The Prospect of Punishment and the Strategic Escalation of Civil Conflicts *

Casey Crisman-Cox†

February 14, 2023

Abstract

Why do some intrastate disputes escalate to conflict or war while others do not? One potential answer to this question is that intrastate groups may face punishments for backing down from challenges, but this idea is largely unconsidered in past work on intrastate conflict. Using several recently released datasets on non-violent disputes and newly-developed structural tools for identifying the costs for backing down in disputes, I consider this explanation. I find that, on average, groups expect to be punished for backing down, and that in many observations there is a U-shaped relationship between punishments and the likelihood of dispute onset. Increasing these punishments frequently increases the probability that groups challenge the state, while also increasing the likelihood that governments give into the challenge. These results suggest that prospect of punishment may actually strengthen a group's bargaining position and can lead to more challenges against the state.

*Thanks to Tyson Chatagnier, Matt Fuhrmann, Michael Gibilisco, Bethany Lacina, Brad Smith, and the participants of the 2020 Texas Triangle Conference for their helpful remarks. Additionally, Niels Appeldorn and Swarup Das provided invaluable research assistance for this project. All errors are my own.

†Texas A&M University. Email: c.crisman-cox@tamu.edu

1 Introduction

Civil conflict remains one of the most frequent and deadly forms of armed conflict, but empirically, we still know comparatively little about the bargaining processes and failures that lead to them. Challenging a state is clearly a risky proposition for groups due to strong asymmetries in resources and capacity for violence, so why do some groups make these challenges while others do not? Likewise, why do some states, in spite of their economic and military advantages, settle these situations without violence? These are fundamental questions that remain open as we still have little theory or data on the bargaining steps between civil unrest and civil conflict.

One possible explanation focuses on the prospect of punishment. That is, when states resist challenges, groups must either escalate into a costly civil conflict or back down from their demands. To the extent that backing down may be costly to groups, a strategic trade-off appears and questions arise.[1] Specifically, does the prospect of punishment deter potential challengers to the state, or does it embolden them by endowing their threats with enhanced credibility? Additionally, how do the strategic aspects of these punishments affect decisions to not only challenge the state but also to escalate from non-violent disputes to full-fledged civil conflict? These strategic concerns lead to two theoretical and empirical questions: Do groups expect to be punished for backing down, and what effect does the expectation of punishment have on the onset and escalation of intrastate disputes?

The prospect of punishment for backing down from during a dispute has a rich history in international relations. However, punishments are rarely considered in the study of intrastate conflict escalation for both theoretical and practical reasons. Theoretically, the prospect of a punishment for backing down was historically thought to be unique to democratic governments in interstate disputes (e.g., Fearon 1994). This understanding suggests that intrastate groups are unlikely to be constrained in this way. However, recent work supports the idea that punishments for

[1] Whether backing down during an intrastate dispute is costly to challengers is itself an open question. However, there are various examples of groups receiving punishment at other points in the conflict process (e.g., Cunningham and Sawyer 2019; Prorok 2016). An historical example comes from Irish republicans who in 1921 signed a treaty with the British that fell short of Irish demands for independent and unified 32 county Irish republic. Afterward, the group's membership splintered, key leaders were assassinated (e.g., Michael Collins), and the 1922 Irish Civil War began (Walsh 2002). All of these can be seen as costs for backing down from key demands.

backing down appear across regime types (e.g., Crisman-Cox and Gibilisco 2018; Weeks 2008) and go beyond interstate conflict (e.g., Crisman-Cox and Gibilisco 2021; Whang, McLean and Kuberski 2013). Given these results, we may suspect that punishment mechanisms apply to a wider set of institutions and settings than just interstate disputes involving a democracy. Indeed, recent studies have explored whether the threat of punishment affects the outcomes and duration of civil conflict (Prorok 2016, 2018). One natural question emerges: Does the prospect of punishment affect a group's decision to challenge the state or escalate their challenge into a civil conflict?

Practically, punishments for backing down have not appeared in work on civil conflict onset because the data seldom identify cases where one side takes an "off ramp" and backs down rather than escalates. Standard civil conflict datasets record violent internal disputes, but almost never consider cases where the dispute ends before violence begins.[2] As a result, the ability to consider what happens when groups back down has been historically limited.

