the target of other territorial disputes (Prorok and Huth 2015). By upholding the norms, these states reap the benefits of their enforcer status (Thompson 2009). In territorial disputes, the primary sanction that operates is non-recognition. It withholds legal effect from the seizure and denies the seizing state the benefit it acted for (Zacher 2001; Fazal 2007; Hathaway and Shapiro 2017; O'Mahoney 2018). Non-recognition is cheap for any one state and worthless unless enough of them do it together.

Enforcement of this kind has two stages, and the second one is conditional on the first. First, third parties must agree on whether the act violates the prohibition on the use of force. Then, they can decide whether and how to punish the act. The first step, however, is far from automatic in territorial disputes. As the previous section showed, a state that takes a long-claimed territory can present the act as an acquisition by force or as the completion of a title already consolidating, and the doctrine supplies no rule for choosing between them. Where non-recognition has failed to materialize, this failure has commonly come at this first stage rather than the second. It is not because states were unwilling to punish a violation, but because they could not agree that one had occurred (O'Mahoney 2018).

The strength of the challenger's legal claim has its main effect in the first stage. A claim with no basis leaves nothing to argue about. The act is an illegal seizure, a conquest, and it is more readily classified as such. A claim with genuine foundations makes the case contestable, and the political cost of backing a sanction rises when a rule's application is unclear, or the violator's conduct is ambiguous (Thompson 2009).

As that cost rises across enough states, the coalition now has to pay higher enforcement costs, and the one required to make non-recognition bite may not materialize or may become more costly.

Failure to enforce does not mean the legal claims are a legal defense to the charge of violating international law. Nothing in international law excuses a violation of Article 2(4), and no third party is invited to weigh the merits of the underlying title before deciding whether the prohibition was breached. The challenger should be culpable for using force. With stronger legal claims, what varies is the consequence: punishing a challenger would defend another legally weak position adopted by the target, and the challenger's action can be categorized as the exercise of legal ownership by some actors. This disagreement itself weakens the coalition needed to impose the punishment. This is mitigation as a property of broader enforcement structure, not as an individual act of third-party judgment or adjustment to the response.

A challenger can anticipate this because it can read its own claim. Both the claims and the law are public. The ICJ jurisprudence that governs the disputes is commonly known (Sumner 2004). The materials that make a claim strong (treaties, colonial administrative lines, prior arbitral and judicial rulings, and long patterns of administration) are matters of public record, and states' own internal assessments of their legal position track outside assessments closely.²

b. Legal Claims Facilitating Peaceful Settlement, Restated

The value of a strong legal claim on the challenger's or target's side in the settlement route is well established. Huth, Croco, and Appel (2011, 2013) argue that international law supplies a focal point that helps disputants overcome the distribution problem caused by the difficulty of choosing among the many ways a disputed territory might be divided. Their 2011 argument rests on two characteristics of law. First, it is established by treaties, customs, and court rulings and commonly known among states. Second, it supplies a common basis for assessing the relative merits of competing claims. Building on these characteristics, Huth et al. claim that a focal point emerges when two conditions hold together: the relevant principles must be clear and well-established, and the parties' claims must be asymmetric.

They find strong support for the pacifying focal point argument. Challengers with strong legal claims are 58% more likely to seek negotiations, and disputants with a clear legal asymmetry are 57% more likely to reach a settlement (Huth et al. 2011). Furthermore, the settlement terms track the law as the focal point argument predicts: a challenger with strong legal claims is 389% more likely to secure favorable terms (Huth et al. 2013). A state with a good claim that reaches the table can expect to leave it with something.

² Huth, Croco, and Appel (2011) built their coding on the strength of a challenger's and target's legal claim from public expert sources and court rulings. Internal British archival assessments matched their coding in more than 90 percent of the cases.

These findings show that, for a challenger with a strong claim, force is not the only option. In fact, the challenger with a strong legal claim has a set of choices, and peaceful choices have low costs and high returns. This is the pacifying face of law in territorial disputes.

The account that Huth, Croco, and Appel provide also assigns third parties a role in the challenger's decision. Among the reasons a legally advantaged challenger should prefer negotiation to force, they invoke the post-1945 prohibition on acquiring title by force and observe that such aggression "might be countered by other states with sanctions or even direct military action," besides undermining the bargaining leverage the claim would otherwise supply (2011, 421). Although this account is consistent with the conceptual framework that I constructed, it is incomplete.

Huth, Croco, and Appel (2012) take the argument further in their work on escalation. There, a state with a strong legal claim escalates with confidence, knowing that third parties will support it because its claim tells the international community which disputant is acting lawfully. A state with weak claims expects the opposite: that others may align against it instead. Anticipated third-party reaction varies with the strength of the claim in their account, which is the premise I build on.

I have two points of disagreement on the mechanism of escalation. In the 2012 paper, they limit the logic of military escalation to the scenario in which force is used against the disputant with a strong legal claim. Only then can the disputant use force with impunity. Here, the strength of legal claims shapes the third-party support. It is not clear to me why the same mechanism of variation in third-party reactions would not apply to punishment of the side that initially uses force. If it is the expected support from third-party actors that shapes the military response to the initial use of force, then the expected punishment can, in turn, shape the initial use of force, and, prior to that, the legal claims can shape that expected punishment.

The second point concerns the channel of third-party reaction in relation to the challenger's legal claims regarding the use of force. Building on the focal point argument, Huth et al. expect a disputant with a strong legal claim to be supported by third parties if the other disputant uses force. A challenger that is using force or responding to force with a strong legal claim is unlikely to expect this support. The challenger, whose claim causes the dispute in the first place, would want them to stay out of it rather than mobilize to punish its use of force. International law does not impose an obligation on third parties to support a challenger to get the territory it claims; it imposes an obligation to punish the challenger's use of force to settle the dispute.

This conceptual point undermines the bidirectional expectation generated by the focal point hypothesis, under which both the challenger and the target should expect support once attacked. Their own results are consistent with my objection. Huth et al. (2012) find no significant effect of legal claims on challenger escalation and attribute the null to the small number of legally advantaged challengers that reach that stage.

Both their explanation and mine fit the findings, and because I observe the full population of militarized disputes rather than only those that reach escalation, the two can be distinguished here in this paper.

I build on this to argue that the third-party reaction that a challenger anticipates is not support but punishment; that the likelihood of that punishment falls as the challenger's own claim grows stronger, because a contestable case is harder for third parties to classify and coordinate against; and that this anticipation operates when the challenger decides whether to use force, not only after force has been used against it.

Neither I nor Huth, Croco, and Appel (2011, 2012, 2013) test the third-party side of the story. It is beyond the scope of this paper, which focuses on disputants' behavior. Still, the way enforcement falls affects disputants' behavior. A challenger looking down the decision tree anticipates third parties' reactions, and the theory would be incomplete without the enforcement logic I offer here as the rationale for expecting the pattern.³ In my conceptual framework and argument, I extend my theory beyond the expectation generated by the prohibition on the use of force to investigate the effects of legal claims when the challenger violates the prohibition.

c. Military Capabilities and Legal Claims

I have presented two accounts of legal claims in a territorial dispute: the one that facilitates peaceful resolution and the one that drives conflict and leads to the use of force. These accounts can both be true under different conditions. A claim itself is not decisive in the resolution of disputes. If it were, all disputes with a strong legal claim would end the same way, but they do not. Legal claims are arguments based on international law, and they can be held in reserve or invoked by disputants when they depart from the status quo in an attempt to acquire territory. Legal claims are stable across time. Therefore, it is hard to explain the challenger's changing behavior with them alone.

When a challenger wants to take advantage of its legal claims and use force, another important consideration is military capability. For a challenger that is much less powerful and cannot use force, its legal claim will not be helpful. Without any risk of escalation, the target is under no obligation or pressure to agree to any settlement. This challenger is likely to maintain the status quo not as a choice but as a condition in which it finds itself. My theory predicts, and this is the prediction that distinguishes it, that the strength of a legal claim should not do much in this region, and that the peacefulness we observe among weak challengers is a fact about their capabilities rather than a fact about law.

As the challenger becomes militarily powerful enough to fight but not to win cheaply and decisively, its legal claims will begin to shape its behavior. Now, it can use its claim at the table to resolve the dispute through peaceful means or mitigate any third-party reaction that might follow the use of force. If it uses

³ In my dissertation, I investigate the effect of legal claims on third-party interventions in territorial disputes.

force with a strong or mixed legal claim, then it expects the third-party punishment to be lower than it would be for a challenger using force with a weak legal claim. Still, fighting is costly in this part of the power spectrum, and the two uses are not worth the same. At the table, a strong claim can deliver the territory itself, and on terms the law supports. The option to use force delivers only a lower expected punishment on a military contest the challenger must still win. A challenger that can fight but cannot win cheaply is therefore trading a large and probable return for a discount on a costly gamble, and I expect its stronger legal claim to pacify it relative to a challenger with a weaker claim.

From the target's perspective, this level of power distribution is where it must start taking the challenger's strong claims seriously, given that the challenger is capable of using force and can do so with a lower expected level of punishment due to a coordination failure among third parties. Faced with strong legal claims and wavering outside support, the cost of fighting may appear higher to the target and, thus, make a peaceful settlement less costly. Indeed, targets are more likely to give up a dispute when the challenger holds a strong legal claim (Schultz and Goemans 2019). Also, many scholars show that territorial concessions require a delicate approach and are not easy to make because they are very salient domestically (Huth 1996; Goddard 2006; Allee and Huth 2006; Mattes 2018). Therefore, it is hard to say that all disputes in this area of the challenger's power will end with settlements, and any claim about the peaceful settlements is beyond the scope of this paper. The point I need is narrower than a prediction about settlement. A capable challenger with a strong claim has a realistic expectation of resolution, and a capable challenger with a weak claim does not.

When we reach a point where the challenger can confidently win a military confrontation, its calculations change again. Most challengers at this level of advantage never have to fight, since targets are considerably more likely to give up a dispute outright when they face a much more powerful opponent (Huth 1996; Schultz and Goemans 2019). What separates the overwhelmingly powerful challengers that use force from those that do not is not the military advantage, which they all have, but the claim. Force is now a low-cost option compared with a lengthy settlement or dispute-resolution process that carries political costs and the risk of losing in court, and a strong claim leaves the act contestable enough that the coalition required to make non-recognition bite does not assemble. The value of acting keeps rising with capability while the value of waiting stays capped at the territory, and the cost of waiting adds up. Instead of waiting for the target to hear its claims and cooperate, the challenger can act without anyone's consent.

On the other side of the dispute, the target has its legal claim that impacts third parties' enforcement. Huth, Croco, and Appel (2011, 2012, 2013) very often combine these claims in the category of asymmetric legal claims when the directionality is not important. In this category, the observations with the target's strong legal claim are more numerous, and they are an important part of the story. Although the target's

claim is connected with the challenger's claim, they do not exist in a linear spectrum.⁴ In some cases, the target can have a strong legal claim while the challenger's claim is mixed (e.g., one based on customary principles) or weak (only political, ethnic, or historical claims). That is why it is important to theorize the effect of the target's claims along with the challenger's claim. I also build on this point and incorporate it into my theory, and the logic mirrors the logic of the challenger's claim in the shadow of third-party enforcement.

When a target has a strong legal claim, it removes the ambiguity that the challenger relies on. Where the record identifies the target as the rightful owner, the challenger's seizure is hard to present as the completion of a title already in the making, and the two departments of the law of title stop pulling in opposite directions. Consequently, characterizing the challenger's actions and coordinating enforcement around it become easier.

This has a consequence worth stating in advance, because here, it departs from the logic I used for the challenger's claim. The challenger's capability determines whether force is available and what it is worth. The military capability has no bearing on how third parties characterize the act once force has been used. The target's claim operates entirely at the points where force is used. I therefore expect a strong target claim to reduce the likelihood of force independent of either the challenger's or target's military power, and I expect no interaction between the target's claim and the challenger's capability. This is one of the places in the theory where my prediction is a null.

d. Hypotheses

This paper's theory is about legal claims, with military capability as the condition that determines what a claim does. Capability enters every model I estimate, but I do not test it as a hypothesis in its own right. The four hypotheses are symmetric: each side's legal claim appears once on its own and once in interaction with the challenger's capability, so that all four possibilities are stated and tested rather than only the ones the theory expects to find.

I established that the challenger's legal claim affects its choices to maintain the status quo, pursue talks for peaceful resolution, or use force in interaction with its military power. Because the challenger's claim works in opposite directions at different levels of capability, I expect its effect to disappear when it is not conditioned on capability. The first hypothesis is therefore a prediction of no or minor effect.

Hypothesis 1: The strength of a challenger's legal claim, on its own, does not change the likelihood of force in disputes.

⁴ In my extended dataset built on Huth, Croco, and Appel (2011), challengers have strong legal claims in 16 disputes and the target in 50. The challenger's and target's legal claims correlate at -0.56 at the dispute level. In 7 disputes, neither side has a good claim. These descriptive numbers show that they are not linear in coding. The most important symmetry is that there is no dispute in which both sides have a strong legal claim.

Although the hypothesis on the independent effect of the challenger's legal claim is straightforward, it is complicated to formulate the hypothesis on its interaction with power. The expectations change at different levels of military power. Still, the main intuition is two-fold. First, a stronger claim should restrain the challenger when its military advantage is not decisive. Second, the same claim should embolden it once the advantage gets decisive.