In this paper, I use newly introduced data from Bartusevičius and Gleditsch (2018) and Acosta (2019) to push deeper into the study of how civil disputes escalate using new methods introduced by Crisman-Cox and Gibilisco (2021). These data focus on both violent and non-violent civil disputes and allow for an analysis of why some internal disputes escalate to conflict. I do this by fitting a formal model of dispute escalation to these data. In the model, punishment takes the form a (dis)utility that groups receive for not escalating to conflict after making a challenge. With the fitted model, I address the above questions about the existence and effects of punishment during the pre-conflict stage of an intrastate dispute. These questions tie the intrastate conflict literature into the broader conflict literature in a new and direct way by exploring how the prospect of punishment affects the dynamics of intrastate disputes. Specifically, I look at the decision-making processes of governments and groups from the moment of contestation to form a unified theoretical and statistical model of how civil disputes turn into civil conflicts.

Several results emerge from this approach. First, I find that groups expect to be punished

[2]Throughout, I will use the terms peace or civil peace to refer to cases where neither a violent civil conflict nor a non-violent civil dispute occur. This understanding differs from past work, which typically classifies disputes that do not escalate to conflict as peace. Additionally, the term dispute here refers to unarmed internal challenges to the state that may or may not escalate to conflict.

if they challenge the state and back down. Second, the average marginal effect of making these punishments more severe is nearly a 20% increase in the probability that groups challenge the state. On the surface, this emboldening effect may be surprising. After all, a challenge carries the risk of having to face these increased punishments down the road. Likewise, the main consensus from the interstate conflict literature is that increased punishment threats should decrease the likelihood of a challenge (e.g., Kurizaki and Whang 2015; Partell and Palmer 1999). However, this emboldening result is consistent with recent theoretical and empirical results from Crisman-Cox and Gibilisco (2018) who identify this effect in interstate conflict. Raising the expected punishment for backing down increases the credibility of a challenge (i.e., it raises the risk that the group chooses to escalate to violence should the state stand firm), states know this and become more likely to settle with a group rather than risk civil conflict. Groups exploit this effect by challenging more, despite the fact that their options in the face of the state resistance are worse. In this sense, stronger punishment threats lead to more gambling because they raise the likelihood of a non-violent success, even as they make the alternatives worse.

This overall marginal effect masks considerable heterogeneity, however. In nearly 40% of cases, punishments for backing down have a U-shaped relationship with the probability of dispute onset. Specifically, when expected punishments are small, increasing them tends to deter most groups from initiating challenges. This deterrence effect is a direct effect wherein an increase in expected punishments makes a group less willing to initiate a challenge due to concerns about having to face that punishment for backing down. However, if punishments are large, the opposite effect emerges; increasing punishments makes challenges against the state *more* likely. In these cases, increased punishments for backing down serve as a commitment device to groups. This enhanced credibility has a coercive effect on governments: raising the expected punishment for backing down increases the likelihood that governments give in when challenged, and as such makes challenging more attractive to groups. Overall, these changes in each side's incentives tend to increase the risk of overall armed conflict.

With these results, I contribute to the study of civil conflict in several ways. First, I provide new evidence that substate groups expect to be punished if they back down from a challenge

against the government, by using newly coded data to show that these punishment concerns are both present and affect the escalation process. This evidence builds on recent work that has found that groups anticipate punishment if they back down during an armed conflict (Prorok 2016). Second, I find a novel and non-monotonic relationship between expected punishments for backing down and the probability of an intrastate dispute onset. Finally, I observe several correlates of expected punishment. In the main specifications, an increase in the number of other intrastate group is associated with harsher expected punishments, matching expectations from previous work on competition among groups (e.g., Conrad and Greene 2015). Likewise groups with a political organizations tend to have larger expected punishments, on average. This variation in expected punishments can explain why some intrastate disputes escalate to conflict while others do not, as it directly affects decisions to initiate and escalate disputes.

Of course, the prospect of punishment is not the only theory that may explain intrastate challenges and their escalation to civil conflict. As in the interstate setting there can be many competing models of dispute onset and crisis escalation. Future work should consider the empirical strength of competing explanations by estimating competing structural models of the intrastate conflict process that can then be compared to the fitted model here using model-fit statistics. This comparative model testing is an important part of evaluating the empirical content of theories by fitting competing models to the same cases; this paper starts this process for intrastate disputes by focusing on a theory based on the prospect of punishment.

2 Punishment, selection, and conflict

Previous research on civil conflict focuses mainly on the direct transition from peace to armed violence. In recent years, however, there has been a push to better understand the dispute escalation stage that exists between peace and conflict. For example, Cunningham (2013) considers how various political and economic factors affect a group's decision to use conventional politics, non-violent protest, or violent conflict, while Bartusevičius and Gleditsch (2018) consider a selection model for groups who first decide whether to initiate a dispute against the state and then whether to escalate to violence or not. This growing push to include non-violent disputes as an additional outcome to the peace-conflict dichotomy provides a fuller picture of the conflict process, and it

builds on a long tradition of interstate conflict scholarship. Both the theoretical and empirical work on interstate conflict provide guidance we might how to consider this intermediate stage between peace and armed conflict.

On the theoretical side, there is a long history of crisis bargaining models starting with Fearon (1994) that provide the workhorse framework for understanding the escalation from peace to crisis to dispute. On the empirical side, Reed (2000) uses selection models to capture these intermediate stages between peace and conflict (similar to what Bartusevičius and Gleditsch (2018) do for intrastate conflict). Building on this escalation framework, the theoretical and empirical models used by Signorino and Tarar (2006), Lewis and Schultz (2003), Crisman-Cox and Gibilisco (2018), and others focus on the strategic logic of dispute escalation by fitting formal models that capture the steps between peace and conflict. This line of work highlights the multi-stage bargaining process that leads to conflict, where both sides face choices about escalating or deescalating. As such, multi-stage empirical tools that match our theoretical understanding of the conflict process provide important links between theory and data.

Using bargaining failure to explain conflict is not unique to the interstate system; observed civil conflict is frequently framed as a bargaining problem, too (e.g., Arena and Hardt 2014; Crisman-Cox 2022). One reason why actors in the interstate system may fail to find an agreeable alternative to fighting is that they are constrained by threats of punishment for backing down from a challenge (Fearon 1994). In the interstate setting, such costs typically result from forces within the state punishing actors who are caught making empty threats. Indeed, the prospect of punishment is thought to deter states from initiating interstates disputes, while also leading them into costly wars when they do.

Only recently, however, have scholars considered whether non-state actors face similar punishment constraints. The most direct look at punishment in civil conflict comes from Prorok (2016, 2018) who considers whether groups are punished for backing down during a conflict and the effects of those punishment concerns on civil conflict outcomes and duration. She argues that groups are concerned about both internal (within the group) and external (from the state) punishments and weigh these factors when deciding to continue a conflict or back down. (2016, 73).

Regarding the former, internal punishments may take on multiple forms. Cunningham and Sawyer (2019) study the internal structures of groups and note that many violent rebel groups have selection mechanisms for choosing their leaders, e.g., elections or some other selectorate made up of high ranking officials. Likewise, they argue that even among tightly controlled groups where leaders choose their own successors, the choice appears to be done with at least the tacit approval of a selectorate. In the interstate setting, selectorates are a key mechanism for imposing costs for backing down, and as such, their presence in substate actors suggests that punishment opportunities exist within these groups and that leaders need to be cognizant of their internal support. A second path for internal punishments can from supporters and fundraisers voting with their feet by choosing which group(s) they want to support against the state, as in outbidding theories of intrastate conflict. In this kind of intergroup competition, backing down may be costly as it signals a lack of resolve to potential supports; when a group backs down, it may lose supporters or resources to one or more rival group. Regarding the latter, external punishments take the form of state actions that punish groups for challenging the state at all such as repression or other consequences (e.g., assassination or imprisonment of key figures).

Prorok finds that the prospect of these various punishments concerns lead groups to engage in the same type of "gambling for resurrection" dynamics that are common to state actors. These findings provide the best evidence to date that the prospect of punishment is a factor in determining how and when civil conflicts end. However, concerns about punishment *during* a violent conflict raise an important question about their role, if any, in the selection stage before armed begins conflict. If the prospect of punishment leads a group to continue fighting an otherwise costly conflict, as she finds, then I expect that these types of punishment threats are an important factor in a group's decision about whether to challenge the state in the first place.

Following Prorok, I remain largely agnostic about which of the above sources (supporters, selectorates, or the state) or combinations thereof generates the expectation of punishment. In the theoretical model, punishment appears as a (dis)utility that groups receive for backing down after challenging the state, which leaves the door open for any or all of these sources to come into play. This broad conceptualization maintains the tractability of the model and allows for identification of

expected punishment without imposing a specific mechanism. Modeling punishments for backing down in this minimalist way matches previous structural work on identifying punishments in the interstate setting (Crisman-Cox and Gibilisco 2018; Kurizaki and Whang 2015). In the empirical model, different covariates can be included to try and identify the relative impact of these different punishment sources.