Hypothesis 2: The effect of a challenger's legal claim on the likelihood of force depends on its military capability. Where the challenger is far less powerful, its stronger claim has no effect. Where it is closer to or at power parity, a stronger claim restrains it. Where it is much more powerful, a stronger claim emboldens it.

The hypotheses about the target's legal claim are also straightforward, as I do not expect its effect to be conditional on the military power. Independent of the power balance between the challenger and the target, the target's strong legal claim helps third parties come to a clear conclusion about the nature of the use of force and solve their coordination problems. The main variation in the effect of the target's legal claim should come from the variation between its categories.

Hypothesis 3: A stronger legal claim on the target's side reduces the likelihood of force.

Hypothesis 4: The effect of the target's legal claim does not depend on the balance of military power.

e. What Else Could Produce This Pattern?

A finding that force is less likely where the target holds a strong claim does not by itself require anything I have argued. Two accounts predict it for other reasons. The first is coordination. Carter and Goemans (2011) show that borders drawn along inherited administrative frontiers are markedly less likely to be disputed and, once disputed, markedly less likely to become violent, because a familiar line is focal, cheap to administer, and expensive to redraw. Their variable is the precedent behind the boundary rather than anyone's claim, but precedent of exactly this kind (colonial delimitation, *uti possidetis*, and prior adjudication) is among the material that makes a legal claim strong in Huth, Croco, and Appel's coding. The two overlap by construction. The second account is Huth, Croco, and Appel's own, in which better title raises the reputational and audience costs of attacking it.

What separates both from my argument is not the sign but the form. Each makes law an additive cost. The coordinative value of a familiar line does not rise or fall with the capabilities of the states on either side of it, and Carter and Goemans's design has no place for capability at all: no measure of it appears in their models, and great power involvement, the nearest thing they include, does nothing. A reputational penalty for attacking good title is likewise a penalty at every level of capability. On my account, military capability decides whether a claim can be converted into anything, so the claim's effect must move with capability,

and on the challenger's side it must change sign. The interaction is therefore not a refinement of the additive story. It is what separates entitlement from constraint, and it is why the second hypothesis rather than the first carries the argument.

Carter (2010) is a third account that provides a close rival explanation with an interactive component. He shows that targets consolidate their hold on disputed territory by establishing garrisons, developing their resources, or passing legislation to incorporate it. Challengers facing a consolidated target escalate less often. What he calls consolidation, the law calls *effectivités*. Because *effectivités* help make a title strong, his variable and mine overlap wherever a target's claim rests on its own conduct. His explanation and mine differ in which interaction they require. On Carter's evidence, consolidation is a strategic response. Targets consolidate when the territory is worth the effort, and they consolidate more as the challenger they face grows stronger. His deterrent effect should therefore be largest where the challenger is most capable. Mine should not vary with capability at all, because how a seizure is characterized and how third parties respond do not depend on the military capabilities. My fourth hypothesis is where our accounts diverge.

Although not directly related to the use of force in territorial disputes, Goddard (2006) provides the fourth account, offering predictions about the resolution of territorial disputes through a theory focused on the target's side. On her account, territory becomes indivisible when actors legitimize their claims in ways that leave them unable to recognize an opponent's demand as legitimate. A target with unimpeachable title and an audience that knows it should then be the one least able to concede. What locks an actor in is not the quality of its claim but the narrowness of its resonance. Legitimation mobilizes some coalitions and severs ties with others. As an actor comes to depend on fewer coalitions, it loses flexibility until, at the extreme point, it holds a single legitimate position and must treat every other claim as illegitimate. The main issue in her account is the form that claims take. A claim resting on treaty text, prior adjudication, and customary principle is the broad register rather than the narrow one, since it is common knowledge among states and supplies a shared basis for comparing competing claims. Her lock-in mechanism describes a specific type of claims. They are relational and not primarily based on international law. In this form, they cannot inform third-party actors and shape their response.

Fifth, Bell (2017) tests the effect of territorial claims in interaction with military power on the probability of MID initiation. He finds that a militarized dispute is more likely when the state that does not control the disputed territory possesses a greater share of military capabilities. This specification closely resembles mine with important distinctions. First, he compares dyads with territorial claims and ones without claims. This design cannot explain the variation within the sample of territorial disputes that I am exploring. Second, his design imposes a linear model on military capabilities that does not account for the changing effect of the variable at multiple points. My design solves both problems by exploring the

variation in the sample of territorial disputes and using a non-linear model to explore the effect of power in interaction with legal claims on the MID outcomes.

Finally, Gent and Shannon (2014) provide an account of the interaction of military power and the court rulings. They find that court rulings track power: states carry out the rulings when their results reflect the balance of power between the disputants. Although this interaction inspired my design, mine is an extension of the design into the entire pool of disputes that do not go to any court and possibly trigger enforcement in the sense that Johns (2012) proposes.

4. Research Design and Data Analysis

I test my hypotheses through a number of logistic regression models and other quantitative tests using my extended dataset built on Huth and Allee (2002) dispute universe and the legal claims coded by Huth et al. (2011) with a temporal scope from 1945 to 2014. In this section, I explain my data collection and extensions, my independent and dependent variables along with controls and model choices in this study.

I extended the temporal scope of Huth and Allee (2002) and Huth et al. (2011) following their own rule for the eligibility of a dispute to be considered a territorial one. A dispute exists when officials of one state claim territory held by another state, and the holder rejects the claim. The state pressing the claim is the challenger, and the state in possession is the target.

I built on the legal claims dataset of Huth et al. (2011) instead of the ICOW dataset for a number of reasons. First, legality coding is not available off-the-shelf, and ICOW does not provide any beyond the salience variable. This dataset provides two separate legality variables for the challenger and the target. This formulation allows me to test my theory more precisely. Second, legal claims dataset of Huth et al. (2011) is built on the dataset of Huth and Allee (2002), which provides narratives about the disputed territory. Using and building on these narratives, I extended it until 2014. Without this foundation, an extension would be more arduous, with narrative building from scratch for this purpose.

My unit of analysis is the dispute-year. Because a dispute can lapse and revive rather than run continuously, I organize the data into episodes, each covering a period of continuous activity within a dispute, and the panel records one row per episode-year. The full panel holds 4,844 dispute-years across 253 episodes and 168 disputes. The estimation sample holds 4,539 dispute-years across 162 disputes, after dropping observations where the power measure is unavailable (102 rows) and the Polity scores needed for joint democracy are missing (203 rows). Because a single pair of states can host multiple disputes or generate a new dispute by taking over territory and becoming the target, standard errors are clustered on groups of disputes that share the same disputed territory, so that a territory contested by successive

sovereigns, claimed in both directions, or generating several bilateral disputes counts once. This yields 136 clusters across 162 disputes.⁵

a. Dependent Variable: Territorial MIDs

I use militarized interstate disputes as the source of the dependent variable. In the main models of the paper, the outcome is binary and defined at the dispute-year level. If at least one territorial militarized dispute between these two states over this disputed territory occurs during this year of this dispute, it is coded 1, and if not, 0. The measure includes threats, displays, and uses of force and covers both newly initiated and continuing MIDs. In the estimation sample, I have 583 positive observations out of 4,539 dispute-years (12.8 percent). 94 of the 162 disputes see force at least once.

As the outcome variable, I am using the presence of an MID in a dispute year, not an initiation by the challenger. The MID data record which side took the first militarized action. In a territorial dispute, that is quite often the target, responding to an encroachment, a survey party, a settlement, or a patrol that was itself not militarized. In many narratives of MIDs, who moved first is genuinely disputed, with multiple accounts pointing in opposite directions. In the MID datasets, side A identifies who moved first with force, not who is seeking to revise the territorial status quo. After reading thousands of narratives, I concluded that using side A as the initiator would exclude cases in which the target preempts or in which coders make a choice to assign the first act to the target. I therefore record a dispute-year as positive when a territorial militarized dispute occurred, and the challenger is the revisionist party, without conditioning on who fired first. The theory concerns the challenger's decision to press its claim; the measure records the dispute's militarization.

For this project, I collected my own territorial MIDs from the MID project (MID v5.0) (Palmer et al. 2022) and the corrected MID project of Gibler, Miller, and Little (Gibler et al. 2016) instead of using these datasets directly. There are two reasons for my own collection. First, these datasets code for the disputants' territorial revisionism. However, I came to realize that this coding is occasionally inaccurate and misses or misclassifies territorial MIDs. Instead of using these off-the-shelf options, I decided to collect my own MIDs from the MID narratives of Gibler (Gibler 2018), the narratives of the MID project, and other secondary sources. This allowed me to find the MIDs that were missed by others.

Second, both MID projects are keyed to the dyad rather than to the dispute. A pair of states can carry more than one territorial dispute at the same time, and in my panel, 149 dyad-years across 32 disputes do so. In those years a dyad-keyed source registers that a territorial militarized dispute occurred but cannot say

⁵ There are three criteria for clustering: (1) Shared dispute legality coding: Dispute episodes are clustered under this rule if they have a shared assigned legality coding. (2) Shared disputed territory: Dispute episodes are clustered under this rule if they have the same disputed territory across episodes and possibly succeeding actors. (3) Connected disputes between the same dyad: Dispute episodes are clustered under this rule if they are between the same dyad (unordered) and share a close geographical connection.

which dispute it belonged to, which is the only thing my design needs to know. When I was collecting the MIDs from the narratives, I matched them with specific territorial disputes. When the MID narratives were unclear, I investigated further to ensure that they corresponded to the correct territorial disputes. When unsure, I adopted a conservative position and left these MIDs out of my coding.

In the estimation sample, my coding yields 583 positive dispute-years, compared with 434 in the corrected MID data and 463 in MID v5.0. 184 of those dispute-years, spread across 56 disputes, appear in my coding but not in either of these MID datasets. Across the full panel, the figure is 199 dispute-years across 57 disputes, of which 167 fall before 2001, which is the window in which the existing legal-claims literature has worked. The correction runs in both directions. In 88 dispute-years across 29 disputes, at least one of two projects flagged a territorial MID for the dyad-year that I do not attribute to that dispute, and in 66 of those both projects flag it, usually because the year carries a concurrent dispute to which the event belongs.

The practical consequence is that roughly a third of the MID-years with force in this analysis are invisible to the sources the literature has relied on, and they are not randomly distributed: they concentrate in the earlier decades and in dyads carrying multiple simultaneous claims. Still, I use the MIDs from the corrected MID project and MID v5.0 in the robustness check to show that the results are not the product of my own DV.

Also, I test the theory by using a more narrowly formulated DV: high hostility MIDs. Militarized interstate disputes are coded on a five-level hostility scale, and levels four and five correspond to the use of force and to war. I therefore report a second dependent variable that requires the challenger's hostility to reach level four, which holds in 422 of the 583 positive dispute-years. Every result below is estimated on both and presented in the results or robustness check, so that no finding rests on the inclusion of threats and displays alongside the actual use of force.

There are two issues to be disclosed in my choice of dependent variable. First, hostility levels in the participant files are recorded once per participant per militarized dispute rather than once per year. Each year of a multi-year dispute inherits that dispute's peak level. This only affects the high-hostility DV, which is in the robustness check. Second, my measure is the use of force in territorial disputes, not conquest or attempts at conquest. The characteristic modern method of taking territory is a bloodless *fait accompli* against unpopulated and ungarrisoned ground (Altman 2017, 2020), and they could be an alternative DV in this design. I decided against it due to the omission of some conquest attempts I found in the MID narratives and Altman's exclusionary criteria, which exclude militarized confrontations that the target may preemptively initiate. In many cases, the target sees the challenger's preparation and decides to consolidate its position by eliminating the challenger's offensive capabilities. By definition, these confrontations are not in the dataset of conquest attempts.

b. Independent Variable: The Strength of Legal Claims

I use the legal claim variables from the legal claims dataset of Huth et al. (2011) as a continuous and categorical variable. Working from expert sources on each dispute and from the principles of treaty and customary international law, they place each side's claim on a three-point scale running from weak through mixed to strong. The materials are the ones an international tribunal would weigh: treaty text, colonial delimitation, *uti possidetis*, prior arbitral and judicial rulings, and long patterns of administration. The scale captures the argumentative resources a state has available, not the legal truth of its position, and I use it in that sense throughout. A state with a strong claim is one that can make a case its opponents find hard to answer, which is precisely what my theory requires.

In my extended panel, I use the same values of legal claims in a dispute in the extension period (2001-2014). Given the stability of legal claims and the legal principle that the crystallization of disputes prevents both sides from improving their legal position in a meaningful sense, this automatic use of the same values for challengers' and targets' legal claims makes sense from the legal perspective.

The two sides are coded separately and are not complements. In the estimation sample, the challenger holds a strong claim in 16 of 162 disputes and the target in 50, which tracks Huth, Croco, and Appel's (2013) own figures of 14 and 48, respectively. The two variables correlate at -0.56 across disputes rather than at -1 , and the two claims sum to zero, one, two, or three rather than to a constant. There are seven disputes in which both sides have a weak legal claim, and none in which both sides have strong claims. Because of this asymmetry, the target's claim needs its own hypotheses rather than appearing as the inverse of the challenger's. Table 1 and 2 show the distribution of disputes and dispute-year observations across the categories, and they demonstrate that there is no symmetry between legal claims.

Table 1: Distribution of Disputes at Dispute Level (% across rows)

Dispute level	Target weak	Target mixed	Target strong
Challenger weak	10% (7)	31% (21)	59% (40)
Challenger mixed	14% (11)	73% (57)	13% (10)
Challenger strong	88% (14)	12% (2)	0% (0)

Table 2: Distribution of Disputes at Dispute-year Level (% across rows)

Dispute-year level	Target weak	Target mixed	Target strong
Challenger weak	8% (149)	36% (655)	55% (999)
Challenger mixed	17% (429)	69% (1700)	13% (326)
Challenger strong	70% (198)	30% (83)	0% (0)

A natural suspicion is that legal claims are a restatement of material power, either because powerful states write the law that benefits them or because on-the-ground control manufactures title. The data do not support it. In my sample, the challenger's claim correlates with the power ratio at -0.09 at the dispute-year

level and -0.08 across disputes, and the target's at 0.12 and 0.15 , respectively. Each side's claim correlates with its own capability at -0.03 and -0.13 . If anything, the more powerful state holds the slightly weaker case. Prorok and Huth (2015) report essentially the same small correlation between capabilities and the strength of legal claims (-0.15), and explain why: many relevant legal principles predate the disputes they govern.

In Table 3, I present the outcomes by the challenger's legal claim. The peaceful resolution category combines the categories of bilateral agreement, legal dispute resolution, and other peaceful dispute-ending events such as claim renunciation or troop withdrawal. Disputes in which the challenger holds a mixed claim run longest and accumulate the most militarized disputes, and disputes with a strong challenger claim run shortest and accumulate the fewest. Per year at risk, however, the three categories are almost identical, at 12.8, 12.8, and 13.5 percent. What the per-dispute counts mostly record is how long a dispute lasts.

Table 3: Distribution of Outcomes by Challenger's Claim

Challenger's claim	Disputes	Peaceful resolution	Bilateral agreement	Legal dispute resolution	MIDs per dispute	MID-years per dispute	Dispute-years per dispute	Force rate
weak	68	63.2% (43/68)	38.2% (26/68)	8.8% (6/68)	2.4	3.4	26.8	12.8%
mixed	78	56.4% (44/78)	43.6% (34/78)	7.7% (6/78)	3.2	4	31.5	12.8%
strong	16	81.2% (13/16)	68.8% (11/16)	12.5% (2/16)	1.5	2.2	16.2	13.5%

Disputes are assigned to claim categories based on their initial values. One dispute changes claim values later and move from weak to strong. Force rate is the number of dispute-years with a territorial militarized dispute divided by all dispute-years in that category.

Table 4: Distribution of Outcomes by Target's Claim

Target's claim	Disputes	Peaceful resolution	Bilateral agreement	Legal dispute resolution	MIDs per dispute	MID-years per dispute	Dispute-years per dispute	Force rate
weak	32	75% (24/32)	68.8% (22/32)	6.2% (2/32)	3.3	4.2	23.6	17.7%
mixed	80	57.5% (46/80)	42.5% (34/80)	7.5% (6/80)	3	4.1	30.5	13.3%
strong	50	60% (30/50)	30% (15/50)	12% (6/50)	1.8	2.5	26.9	9.3%

Disputes are assigned to claim categories based on their initial values. One dispute changes claim values later and move from strong to weak. Force rate is the number of dispute-years with a territorial militarized dispute divided by all dispute-years in that category.

In Table 4, several important points become apparent. First, the force rate drops as the strength of the target's legal claim increases. We can see the deterrent effect of the target's claim descriptively. Second, legal dispute resolution increases along with the strength of targets' legal claim. This shows the role of consent in legal dispute resolution. Third, peaceful resolution is most common where the target's claim is

weak, at 75.0 percent against 60.0 percent where it is strong, and the difference is entirely bilateral agreement, 68.8 percent against 30.0 percent. A target with a poor case concedes at the table. A target with a good one goes to court instead. The two findings are the same fact seen twice, and both turn on consent.

Disputes with the challenger's different legal claims behave differently. Disputes in which the challenger has a strong legal claim are more likely to get resolved and leave the dataset early. This creates an imbalance in the dataset. Disputes in which the challenger has a weak or mixed legal claim supply a lot more observations than the ones with a strong legal claim challenger. When unmanaged, these long and frozen disputes dominate the sample and dictate the results of the models. As Table 3 shows, the number of observations coming from the disputes in which the challenger has a strong legal claim is considerably lower than other claim levels. We can see their imbalanced effect in Table 5.

Table 5: Sample Representation by Legal Claim

Claim	Disputes	Dispute-years	Share of Dispute-years	Share of Disputes
Challenger weak	68	1824	40.2%	42%
Challenger mixed	78	2455	54.1%	48.1%
Challenger strong	16	260	5.7%	9.9%
Target weak	32	755	16.6%	19.8%
Target mixed	80	2438	53.7%	49.4%
Target strong	50	1346	29.7%	30.9%

Shares are computed over the 4,539 dispute-years and 162 disputes of the estimation sample. The three challenger rows partition the sample, as do the three target rows, so each set of shares sums to 100 percent. Disputes are assigned to a claim category by their initial coding.

c. Power

I measure relative capability with Souva's (2023) Material Military Power. MMP is built from weapons stocks, naval tonnage, generation-weighted combat aircraft, armor, ballistic missiles, and nuclear warheads, and Souva describes it as capturing "actual and not latent military power."⁶ That composition is why I prefer it. The Composite Index of National Capabilities combines six components: total population, urban population, iron and steel production, primary energy consumption, military expenditure, and military personnel. Four of the six are demographic or industrial, and all four plausibly track how much a state wants territory as well as how easily it could take it. MMP contains none of them. MMP contains none of them. Since my argument treats capability as what converts a legal claim into action rather than as a motive for acting, the right measure is one less likely to generate this motive. I also re-estimate every model with the

⁶ MMP's composition changes twice inside my window: air power enters only in 1965, and the land component switches from an expenditure-derived proxy to a weapons count in 1975, with the two correlating at 0.71. Thirty percent of my estimation sample falls before 1965 and forty-six percent before 1975. I did not find discontinuity at any break. The measure's missingness has the same source: Souva notes that only the expenditure-derived land component has missing values, and 96 of my 102 missing rows fall before 1975.

military capability ratio used by Huth, Croco, and Appel, the average of the challenger's dyadic share on three indicators: military expenditure, military personnel, and expenditure per soldier.

I divide the challenger's share by the dyadic total and then center the result at parity, so the power ratio runs from -0.5 , where the target is overwhelmingly stronger, to $+0.5$, where the challenger is. Centering matters for interpretation: at zero the two states are evenly matched, and the coefficients on the other variables are read at that point rather than at an empty end of the scale. I include the squared term throughout, because the theory expects the effect of capability to turn rather than to run in one direction.

d. Controls

I include four controls and a set of duration terms in the main models, with two enforcement proxies entering a separate pair of models, chosen to cover the standard alternative accounts of militarization in a territorial dispute without absorbing the variation the hypotheses are about.

Rivalry: Territorial disputes embedded in a broader antagonism militarize more readily, and rivalry is the strongest control in the models below. I do not use the fixed rivalry lists, because they were built from earlier militarized dispute data that contained the coding errors Gibler, Miller, and Little (2016) document, and because a count of disputes in a fixed window makes the variable a near-restatement of the outcome. Following Diehl and Goertz (2000), I treat rivalry as a state that begins and then persists. A dyad becomes a rival pair in a year when it has experienced five or more militarized dispute onsets of any kind in the preceding twenty years, and it remains one as long as at least one onset has occurred in the preceding ten years. The state machine runs over a complete year grid for each dyad rather than only the years present in the panel, so that a gap between episodes of the same dispute cannot reset a rivalry that never lapsed.

Alliance: Coded as one when the disputants share a defense pact or a consultation pact in that year, from ATOP 5.1 (Leeds et al. 2002). States that have already committed to each other's security have reason to keep a territorial disagreement below the militarized threshold.

Joint democracy: Coded as one when both states score six or above on the Polity2 index (Marshall and Gurr 2020). The threshold is conventional, and the interpolated series resolves the standard interruption codes.

Contiguity: Coded as one for any form of direct contiguity in the Correlates of War classification: land border or separation by up to four hundred miles of water (Stinnett et al. 2002). Neighbors fight more for a number of reasons, and the disputes in this sample are overwhelmingly between neighbors. The variable also serves to isolate the handful that are not.

Temporal Dependency: Dispute-years are not independent. A dispute that saw force last year is far more likely to see it again, and the risk decays as the peace lengthens. I therefore include the number of years since the last force-year in the same dispute, with its square and cube, following Carter and Signorino

(2010). The clock is kept within the dispute and the dyad, and it resets when the opposing state changes, since a dispute thread can survive a state succession. Spells that are already running in 1945 begin their clock at the start of the panel and are flagged as left-censored.

Enforcement, entered additively: Two further variables proxy the intensity of third-party enforcement: a post-Cold War indicator, and the annual mean of the five permanent Security Council members' ideal points from Bailey, Strezhnev, and Voeten (2017), which captures how far the states capable of organizing a response are aligned with each other. They appear as additive controls in a separate pair of models rather than in the main specifications. Both are year-level quantities, and a year-level variable cannot identify moderation of a dispute-level effect, so I make no claim about enforcement conditioning anything. Including them also costs 119 observations.

e. Models, Estimation, and Weighting

Each hypothesis is tested on a discrete-time logit of the binary outcome at the dispute-year, the standard specification for binary time-series cross-section data (Beck et al. 1998), with the cubic peace-year polynomial described above carrying duration dependence (Carter and Signorino 2010). Standard errors are clustered on groups of disputes that share the same disputed territory, so that a territory contested by successive sovereigns, claimed in both directions, or generating several bilateral disputes counts once. This yields 136 clusters across 162 disputes. Because two of the four hypotheses involve more than one coefficient, the claim entering in interaction with both the linear and the squared power terms, I test each with a joint Wald test rather than reading significance off individual terms, and I report the substantive quantities as predicted probabilities and marginal effects evaluated across the power range rather than as coefficients, since a coefficient on an interacted term is not interpretable on its own.

Disputes in this sample range from one year to seventy, so pooling dispute-years allows a handful of long-running disputes to dominate the estimates. I therefore report every hypothesis in two forms: weighted models and unweighted models. The unweighted arm treats each dispute-year as one observation and answers what happens in a typical dispute-year. The weighted arm weights each observation by the inverse of its dispute's length, so that every dispute contributes roughly equally, and answers what happens in a typical dispute. The two are different estimands rather than a robust and a fragile version of one, and neither is the right answer on its own. The price of weighting is precision: the Kish effective sample size falls from 4,539 to about 1,678. Every hypothesis is estimated and reported in both arms, so that no test appears only in the arm where it is supported. Which arm carries each finding is itself a prediction of the design rather than a choice made after the fact, and I report the full gradient across five weighting schemes in the robustness checks below.

Weighting matters most for the disputes in which the challenger holds a strong claim. They are 5.7 percent of the dispute-years and 9.9 percent of the disputes, so pooling gives them a little over half the say their number warrants, and weighting restores it.

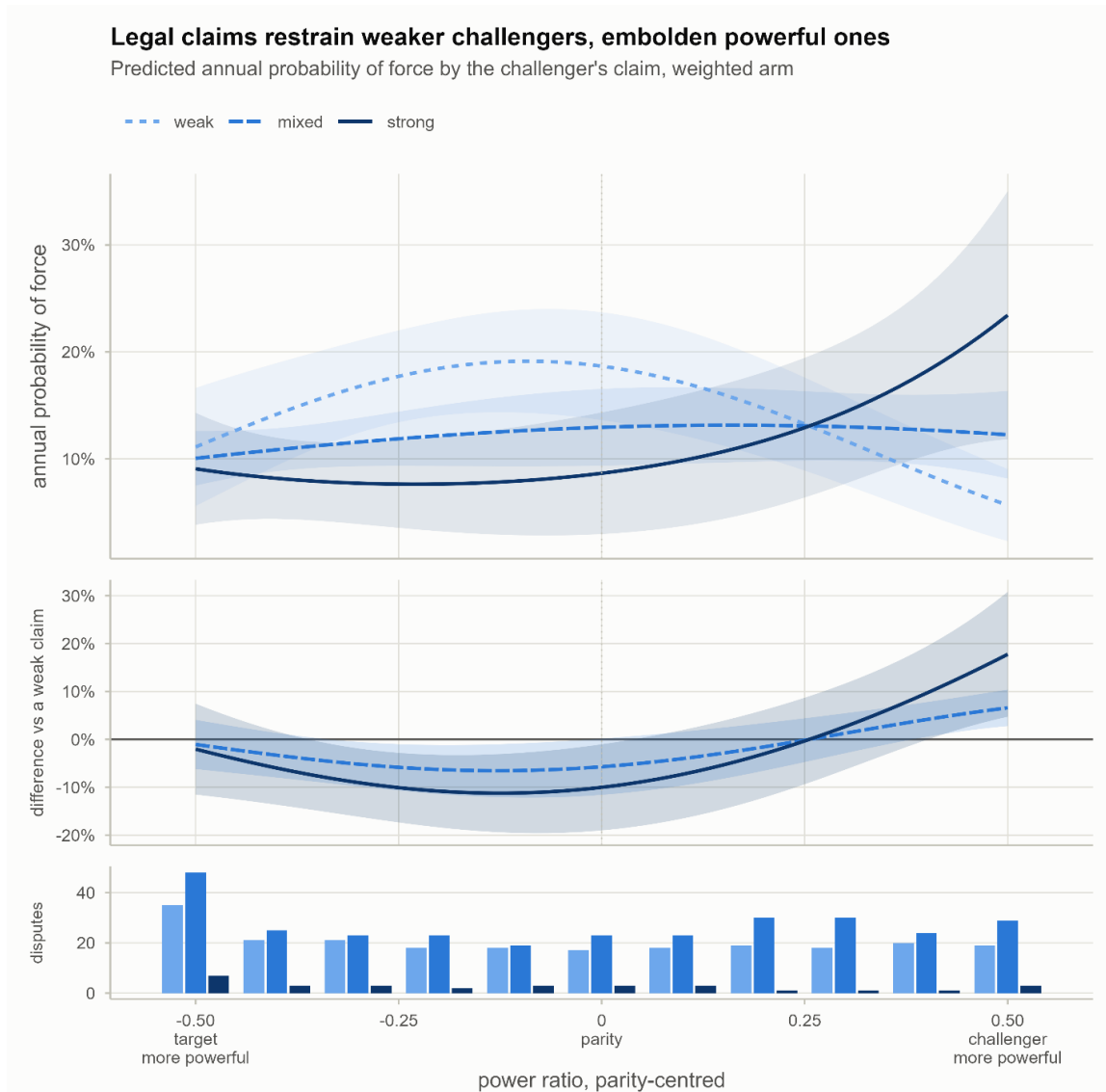
I calculate inverse-dispute-size weights using the number of dispute-years in the full analysis panel before exclusions for missing covariates. This preserves each retained observation's original weight when other years from the same dispute are excluded. Recalculating weights after these exclusions would instead redistribute a dispute's total weight across its surviving observations, potentially assigning substantial weight to a small and unrepresentative portion of its history. Under my approach, disputes with incomplete covariate coverage receive less total estimation weight than fully observed disputes. I therefore assess sensitivity to both the weighting rule and the extent of missingness.

5. Results

Tables 6 and 7 present the model ladders for each arm, Table 8 the joint test of every hypothesis in both arms, and A2 and A3 the substantive quantities. Figures 1 and 2 carry the two main findings. The findings provide support for the hypotheses. First, the challenger's legal claim does not appear to play a role on its own, exactly as predicted. Second, there is an association between challenger legal claims and militarization that varies with relative military capability in the weighted models. It reverses sign across the range of military capability, restraining a challenger that cannot win cheaply and emboldening one that can. Third, the target's legal claim lowers the likelihood of force by about four percentage points in the unweighted models, and, fourth, it does so uniformly, without interacting with capability.

I find support for the first hypothesis: The strength of a challenger's legal claim, on its own, does not change the likelihood of force in disputes. Indeed, the challenger's claim, alone, does nothing. The joint test returns insignificant results. On the probability scale, the average marginal effect is -0.6 of a percentage point, and moving a challenger from the weakest to the strongest claim changes the annual probability of force from 13.3 to 12.1 percent, a difference of 1.2 points with a standard error twice its size. This is the hypothesis rather than a failed test of it. An unconditional estimate of the challenger's claim averages a negative effect at one end of the capability range against a positive one at the other.

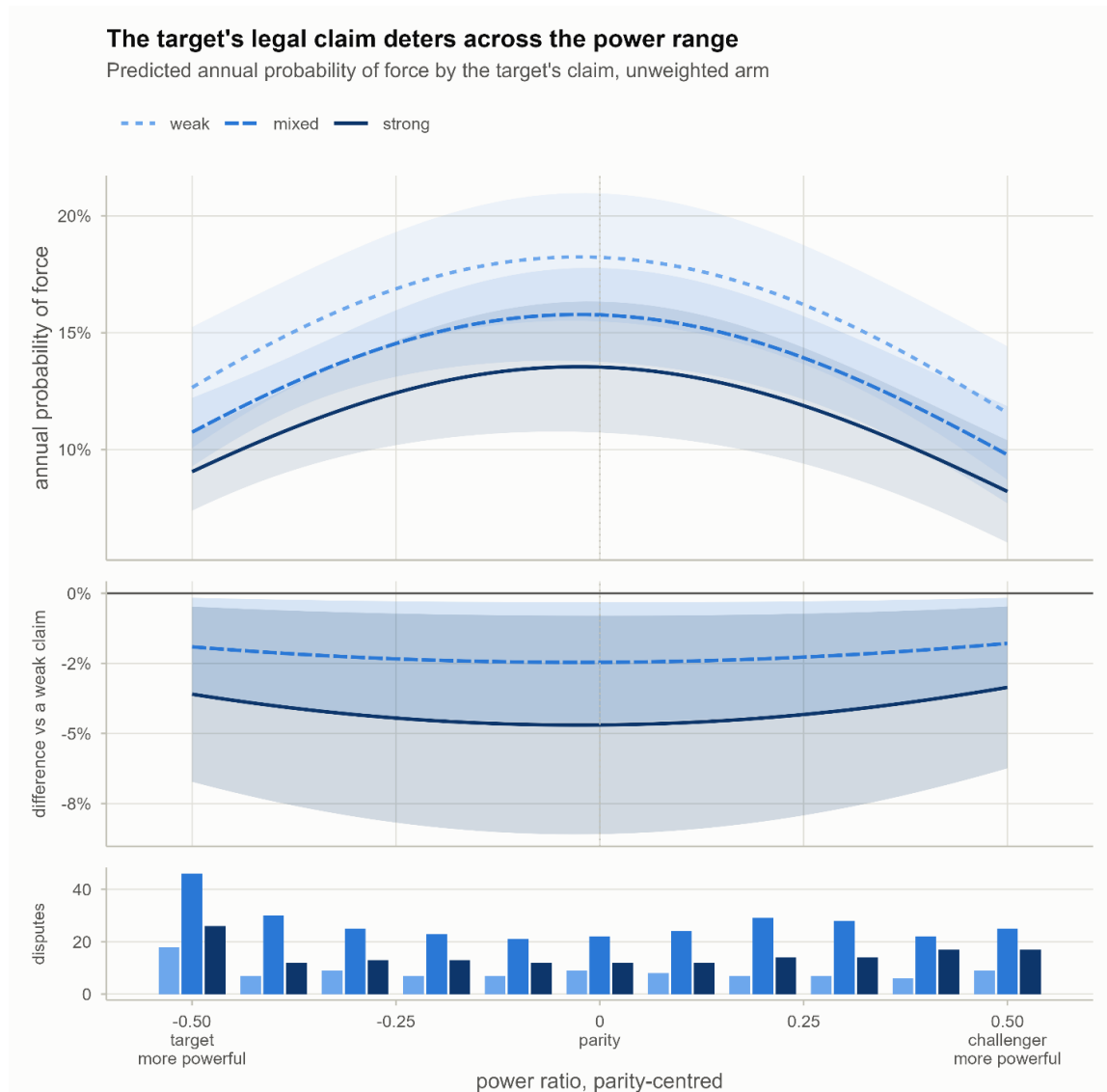
The null finding for H1 does not mean that the challenger's legal claim has no effect. I find significant support for the second hypothesis: The effect of a challenger's legal claim on the likelihood of force depends on its military capability. Figure 1 shows that the challenger's claim reverses sign across the capability range. In the weighted arm, the joint test on the interaction returns $\chi^2(2) = 11.31$, $p = .003$. A permutation test with 2,000 draws returns significant results, showing that the result is not the product of asymptotic behavior of a clustered covariance. The categorical specification, which imposes no linear spacing on the three claim levels, returns $\chi^2(4) = 12.53$, $p = .014$.



Sample-average predicted probabilities from the weighted specification (Table 6, model 4), with the challenger's claim set to each level for every observation and power held at the value on the horizontal axis. Middle panel: the same quantity as a difference against a weak claim, where the uncertainty the two levels share cancels and the band narrows. Intervals use the cluster-robust covariance. The lower panel counts disputes, not dispute-years, in each power bin: where a claim level has few or no disputes at a power value, the curve above is extrapolation. Bands are 90% intervals throughout.

Figure 1

Figure 1 shows the changing effect of the interaction, and M5b traces it. Where the challenger is far weaker than the target, the slope on its claim is -0.25 and indistinguishable from zero. It becomes restraining as the challenger acquires the capacity to fight, reaching -0.59 ($p = .043$) where the target retains a modest advantage and remaining negative through parity. It fades to nothing as the challenger's advantage grows, and turns positive at the top of the range, reaching $+0.71$ ($p = .026$). In this part of the interaction, we start to observe the emboldening effect of the legal claims.



Sample-average predicted probabilities from the unweighted specification without the target-by-power interaction (Table 6, model 4), the arm's primary specification. The claim's effect is constant on the log-odds scale here; the curvature and the varying width of the gap come from the link function and from how baseline risk changes across the power range, not from an estimated interaction. Middle panel: the same quantity as a difference against a weak claim, where the uncertainty the two levels share cancels. Intervals use the cluster-robust covariance; bands are 90% intervals throughout. The lower panel counts disputes, not dispute-years, in each power bin.

Figure 2

The estimated differences are substantively substantial. In the weighted model, moving from a weak to a strong challenger claim is associated with an annual probability of militarization approximately 9 percentage points lower where the target is more powerful and 13.6 percentage points higher where the challenger is far more powerful. Figure 3 also reports the weak-to-mixed comparison, which shows the same directional reversal with smaller estimated differences. The strong-claim estimates at high capability rest on relatively few disputes, but the weak-claim category is well populated and support the same conclusion.

What a legal claim does to the annual risk of force

Change in the predicted probability of force. Thick line: 90% interval. Thin dashed: 95%.

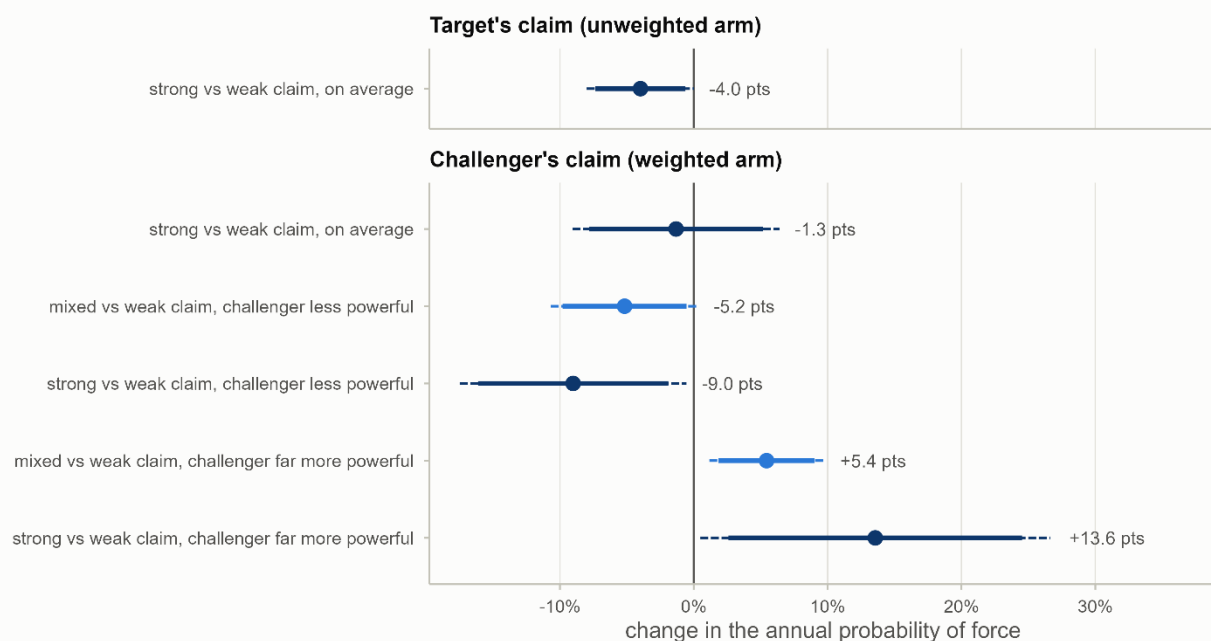


Figure 3

I find considerable support for Hypothesis 3: A stronger legal claim on the target's side reduces the likelihood of force. In the unweighted arm, the estimated coefficient is -0.225 ($SE = 0.115$, $p = .049$). Moving from a weak to a strong target claim corresponds to a decline in predicted annual probability from 15.0 to 11.0 percent; the test of this probability difference yields $p = .051$. The negative association is substantively meaningful, but its statistical precision varies across specifications, including $p = .062$ with a natural spline for temporal dependence and $p = .076$ under the reported permutation test.

Figure 2 shows this effect across the power range under the reported specification. The strong-versus-weak gap is -4.0 points where the challenger is much less powerful, -4.7 at parity, and -3.8 where it is much more powerful, and the move from weak to mixed runs between -2.0 and -2.5 . The gap widens near parity. Because the effect is a single coefficient on the log-odds scale, the same shift produces a larger change in probability where the baseline risk of force is higher, which is near parity. Whether the effect genuinely varies with capability is the fourth hypothesis, tested below with the interaction term.

The challenger interaction is supported under weighting, but not without it; the target effect is supported without weighting, but not under it. This is not two results and a choice between them. It is a gradient the design predicted before it was estimated, and A1 in Appendix A reports it across five weighting schemes. As weight moves from the average dispute-year toward the average dispute, the challenger interaction strengthens monotonically, from $\chi^2 = 0.77$ unweighted, through 3.71 under square-root weights, to 11.31 under full inverse-size weights. Over the same sequence, the target effect weakens monotonically,

Table 6: Arm A, Dispute-Year Unweighted, Clustered SE, DV= my own

Term	(1) controls	(2) + target	(3) + challenger	(4) + power	(5) + target x pw
Target's Claim		-0.190+ (0.109)	-0.237+ (0.123)	-0.225* (0.115)	-0.014 (0.172)
Challenger's Claim			-0.098 (0.132)	-0.069 (0.132)	-0.055 (0.130)
Military Ratio				-0.125 (0.197)	-0.439 (0.357)
Military Ratio Sq				-2.434*** (0.673)	-1.115 (1.153)
Target's Claim x Military Ratio					0.323 (0.287)
Target's Claim x Military Ratio Sq					-1.442 (1.029)
Rivalry	0.904*** (0.151)	0.870*** (0.154)	0.865*** (0.153)	0.733*** (0.145)	0.735*** (0.147)
Alliance	-0.296* (0.143)	-0.265+ (0.147)	-0.254+ (0.145)	-0.277* (0.140)	-0.258+ (0.142)
Joint Democracy	-0.107 (0.167)	-0.135 (0.168)	-0.112 (0.164)	-0.143 (0.160)	-0.148 (0.150)
Contiguity	0.737** (0.248)	0.726** (0.249)	0.729** (0.247)	0.621* (0.252)	0.585* (0.250)
Peace Years	-0.504*** (0.058)	-0.503*** (0.058)	-0.504*** (0.058)	-0.499*** (0.058)	-0.500*** (0.058)
Peace Years Sq	0.024*** (0.005)	0.024*** (0.005)	0.024*** (0.005)	0.024*** (0.005)	0.024*** (0.005)
Peace Years Cu	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
(Intercept)	-1.412*** (0.268)	-1.205*** (0.302)	-1.094** (0.362)	-0.682+ (0.387)	-0.875* (0.390)
Observations	4539	4539	4539	4539	4539
Clusters	136	136	136	136	136
Positives	583	583	583	583	583
Specification (4) is the arm's primary; (5) adds the target-by-power interaction so it is tested in the same estimand the target claims are made in, not only in the weighted arm. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15					

Table 7: Arm B, Dispute-Year Weighted by Inverse Dispute Size, Clustered SE, DV= my own

Term	(1) no interaction	(2) linear	(3) quadratic	(4) both	(5) both + target
Challenger's Claim	-0.143 (0.249)	-0.035 (0.241)	-0.514~ (0.320)	-0.514~ (0.315)	-0.430 (0.336)
Target's Claim	-0.062 (0.175)	-0.022 (0.168)	-0.093 (0.166)	-0.054 (0.153)	0.065 (0.265)
Military Ratio	-0.065 (0.366)	-0.627 (0.503)	-0.001 (0.365)	-0.812~ (0.510)	-2.103* (0.937)
Military Ratio Sq	-2.051+ (1.067)	-1.780+ (1.053)	-3.976* (1.552)	-4.482** (1.471)	-4.530~ (2.982)
Challenger's Claim x Military Ratio		0.795+ (0.482)		1.067* (0.442)	1.614** (0.585)
Challenger's Claim x Military Ratio Sq			2.651~ (1.736)	3.679* (1.655)	3.773+ (1.995)
Target's Claim x Military Ratio					0.860+ (0.489)
Target's Claim x Military Ratio Sq					-0.091 (1.683)
Rivalry	0.791*** (0.236)	0.827*** (0.232)	0.847*** (0.230)	0.913*** (0.218)	0.951*** (0.218)
Alliance	-0.493* (0.230)	-0.529* (0.229)	-0.511* (0.230)	-0.566* (0.230)	-0.551* (0.229)
Joint Democracy	-0.191 (0.300)	-0.207 (0.301)	-0.184 (0.295)	-0.218 (0.298)	-0.206 (0.293)
Contiguity	0.245 (0.376)	0.341 (0.369)	0.199 (0.378)	0.294 (0.362)	0.270 (0.353)
Peace Years	-0.418*** (0.063)	-0.421*** (0.062)	-0.419*** (0.064)	-0.427*** (0.063)	-0.431*** (0.062)
Peace Years Squared	0.020*** (0.005)	0.020*** (0.005)	0.020*** (0.005)	0.021*** (0.005)	0.021*** (0.005)
Peace Years Cu	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
(Intercept)	-0.831~ (0.569)	-1.050+ (0.558)	-0.504 (0.573)	-0.638 (0.551)	-0.830 (0.632)
Observations	4539	4539	4539	4539	4539
Clusters	136	136	136	136	136
Positives	583	583	583	583	583
Specification (4) is the arm's primary. Joint tests of the interaction terms are in Table 8. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15					

Table 8. Joint Wald tests of each hypothesis, estimated in both arms

Hypothesis	df	Chi2	Chi2	p-value	p-value
		unweighted	weighted	unweighted	weighted
H1 Challenger's claim	1	0.28	0.33	0.599	0.564
H2 C's claim x power (lin)	1	0.22	2.72	0.637	0.099 +
H2 C's claim x power (qua)	1	0.72	2.33	0.396	0.127 ~
H2 C's claim x power (joint)	2	0.77	11.31	0.681	0.003 **
H3 Target's claim	1	3.87	0.12	0.049 *	0.724
H4 T's claim x power (joint)	2	3.72	2.21	0.156	0.332
Power, linear	1	0.41	0.03	0.524	0.858
Power, quadratic	1	13.10	3.70	0.000 ***	0.055 +
Power, joint	2	13.13	3.70	0.001 **	0.158

Identical right-hand side within a row; only the weighting differs. Both columns are reported for every hypothesis so that no test is shown only in the arm where it is supported.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$, ~ $p < 0.15$

from $\chi^2 = 3.87$ through 1.92 to 0.12. Neither series has an exception, and the same pattern holds when weights are built on the clustering unit rather than the dispute.

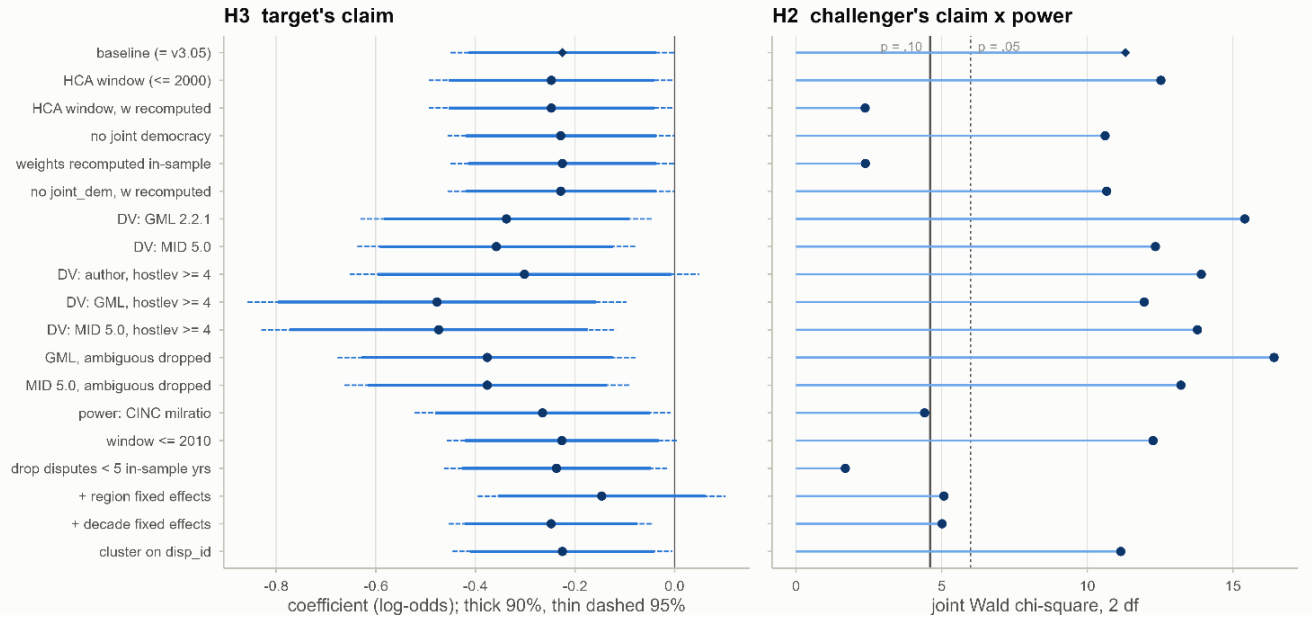
The two effects live at different levels, which is why they respond to weighting in opposite directions. The challenger's emboldening is a between-dispute contrast: it distinguishes disputes in which a capable challenger holds a good case from those in which it does not, and weighting each dispute equally is what brings that contrast into view rather than letting a handful of long-running disputes set the estimate. The target's deterrence operates within long disputes and accumulates as a dispute ages, so down-weighting long disputes removes precisely the observations in which it is visible. Had the gradient run the other way, or had it been flat, the design would have been wrong about where each effect lives.

Hypothesis 4 is a null expectation between the target's claim and military power ratio, and it is supported. The target's claim does not interact with capability. $\chi^2(2) = 3.72$, $p = .156$ in the unweighted arm and 2.21, $p = .332$ in the weighted arm. None of the eighteen robustness variants brings it below .05. This finding is consistent with H4, although it does not decisively establish that the interaction is absent.

Figure 3 summarizes the substantive differences presented in Figures 1 and 2. In the unweighted model, stronger target claims are associated with a lower average probability of militarization. In the weighted model, the small average association for challenger claims accompanies conditional estimates that reverse

Both findings across every specification choice

Diamonds mark the specification reported in the paper. Thick line / solid rule: 90%. Thin dashed: 95%.



Left: the target's claim keeps its sign in all 19 specifications; the interval clears zero in 18 of them at 90% and in 16 at 95%. Right: the joint test of the challenger-by-power interaction against both thresholds; 15 of 19 variants clear $p = .10$ and 13 clear $p = .05$. Both levels are drawn throughout so the narrower interval reported in the text can be checked against the wider one. Every deviation is discussed in the text.

Figure 4

direction across the power range: stronger claims are associated with restraint among less powerful challengers and emboldenment among substantially more powerful challengers.

6. Robustness Checks

Table B2 in Appendix B re-estimates the baseline alongside eighteen alternative specifications, and Figure 4 plots both findings across all of them. B3 summarizes. The challenger interaction clears .05 in twelve of eighteen and .10 in fourteen, with a median p of .004. The target effect keeps its sign in all eighteen, clears .05 in fifteen and .10 in seventeen, with a median p of .044. The first hypothesis remains null in fifteen of eighteen and the fourth in all eighteen.

Substituting either DV source (corrected MIDs or MID v5) for my own leaves both findings intact and, for the target, strengthens them: the interaction runs between $\chi^2 = 12.34$ and 15.41 across the corrected MIDs and MID v5.0, and the target coefficient moves from -0.225 to -0.338 and -0.358 . Restricting to high-hostility events does the same, with the target coefficient reaching -0.474 with MIDv5 high-hostility MIDs. Dropping the dyad-years in which more than one dispute is active, so that no attribution judgment of mine is doing any work, gives the strongest results in the table. Neither finding is an artifact of the hand coding.

Truncating the dataset in 2000, the window the existing literature uses (e.g. Huth et al. 2011), leaves both intact. So does truncating at 2010 (when the corrected MIDs end), dropping the joint democracy requirement, and clustering on the dispute rather than the dyad.

Three specifications weaken my challenger interaction, and they are related. Dropping disputes with fewer than five in-sample years takes it to $\chi^2 = 1.70$, $p = .428$. Recomputing the inverse-size weights inside the estimation sample rather than on the full panel takes it to 2.38, $p = .304$. Region fixed effects take it to 5.08, $p = .079$, and decade fixed effects to 5.02, $p = .081$. Substituting a capability ratio built from CINC takes it to 4.42, $p = .110$.

The challenger interaction is sensitive to recalculating weights within the estimation sample, although this procedure leaves the observations unchanged. The joint test falls from $\chi^2(2) = 11.31$, $p = .003$, to 2.38, $p = .304$. Additional diagnostics locate this sensitivity among disputes with substantial missingness, whose surviving observations receive considerably greater weight under recomputation. Excluding the three disputes retaining less than half their full-panel observations yields $\chi^2(2) = 12.66$, $p = .002$, under recomputed weights. Among disputes retaining at least 80 percent of their observations, the joint tests are nearly identical under full-panel and recomputed weights: 12.64 and 12.78, respectively.

Restricting the sample to disputes with at least five observed years produces a separate sensitivity. This restriction removes 17 disputes accounting for 10.5 percent of the baseline weighted sample, and the interaction is no longer statistically distinguishable from zero under either weighting rule. The finding therefore depends on including short disputes in the population being analyzed. By contrast, omitting each territorial cluster individually leaves the baseline joint interaction significant in every replication, with the largest p-value equal to .011. These diagnostics distinguish sensitivity to including short disputes collectively from dependence on any single cluster.

The fixed effects are a different matter, and I read them as substantive. Regions with settled legal baselines, Latin America and post-1964 Africa, contain strong target claims and little force; Asia and the Middle East show the reverse. Region proxies the availability of a determinate legal baseline. However, both findings weaken under regional controls, and I cannot fully separate the two readings.

The target effect is vulnerable in one specification: region fixed effects take the coefficient to -0.146 , $p = .249$. Decade effects leave it at -0.248 , $p = .018$. Dropping the short disputes leaves it at -0.237 , $p = .040$, so it does not share the challenger interaction's dependence.

Leaving out one dyad at a time, the coefficient stays between -0.287 and -0.175 and never changes sign, but the p-value crosses .05 in sixty-eight of the one hundred and thirty-six replications and crosses .10 in one. That is what an estimate sitting on the threshold looks like: stable in magnitude, unstable in stars.

The peace-year cubic lets the baseline risk bend with time but holds every covariate effect flat across it, an assumption worth testing directly. Allowing the interaction terms to vary with the log of time, which

is the specification the objection asks for, leaves the challenger interaction at $\chi^2 = 9.17$, $p = .010$, and the four challenger-side time terms are jointly null throughout, $\chi^2(4) = 4.58$, $p = .334$. The target effect does not weaken under a time-varying specification but strengthens, $\chi^2(2) = 13.19$, $p = .001$, because its level coefficient in that model is the effect at duration zero rather than the effect. Evaluated at successive durations it grows steadily, from nothing in the year after a militarized episode to -0.42 at five years and -0.73 at twenty. This is consistent with the account in section 6 of a deterrent operating within disputes and accumulating over their life, and it deserves more than I give it here. I flag it and leave it for future work.

The fourth hypothesis reaches $p = .014$ in one fully time-varying specification. That result rests on a single term, the squared power ratio interacted with time. Removing it returns $p = .317$, and the implied profile across durations is not interpretable, significant at zero years, null at two and five, significant again at forty. I do not treat it as support.

Entering the claims categorically rather than on a scale, the challenger interaction holds at $\chi^2(4) = 12.53$, $p = .014$, though the work is done by the strong category, $\chi^2(2) = 12.01$, $p = .002$, with the mixed category at $p = .065$. On the target side, the categorical specification is weaker than the continuous one, with strong versus weak at $p = .057$, mixed versus weak null, and the joint test at $p = .15$. The target effect is a contrast between strong claims and everything else rather than a graded response across the scale, and the continuous coefficient should be read with that in mind.

Finally, adding a post-Cold War indicator or the annual alignment of the permanent Security Council members leaves both findings in place. Substituting a natural spline for the peace-year cubic moves the target coefficient's p to $.062$ and leaves the interaction unchanged.

In addition to these checks above, I tested the effects of several other factors, such as the legal tradition of the disputants (Powell and Wiegand 2010) as controls and found that the results are robust.

7. Conclusion

The analyses presented here offer several contributions. First and foremost, they show that the mechanism by which international law pacifies territorial disputes also has a permissive branch that allows, and in fact causes, the use of force and militarization of disputes. Huth, Croco, and Appel (2011, 2012, 2013) establish that third parties read the merits of a dispute and that their reaction varies with those merits, and they trace that reaction to the support a legally advantaged state can expect when force is used against it. For a challenger that uses force to take territory it claims, what varies is not the support it receives but the punishment it does not, because a case with genuine legal foundations is harder for outside states to agree is a violation and harder for them to act together. A challenger's legal claim has no effect on the use of force on its own. Conditioned on military capability, it reverses direction, lowering

the annual probability of force from about 17 percent to 8 where the target is the more powerful side, and raising it from about 7 to 21 where the challenger is decisively more powerful.

Second, the target's legal claim deters, and it does so uniformly. A strong claim on the side in possession lowers the annual probability of force by about four percentage points, and by roughly the same amount whether the challenger is much less powerful than the target, evenly matched with it, or much more powerful. This is what a mechanism operating on how third parties characterize a seizure should look like, since characterization does not depend on who was able to carry the seizure out. The absence of an interaction on the target's side is therefore a prediction of the theory.

Third, these findings speak to the place of military power in this literature. Huth, Croco, and Appel (2013) report that the military balance is never significant in their settlement models and read that as a challenge to the conventional wisdom about bargaining leverage. The results here suggest a more specific version of the same point. Inside territorial disputes, the direct effect of relative capability on the use of force is weak, and what capability does instead is determine whether a legal claim can be converted into action at all. Power operates on these disputes conditionally rather than directly, which is why designs that enter it additively as a control find so little.

Fourth, the analysis speaks to the debate over whether international law can be effective without enforcement. Huth, Croco, and Appel note that their results illustrate law exerting influence in the absence of third-party sanctions. The account developed here locates the same phenomenon differently. Enforcement is present as an anticipation and fails in a structured way, at the stage where outside states must agree that a violation occurred rather than at the stage where they decide what to do about it. Law works here not despite the absence of enforcement but through the predictable difficulty of organizing it.

Fifth, the paper contributes an extended dataset. I carry the Huth and Allee dispute universe and the Huth, Croco, and Appel legal codings forward to 2014, and I collect territorial militarized disputes by hand and match each to the specific dispute that produced it. In the estimation sample, this recovers 184 dispute-years of force that appear in neither of the machine-coded sources, and every finding is reported across all three codings.

Two limits bear on how these results should be read. The mechanism is an anticipation, which a dispute-year panel cannot observe; what the analysis shows is that behavior moves as an anticipation account predicts and as the alternatives do not. And the outcome is the use of force in territorial disputes rather than territorial revision, so the bloodless seizure of empty ground does not register, an omission that works against the challenger finding rather than for it.

These findings also suggest three avenues for future research. First, whether third parties in fact respond to legal merit as disputants expect them to is the premise this paper takes as given and does not test; it is the natural next study. Second, the deterrent effect of the target's claim grows as a dispute ages,

which suggests that title accumulates force through the same acquiescence that builds it. Third, the same theory should have implications for peaceful settlement as well as for force, and the two outcomes are best examined together.

Bibliography

- Abramson, Scott, and David Carter. 2016. "The Historical Origins of Territorial Disputes." *American Political Science Review* 110 (4): 675–98. <https://doi.org/10.1017/S0003055416000381>.
- Allee, Todd, and Paul Huth. 2006. "The Pursuit of Legal Settlements to Territorial Disputes." *Conflict Management and Peace Science* 23 (4): 285–307.
- Altman, Dan. 2017. "By Fait Accompli, Not Coercion: How States Wrest Territory from Their Adversaries." *International Studies Quarterly* 61: 881–91. <https://doi.org/10.1093/isq/sqx049>.
- Altman, Dan. 2020. *The Evolution of Territorial Conquest After 1945 and the Limits of the Territorial Integrity Norm*. 74 (3): 490–522.
- Atzili, Boaz. 2012. *Good Fences, Bad Neighbors: Border Fixity and International Conflict*. The University of Chicago Press.
- Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61 (2): 430–56. <https://doi.org/10.1177/0022002715595700>.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (4): 1260–88. <https://doi.org/10.2307/2991857>.
- Bell, Sam R. 2017. "Power, Territory, and Interstate Conflict." *Conflict Management and Peace Science* 34 (2): 160–75. <https://doi.org/DOI:%2010.1177/0738894216650428>.
- Carter, David. 2010. "The Strategy of Territorial Conflict." *American Journal of Political Science* 54 (4): 969–87. <https://doi.org/10.1111/j.1540-5907.2010.00471.x>.
- Carter, David, and Hein E. Goemans. 2011. "The Making of the Territorial Order: New Borders and the Emergence of Interstate Conflict." *International Organization* 65 (2): 275–309. <https://doi.org/10.1017/S0020818311000051>.
- Carter, David, and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–92.
- CIA Research Paper. 1980. *Iran: The Persian Gulf Dispute*.
- Diehl, Paul F., and Gary Goertz. 2000. *War and Peace in International Rivalry*. University of Michigan Press. <https://doi.org/10.3998/mpub.16693>.
- Fazal, Tanisha M. 2007. *State Death: The Politics and Geography of Conquest, Occupation, and Annexation*. Princeton University Press.
- Franck, Thomas M. 1990. *The Power of Legitimacy among Nations*. Oxford University Press.
- Gent, Stephen E., and Megan Shannon. 2014. "Bargaining Power and the Arbitration and Adjudication of Territorial Claims." *Conflict Management and Peace Science* 31 (3): 303–22. <https://doi.org/10.1177/0738894213508710>.
- Gibler, Douglas M. 2018. *Militarized Interstate Dispute Narratives, 1816-2010*. Rowman and Littlefield Publishers.
- Gibler, Douglas M., Steven V. Miller, and Erin K. Little. 2016. "An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001." *International Studies Quarterly* 60 (4): 719–30.
- Goddard, Stacie E. 2006. "Uncommon Ground: Indivisible Territory and the Politics of Legitimacy." *International Organization* 60: 35–68. <https://doi.org/10.1017/S0020818306060024>.
- Goertz, Gary, Paul F. Diehl, and Alexandru Balas. 2016. *The Puzzle of Peace: The Evolution of Peace in the International System*. Oxford University Press.
- Guzman, Andrew. 2008. *How International Law Works: A Rational Choice Theory*. Oxford University Press.
- Hathaway, Oona A., and Scott J. Shapiro. 2017. *The Internationalists: How a Radical Plan to Outlaw War Remade the World*. Simon & Schuster.
- Huth, Paul. 1996. *Standing Your Ground: Territorial Disputes and International Conflict*. University of Michigan Press.

- Huth, Paul, and Todd Allee. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. Cambridge University Press.
- Huth, Paul, Sarah Croco, and Benjamin Appel. 2011. "Does International Law Promote the Peaceful Settlement of International Disputes? Evidence from the Study of Territorial Conflicts since 1945." *American Political Science Review* 105 (2): 415–36.
- Huth, Paul, Sarah Croco, and Benjamin Appel. 2012. "Law and the Use of Force in World Politics: The Varied Effects of Law on the Exercise of Military Power in Territorial Disputes." *International Studies Quarterly* 56 (1): 17–31.
- Huth, Paul, Sarah Croco, and Benjamin Appel. 2013. "Bringing Law to the Table: Legal Claims, Focal Points, and the Settlement of Territorial Disputes Since 1945." *American Journal of Political Science* 57 (1): 90–103.
- Irajpanah, Katherine, and Kenneth A. Schultz. 2021. "Off the Menu: Post-1945 Norms and the End of War Declarations." *Security Studies* 30 (4): 485–516. <https://doi.org/10.1080/09636412.2021.1979842>.
- Jennings, Robert Yewdall. (1963) 2017. *The Acquisition of Territory in International Law*. Manchester University Press. Reprint, Manchester University Press.
- Johns, Leslie. 2012. "Courts as Coordinators: Endogenous Enforcement and Jurisdiction in International Adjudication." *Journal of Conflict Resolution* 56 (2): 257–89. <https://doi.org/10.1177/0022002711414374>.
- Kohen, Marcelo, and Mamadou Hebie. 2011. "Territory, Discovery." In *Max Planck Encyclopedia of Public International Law*. Max Planck Encyclopedias of International Law. April. <https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e1119?prd=OPIL>.
- Leeds, Brett Ashley, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28 (3): 237–60. <https://doi.org/10.1080/03050620213653>.
- Marshall, Monty G., and Tedd Robert Gurr. 2020. "Polity5: Political Regime Characteristics and Transitions, 1800-2018." Center for Systemic Peace. <https://www.systemicpeace.org/inscrdata.html>.
- Mattes, Michaela. 2018. "Chipping Away at the Issues": Piecemeal Dispute Resolution and Territorial Conflict." *Journal of Conflict Resolution* 62 (1): 94–118. <https://doi.org/10.1177/0022002716640899>.
- Mitchell, Sara McLaughlin, and Paul R. Hensel. 2007. "International Institutions and Compliance with Agreements." *American Journal of Political Science* 51 (4): 721–37.
- Mobley, Richard A. 2003. "The Tunbs and Abu Musa Islands: Britain's Perspective." *Middle East Journal* 57 (4): 627–45.
- Nazarahari, Reza. 2005. "An Assessment of the Withdrawal of British Forces from the Persian Gulf (1971) within the Framework of Disputed Islands." In *Anglo-Iranian Relations since 1800*. Ed. by Vanessa Martin. Royal Asiatic Society Books. Routledge.
- O'Mahoney, Joseph. 2018. *Denying the Spoils of War: The Politics of Invasion and Non-Recognition*. Edinburgh University Press. <https://www.jstor.org/stable/10.3366/j.ctt1tqx9nb>.
- Orakhelashvili, Alexander. 2022. *Akehurst's Modern Introduction to International Law*. 9th ed. Routledge.
- Palmer, Glenn, Roseanne W. McManus, Vito D'Orazio, et al. 2022. "The MID5 Dataset, 2011–2014: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 39 (4): 470–82. <https://doi.org/https://doi.org/10.1177/0738894221995743>.
- Pellet, Alain. 2012. "Peaceful Settlement of International Disputes." In *Max Planck Encyclopedia of Public International Law*, VI. Max Planck Encyclopedias of International Law. <https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e70>.
- Powell, Emilia Justyna, and Krista E. Wiegand. 2010. "Legal Systems and Peaceful Attempts to Resolve Territorial Disputes." *Conflict Management and Peace Science* 27 (2): 129–51. <https://doi.org/10.1177/0738894209343979>.

- Prorok, Alyssa K., and Paul Huth. 2015. "International Law and the Consolidation of Peace Following Territorial Changes." *Journal of Politics* 77 (1): 161–74. <https://doi.org/http://dx.doi.org/10.1086/678529>.
- Schultz, Kenneth A. 2014. "What's in a Claim? De Jure versus De Facto Borders in Interstate Territorial Disputes." *Journal of Conflict Resolution* 58 (6): 1059–84.
- Schultz, Kenneth A., and Hein E. Goemans. 2019. "Aims, Claims, and the Bargaining Model of War." *International Theory* 11: 344–74. <https://doi.org/10.1017/S1752971919000071>.
- Sharma, Surya P. 1997. *Territorial Acquisition, Disputes and International Law*. Kluwer Law International.
- Simmons, Beth. 2002. "Capacity, Commitment, and Compliance." *Journal of Conflict Resolution* 46 (6): 829–56. <https://doi.org/10.1177/002200202237931>.
- Simmons, Beth. 2005. "Rules over Real Estate: Trade, Territorial Conflict, and International Borders as Institution." *JOURNAL OF CONFLICT RESOLUTION* 49 (6): 823–48. <https://doi.org/10.1177/0022002705281349>.
- Souva, Mark. 2023. "Material Military Power: A Country-Year Measure of Military Power, 1865–2019." *Journal of Peace Research* 60 (6): 1002–9. <https://doi.org/https://doi.org/10.1177/00223433221112970>.
- Stinnett, Douglas M., Jaroslav Tir, Philip Schafer, Paul F. Diehl, and Charles Gochman. 2002. "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Management and Peace Science* 19 (2): 58–66.
- Sumner, Brian Taylor. 2004. "Territorial Disputes at the International Court of Justice." *Duke Law Journal* 53: 1779–812.
- Tarar, Ahmer. 2016. "A Strategic Logic of the Military Fait Accompli." *International Studies Quarterly* 60: 742–52. <https://doi.org/10.1093/isq/sqw018>.
- Thompson, Alexander. 2009. "The Rational Enforcement of International Law: Solving the Sanctioners'dilemma." *International Theory* 1 (2): 307–21. <https://doi.org/doi:10.1017/S1752971909990078>.
- Zacher, Mark W. 2001. *The Territorial Integrity Norm: International Boundaries and the Use of Force*. 55 (2): 215–50.

Appendix A

A1. Each coefficient family across weighting schemes

family	weighting	chi2	df	p sig	direction
challenger x power (joint)	unweighted	0.77	2	0.681	-0.09, +0.80
challenger x power (joint)	sqrt (disp)	3.71	2	0.157	+0.35, +2.07
challenger x power (joint)	full (disp)	11.31	2	0.003 **	+1.07, +3.68
challenger x power (joint)	sqrt (clus)	2.11	2	0.349	+0.31, +1.64
challenger x power (joint)	full (clus)	8.43	2	0.015 *	+1.15, +2.80
target claim -> force	unweighted	3.87	1	0.049 *	-0.23
target claim -> force	sqrt (disp)	1.92	1	0.165	-0.17
target claim -> force	full (disp)	0.12	1	0.724	-0.06
target claim -> force	sqrt (clus)	3.03	1	0.082 +	-0.22
target claim -> force	full (clus)	1.27	1	0.260	-0.19
target x power (joint)	unweighted	3.72	2	0.156	+0.32, -1.44
target x power (joint)	sqrt (disp)	3.32	2	0.191	+0.21, -1.74
target x power (joint)	full (disp)	2.21	2	0.332	+0.03, -2.25
target x power (joint)	sqrt (clus)	2.41	2	0.300	+0.16, -1.54
target x power (joint)	full (clus)	1.50	2	0.472	-0.13, -1.99
challenger claim -> exit	unweighted	0.03	1	0.866	-0.03
challenger claim -> exit	sqrt (disp)	1.07	1	0.302	+0.21
challenger claim -> exit	full (disp)	4.14	1	0.042 *	+0.39
challenger claim -> exit	sqrt (clus)	1.10	1	0.295	+0.22
challenger claim -> exit	full (clus)	5.71	1	0.017 *	+0.47
target claim -> exit	unweighted	4.62	1	0.032 *	-0.32
target claim -> exit	sqrt (disp)	3.93	1	0.047 *	-0.28
target claim -> exit	full (disp)	2.39	1	0.122 ~	-0.24
target claim -> exit	sqrt (clus)	4.18	1	0.041 *	-0.30
target claim -> exit	full (clus)	1.97	1	0.160	-0.23

Joint Wald on the clustered covariance. The design predicts challenger families strengthen and target families weaken as weight moves from the average dispute-year to the average dispute. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15

A2. Substantive size: marginal effects and counterfactual probabilities				
arm	quantity	estimate	std.error	p.value
A (unweighted)	AME, legalt0	-0.0197	0.01	0.049 *
A (unweighted)	Pr(force) legalt0 = 0	0.1497	0.0125	NA
A (unweighted)	Pr(force) legalt0 = 1	0.1288	0.0059	NA
A (unweighted)	Pr(force) legalt0 = 2	0.1099	0.011	NA
A (unweighted)	difference, legalt0 2 - 0	-0.0399	0.0205	0.051 +
A (unweighted)	AME, legalc0	-0.0061	0.0115	0.599
A (unweighted)	Pr(force) legalc0 = 0	0.1327	0.0107	NA
A (unweighted)	Pr(force) legalc0 = 1	0.1266	0.0062	NA
A (unweighted)	Pr(force) legalc0 = 2	0.1207	0.0146	NA
A (unweighted)	difference, legalc0 2 - 0	-0.012	0.0225	0.595
B (weighted)	AME, legalc0	-0.0102	0.0197	0.604
B (weighted)	Pr(force) legalc0 = 0	0.1294	0.0191	NA
B (weighted)	Pr(force) legalc0 = 1	0.1171	0.0106	NA
B (weighted)	Pr(force) legalc0 = 2	0.1161	0.0249	NA
B (weighted)	difference, legalc0 2 - 0	-0.0132	0.0394	0.737
Clustered covariance, each arm's primary specification. Probabilities are sample averages with the claim set to the stated value for every observation; the weighted arm averages with its own weights, so each row is on that arm's estimand. Base rate of force in the estimation sample is 12.8% of episode-years (11.8% weighted). *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15				

A3. Effect of each claim on the log-odds of force, across the power range							
arm	claim	power	slope	se	lo_95	hi_95	p
B (weighted)	challenger	-0.45	-0.249	0.312	-0.861	0.363	0.425
B (weighted)	challenger	-0.3	-0.503	0.261	-1.014	0.008	0.054 +
B (weighted)	challenger	-0.15	-0.591	0.293	-1.165	-0.018	0.043 *
B (weighted)	challenger	0	-0.514	0.315	-1.131	0.104	0.103 ~
B (weighted)	challenger	0.15	-0.271	0.301	-0.862	0.32	0.368
B (weighted)	challenger	0.3	0.137	0.275	-0.401	0.676	0.617
B (weighted)	challenger	0.45	0.711	0.319	0.087	1.336	0.026 *
A (unweighted)	target	-0.45	-0.451	0.17	-0.785	-0.118	0.008 **
A (unweighted)	target	-0.3	-0.241	0.14	-0.515	0.034	0.086 +
A (unweighted)	target	-0.15	-0.095	0.161	-0.41	0.22	0.555
A (unweighted)	target	0	-0.014	0.172	-0.352	0.324	0.935
A (unweighted)	target	0.15	0.002	0.163	-0.317	0.322	0.99
A (unweighted)	target	0.3	-0.047	0.156	-0.352	0.259	0.764
A (unweighted)	target	0.45	-0.16	0.209	-0.571	0.25	0.444
Challenger slope from the weighted arm (Table 7, model 4). Target slope from the unweighted arm (Table 6, model 5, not main model). The main target model (Table 6, model 4) wouldn't vary across power. Thus, I present Table 6, model 5 here. Power is parity-centered. Intervals are 95%; the figures are drawn at 90% and 95%. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15							

Appendix B

B1. Each hypothesis under every specification choice, in the arm it is made in

variant	hypothesis	estimand	estimate_or_chi2	std.error	p	sig	n	events	disputes
baseline (= v3.05)	H1 challenger's claim	A unweighted	-0.069	0.132	0.599		4,539	583	162
HCA window (<= 2000)	H1 challenger's claim	A unweighted	-0.079	0.141	0.575		3,800	498	161
HCA window, w recomputed	H1 challenger's claim	A unweighted	-0.079	0.141	0.575		3,800	498	161
no joint democracy	H1 challenger's claim	A unweighted	-0.104	0.137	0.447		4,742	614	165
weights recomputed in-sample	H1 challenger's claim	A unweighted	-0.069	0.132	0.599		4,539	583	162
no joint_dem, w recomputed	H1 challenger's claim	A unweighted	-0.104	0.137	0.447		4,742	614	165
DV: GML 2.2.1	H1 challenger's claim	A unweighted	-0.272	0.166	0.101	~	4,343	434	162
DV: MID 5.0	H1 challenger's claim	A unweighted	-0.267	0.164	0.104	~	4,539	463	162
DV: author, hostlev >= 4	H1 challenger's claim	A unweighted	-0.205	0.197	0.299		4,539	422	162
DV: GML, hostlev >= 4	H1 challenger's claim	A unweighted	-0.404	0.227	0.075	+	4,343	329	162
DV: MID 5.0, hostlev >= 4	H1 challenger's claim	A unweighted	-0.433	0.218	0.047	*	4,539	346	162
GML, ambiguous dropped	H1 challenger's claim	A unweighted	-0.351	0.162	0.030	*	4,056	390	154
MID 5.0, ambiguous dropped	H1 challenger's claim	A unweighted	-0.324	0.159	0.042	*	4,252	421	154
power: CINC milratio	H1 challenger's claim	A unweighted	-0.129	0.143	0.365		4,558	593	163
window <= 2010	H1 challenger's claim	A unweighted	-0.093	0.131	0.477		4,343	558	162
drop disputes < 5 in-sample yrs	H1 challenger's claim	A unweighted	-0.019	0.133	0.885		4,497	577	145
+ region fixed effects	H1 challenger's claim	A unweighted	0.060	0.147	0.681		4,539	583	162
+ decade fixed effects	H1 challenger's claim	A unweighted	-0.054	0.124	0.661		4,539	583	162
cluster on disp_id	H1 challenger's claim	A unweighted	-0.069	0.135	0.607		4,539	583	162
baseline (= v3.05)	H2 challenger's claim x power	B weighted	11.314		0.003	**	4,539	583	162
HCA window (<= 2000)	H2 challenger's claim x power	B weighted	12.527		0.002	**	3,800	498	161
HCA window, w recomputed	H2 challenger's claim x power	B weighted	2.376		0.305		3,800	498	161
no joint democracy	H2 challenger's claim x power	B weighted	10.612		0.005	**	4,742	614	165
weights recomputed in-sample	H2 challenger's claim x power	B weighted	2.384		0.304		4,539	583	162
no joint_dem, w recomputed	H2 challenger's claim x power	B weighted	10.666		0.005	**	4,742	614	165
DV: GML 2.2.1	H2 challenger's claim x power	B weighted	15.406		0.000	***	4,343	434	162
DV: MID 5.0	H2 challenger's claim x power	B weighted	12.339		0.002	**	4,539	463	162
DV: author, hostlev >= 4	H2 challenger's claim x power	B weighted	13.915		0.001	***	4,539	422	162
DV: GML, hostlev >= 4	H2 challenger's claim x power	B weighted	11.955		0.003	**	4,343	329	162
DV: MID 5.0, hostlev >= 4	H2 challenger's claim x power	B weighted	13.782		0.001	**	4,539	346	162
GML, ambiguous dropped	H2 challenger's claim x power	B weighted	16.412		0.000	***	4,056	390	154
MID 5.0, ambiguous dropped	H2 challenger's claim x power	B weighted	13.214		0.001	**	4,252	421	154
power: CINC milratio	H2 challenger's claim x power	B weighted	4.416		0.110	~	4,558	593	163
window <= 2010	H2 challenger's claim x power	B weighted	12.259		0.002	**	4,343	558	162
drop disputes < 5 in-sample yrs	H2 challenger's claim x power	B weighted	1.695		0.428		4,497	577	145
+ region fixed effects	H2 challenger's claim x power	B weighted	5.080		0.079	+	4,539	583	162
+ decade fixed effects	H2 challenger's claim x power	B weighted	5.016		0.081	+	4,539	583	162
cluster on disp_id	H2 challenger's claim x power	B weighted	11.150		0.004	**	4,539	583	162

variant	hypothesis	estimand	estimate_or_chi2	std.error	p	sig	n	events	disputes
baseline (= v3.05)	H3 target's claim	A unweighted	-0.225	0.115	0.049	*	4,539	583	162
HCA window (<= 2000)	H3 target's claim	A unweighted	-0.247	0.126	0.049	*	3,800	498	161
HCA window, w recomputed	H3 target's claim	A unweighted	-0.247	0.126	0.049	*	3,800	498	161
no joint democracy	H3 target's claim	A unweighted	-0.229	0.116	0.049	*	4,742	614	165
weights recomputed in-sample	H3 target's claim	A unweighted	-0.225	0.115	0.049	*	4,539	583	162
no joint_dem, w recomputed	H3 target's claim	A unweighted	-0.229	0.116	0.049	*	4,742	614	165
DV: GML 2.2.1	H3 target's claim	A unweighted	-0.338	0.150	0.024	*	4,343	434	162
DV: MID 5.0	H3 target's claim	A unweighted	-0.358	0.142	0.012	*	4,539	463	162
DV: author, hostlev >= 4	H3 target's claim	A unweighted	-0.301	0.179	0.092	+	4,539	422	162
DV: GML, hostlev >= 4	H3 target's claim	A unweighted	-0.477	0.194	0.014	*	4,343	329	162
DV: MID 5.0, hostlev >= 4	H3 target's claim	A unweighted	-0.474	0.182	0.009	**	4,539	346	162
GML, ambiguous dropped	H3 target's claim	A unweighted	-0.376	0.153	0.014	*	4,056	390	154
MID 5.0, ambiguous dropped	H3 target's claim	A unweighted	-0.376	0.146	0.010	*	4,252	421	154
power: CINC milratio	H3 target's claim	A unweighted	-0.265	0.131	0.043	*	4,558	593	163
window <= 2010	H3 target's claim	A unweighted	-0.226	0.118	0.055	+	4,343	558	162
drop disputes < 5 in-sample yrs	H3 target's claim	A unweighted	-0.237	0.115	0.040	*	4,497	577	145
+ region fixed effects	H3 target's claim	A unweighted	-0.146	0.127	0.249		4,539	583	162
+ decade fixed effects	H3 target's claim	A unweighted	-0.248	0.105	0.018	*	4,539	583	162
cluster on disp_id	H3 target's claim	A unweighted	-0.225	0.113	0.045	*	4,539	583	162
baseline (= v3.05)	H4 target's claim x power	A unweighted	3.716		0.156		4,539	583	162
HCA window (<= 2000)	H4 target's claim x power	A unweighted	4.469		0.107	~	3,800	498	161
HCA window, w recomputed	H4 target's claim x power	A unweighted	4.469		0.107	~	3,800	498	161
no joint democracy	H4 target's claim x power	A unweighted	3.285		0.193		4,742	614	165
weights recomputed in-sample	H4 target's claim x power	A unweighted	3.716		0.156		4,539	583	162
no joint_dem, w recomputed	H4 target's claim x power	A unweighted	3.285		0.193		4,742	614	165
DV: GML 2.2.1	H4 target's claim x power	A unweighted	3.124		0.210		4,343	434	162
DV: MID 5.0	H4 target's claim x power	A unweighted	4.272		0.118	~	4,539	463	162
DV: author, hostlev >= 4	H4 target's claim x power	A unweighted	2.818		0.244		4,539	422	162
DV: GML, hostlev >= 4	H4 target's claim x power	A unweighted	5.036		0.081	+	4,343	329	162
DV: MID 5.0, hostlev >= 4	H4 target's claim x power	A unweighted	5.799		0.055	+	4,539	346	162
GML, ambiguous dropped	H4 target's claim x power	A unweighted	4.653		0.098	+	4,056	390	154
MID 5.0, ambiguous dropped	H4 target's claim x power	A unweighted	5.054		0.080	+	4,252	421	154
power: CINC milratio	H4 target's claim x power	A unweighted	2.870		0.238		4,558	593	163
window <= 2010	H4 target's claim x power	A unweighted	3.061		0.216		4,343	558	162
drop disputes < 5 in-sample yrs	H4 target's claim x power	A unweighted	3.596		0.166		4,497	577	145
+ region fixed effects	H4 target's claim x power	A unweighted	4.853		0.088	+	4,539	583	162
+ decade fixed effects	H4 target's claim x power	A unweighted	3.461		0.177		4,539	583	162
cluster on disp_id	H4 target's claim x power	A unweighted	4.061		0.131	~	4,539	583	162

For single-term hypotheses the reported statistic is the coefficient with its clustered standard error; for joint hypotheses it is the Wald chi-square on two degrees of freedom and no standard error applies. The first row of each block reproduces v3.05 and is a check that nothing upstream has drifted, not a result. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15

B2. The same hypotheses estimated in the other arm

variant	hypothesis	estimand	estimate_or_chi2	std.error	p sig	n	events	disputes
baseline (= v3.05)	H1 challenger's claim	B weighted	-0.143	0.249	0.564	4,539	583	162
HCA window (<= 2000)	H1 challenger's claim	B weighted	-0.151	0.273	0.580	3,800	498	161
HCA window, w recomputed	H1 challenger's claim	B weighted	-0.159	0.252	0.529	3,800	498	161
no joint democracy	H1 challenger's claim	B weighted	-0.201	0.250	0.422	4,742	614	165
weights recomputed in-sample	H1 challenger's claim	B weighted	-0.212	0.241	0.380	4,539	583	162
no joint_dem, w recomputed	H1 challenger's claim	B weighted	-0.224	0.247	0.363	4,742	614	165
DV: GML 2.2.1	H1 challenger's claim	B weighted	-0.440	0.314	0.161	4,343	434	162
DV: MID 5.0	H1 challenger's claim	B weighted	-0.482	0.311	0.122 ~	4,539	463	162
DV: author, hostlev >= 4	H1 challenger's claim	B weighted	-0.224	0.299	0.454	4,539	422	162
DV: GML, hostlev >= 4	H1 challenger's claim	B weighted	-0.554	0.362	0.126 ~	4,343	329	162
DV: MID 5.0, hostlev >= 4	H1 challenger's claim	B weighted	-0.652	0.350	0.062 +	4,539	346	162
GML, ambiguous dropped	H1 challenger's claim	B weighted	-0.551	0.317	0.083 +	4,056	390	154
MID 5.0, ambiguous dropped	H1 challenger's claim	B weighted	-0.550	0.311	0.077 +	4,252	421	154
power: CINC milratio	H1 challenger's claim	B weighted	-0.079	0.239	0.739	4,558	593	163
window <= 2010	H1 challenger's claim	B weighted	-0.149	0.255	0.558	4,343	558	162
drop disputes < 5 in-sample yrs	H1 challenger's claim	B weighted	0.185	0.171	0.279	4,497	577	145
+ region fixed effects	H1 challenger's claim	B weighted	0.024	0.217	0.913	4,539	583	162
+ decade fixed effects	H1 challenger's claim	B weighted	-0.062	0.192	0.746	4,539	583	162
cluster on disp_id	H1 challenger's claim	B weighted	-0.143	0.248	0.562	4,539	583	162
baseline (= v3.05)	H2 challenger's claim x power	A unweighted	0.768		0.681	4,539	583	162
HCA window (<= 2000)	H2 challenger's claim x power	A unweighted	1.775		0.412	3,800	498	161
HCA window, w recomputed	H2 challenger's claim x power	A unweighted	1.775		0.412	3,800	498	161
no joint democracy	H2 challenger's claim x power	A unweighted	1.037		0.595	4,742	614	165
weights recomputed in-sample	H2 challenger's claim x power	A unweighted	0.768		0.681	4,539	583	162
no joint_dem, w recomputed	H2 challenger's claim x power	A unweighted	1.037		0.595	4,742	614	165
DV: GML 2.2.1	H2 challenger's claim x power	A unweighted	0.211		0.900	4,343	434	162
DV: MID 5.0	H2 challenger's claim x power	A unweighted	0.486		0.784	4,539	463	162
DV: author, hostlev >= 4	H2 challenger's claim x power	A unweighted	2.072		0.355	4,539	422	162
DV: GML, hostlev >= 4	H2 challenger's claim x power	A unweighted	0.767		0.681	4,343	329	162
DV: MID 5.0, hostlev >= 4	H2 challenger's claim x power	A unweighted	0.866		0.648	4,539	346	162
GML, ambiguous dropped	H2 challenger's claim x power	A unweighted	0.204		0.903	4,056	390	154
MID 5.0, ambiguous dropped	H2 challenger's claim x power	A unweighted	0.386		0.825	4,252	421	154
power: CINC milratio	H2 challenger's claim x power	A unweighted	0.119		0.942	4,558	593	163
window <= 2010	H2 challenger's claim x power	A unweighted	1.706		0.426	4,343	558	162
drop disputes < 5 in-sample yrs	H2 challenger's claim x power	A unweighted	0.716		0.699	4,497	577	145
+ region fixed effects	H2 challenger's claim x power	A unweighted	0.535		0.765	4,539	583	162
+ decade fixed effects	H2 challenger's claim x power	A unweighted	1.129		0.569	4,539	583	162
cluster on disp_id	H2 challenger's claim x power	A unweighted	0.806		0.668	4,539	583	162
baseline (= v3.05)	H3 target's claim	B weighted	-0.062	0.175	0.724	4,539	583	162
HCA window (<= 2000)	H3 target's claim	B weighted	-0.068	0.195	0.727	3,800	498	161

variant	hypothesis	estimand	estimate_or_chi2	std.error	p sig	n	events	disputes
HCA window, w recomputed	H3 target's claim	B weighted	-0.029	0.185	0.876	3,800	498	161
no joint democracy	H3 target's claim	B weighted	-0.025	0.182	0.892	4,742	614	165
weights recomputed in-sample	H3 target's claim	B weighted	-0.035	0.180	0.844	4,539	583	162
no joint_dem, w recomputed	H3 target's claim	B weighted	-0.048	0.181	0.792	4,742	614	165
DV: GML 2.2.1	H3 target's claim	B weighted	-0.082	0.230	0.723	4,343	434	162
DV: MID 5.0	H3 target's claim	B weighted	-0.118	0.226	0.600	4,539	463	162
DV: author, hostlev >= 4	H3 target's claim	B weighted	-0.026	0.222	0.906	4,539	422	162
DV: GML, hostlev >= 4	H3 target's claim	B weighted	-0.180	0.261	0.490	4,343	329	162
DV: MID 5.0, hostlev >= 4	H3 target's claim	B weighted	-0.175	0.257	0.495	4,539	346	162
GML, ambiguous dropped	H3 target's claim	B weighted	-0.117	0.234	0.618	4,056	390	154
MID 5.0, ambiguous dropped	H3 target's claim	B weighted	-0.110	0.229	0.632	4,252	421	154
power: CINC milratio	H3 target's claim	B weighted	-0.091	0.186	0.626	4,558	593	163
window <= 2010	H3 target's claim	B weighted	-0.055	0.180	0.760	4,343	558	162
drop disputes < 5 in-sample yrs	H3 target's claim	B weighted	-0.214	0.140	0.125 ~	4,497	577	145
+ region fixed effects	H3 target's claim	B weighted	-0.096	0.155	0.534	4,539	583	162
+ decade fixed effects	H3 target's claim	B weighted	-0.213	0.148	0.150 ~	4,539	583	162
cluster on disp_id	H3 target's claim	B weighted	-0.062	0.188	0.742	4,539	583	162
baseline (= v3.05)	H4 target's claim x power	B weighted	2.207		0.332	4,539	583	162
HCA window (<= 2000)	H4 target's claim x power	B weighted	2.615		0.270	3,800	498	161
HCA window, w recomputed	H4 target's claim x power	B weighted	2.424		0.298	3,800	498	161
no joint democracy	H4 target's claim x power	B weighted	3.975		0.137 ~	4,742	614	165
weights recomputed in-sample	H4 target's claim x power	B weighted	1.666		0.435	4,539	583	162
no joint_dem, w recomputed	H4 target's claim x power	B weighted	4.077		0.130 ~	4,742	614	165
DV: GML 2.2.1	H4 target's claim x power	B weighted	1.297		0.523	4,343	434	162
DV: MID 5.0	H4 target's claim x power	B weighted	1.798		0.407	4,539	463	162
DV: author, hostlev >= 4	H4 target's claim x power	B weighted	2.019		0.364	4,539	422	162
DV: GML, hostlev >= 4	H4 target's claim x power	B weighted	2.591		0.274	4,343	329	162
DV: MID 5.0, hostlev >= 4	H4 target's claim x power	B weighted	2.845		0.241	4,539	346	162
GML, ambiguous dropped	H4 target's claim x power	B weighted	1.057		0.589	4,056	390	154
MID 5.0, ambiguous dropped	H4 target's claim x power	B weighted	0.992		0.609	4,252	421	154
power: CINC milratio	H4 target's claim x power	B weighted	4.221		0.121 ~	4,558	593	163
window <= 2010	H4 target's claim x power	B weighted	2.172		0.338	4,343	558	162
drop disputes < 5 in-sample yrs	H4 target's claim x power	B weighted	3.427		0.180	4,497	577	145
+ region fixed effects	H4 target's claim x power	B weighted	2.952		0.229	4,539	583	162
+ decade fixed effects	H4 target's claim x power	B weighted	3.627		0.163	4,539	583	162
cluster on disp_id	H4 target's claim x power	B weighted	2.516		0.284	4,539	583	162

Reported so that no test appears only in the estimand where it is supported. The right-hand side is identical to B2 within a row; only the weighting differs. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15

B3. How each hypothesis fares across the robustness variants

hypothesis	variants	same_sign	p_below_.05	p_below_.10	median_p	max_p
H1 challenger's claim	18	17	3	4	0.447	0.885
H2 challenger's claim x power	18		12	14	0.004	0.428
H3 target's claim	18	18	15	17	0.044	0.249
H4 target's claim x power	18		0	5	0.144	0.244

Counts exclude the baseline row. same_sign compares against the baseline estimate; for joint tests the chi-square is always positive, so that column is uninformative there and only the p-value counts. *** p<0.001, ** p<0.01, * p<0.05, + p<0.10, ~ p<0.15