3 A structural model of civil dispute escalation

To consider the size and effects of expected punishments more thoroughly, I use a theoretical and empirical framework that directly considers expected punishment as an estimable parameter. Specifically, I rely on the canonical crisis bargaining model from Lewis and Schultz (2003). Importantly, the framework accounts for the stage between "dispute onset" and "escalation to conflict" that international relations has long acknowledged as an important phase of the conflict process and allows for, but does not require, that groups anticipate punishment when backing down from demands.

This model allows us to consider the strategic selection that goes into a group's decision to first challenge and then escalate to violence. Past efforts to consider this in-between stage tend to rely on Heckman-style selection models. While such efforts are important for bringing the escalation processes into the empirical study of civil conflict, there are a few advantages to adopting a structural approach based on a standard conflict model.

First, structural approaches can be used to test for the presence of unobserved, but fundamentally interesting, parameters. Political science has a rich history of using structural models to answer questions about theoretically interesting parameters (e.g., ideal points, reputation for resolve, or audience costs in interstate disputes). Continuing this tradition by using a structural model as a tool to search for punishment effects in the intrastate setting is a natural application of the approach.

Second, the structural approach offers a more direct match between theory and data than standard selection models. Accounting for the strategic nature of dispute escalation provides a theoretically principled way to make inferences from observational data. While no theoretical model exactly captures the true data generating process, a reasonable model allows the researcher

to estimate important quantities of interest and conduct counterfactual or comparative statics exercises at an empirically relevant equilibrium. Such model-based inference is particularly useful in situations where (i) experimental manipulation is infeasible (i.e., we cannot randomize the prospect of punishment) and (ii) the quantity of interest is inherently theoretical and, as such, unobservable outside the context of a theoretical model.

Additionally, this particular structural approach provides an advantage over some past structural approaches to studying the escalation to civil conflict. Specifically, data limits have led recent scholars to use partially observed strategic models to allow their "peaceful" outcomes to reflect both civil peace and civil disputes that do not escalate to conflict (Chatagnier and Castelli 2019; Nieman 2015). However, such approaches do not allow for the identification of punishments for backing down as they do not allow analysts to distinguish among these different non-conflict outcomes. Fortunately, newly available data from Bartusevičius and Gleditsch (2018) and Acosta (2019) allows us to move beyond partially observed structural models as they record a number of disputes that ended without violent. Furthermore, by distinguishing among these intermediate cases, I avoid Schultz's (2001) critique that punishments are hard to identify empirically because cases that do not escalate to armed conflict are typically unobserved. Because the theoretical model starts at the decision to initiate a challenge, it allows groups who do not want to risk punishment the opportunity to not challenge the state. Including this initial step is crucial for identification of the punishment parameters. Throughout, I stick with Lewis and Schultz's (2003) version of the crisis bargaining game, as it (i) focuses directly on both the dispute-to-conflict escalation process, (ii) includes the prospect of punishment as a main parameter of interest, and (iii) is designed with empirical estimation in mind.[3]

Figure 1 presents the model. A group R is bargaining with a government G over some good or policy that G controls. At the start of the interaction, R decides if it wants to challenge the government. In cases where R does not initiate a challenge, the game ends with the status quo

[3]Alternative approaches have developed to the original Lewis and Schultz (2003) version (e.g., Whang 2010). I stick with the Lewis and Schultz (2003) version for two reasons: It is computationally much easier and the main advantage of the Whang (2010) version is improved estimates of how much updating occurs after a challenge, which is not the main focus of this project.

intact (outcome SQ). When the status quo is maintained, the group and the government receive payoffs of S_R and S_G, respectively.

However, if R challenges the government, then G decides to either conciliate the group (outcome CL) or resist the challenge. If the government concedes, then the group receives a victory payoff V_R and the government receives a concession payoff C_G. If the government does not concede, then the group has to decide whether to escalate to civil conflict (outcome CC) or end the dispute without receiving a concession (BD). In the case of the former, armed conflict begins with R and G receiving their conflict payoffs W_R and W_G, respectively. In the case of the latter, the group receives a payoff r for backing down that includes any internal or external punishments incurred, while the government receives a victory payoff V_G.

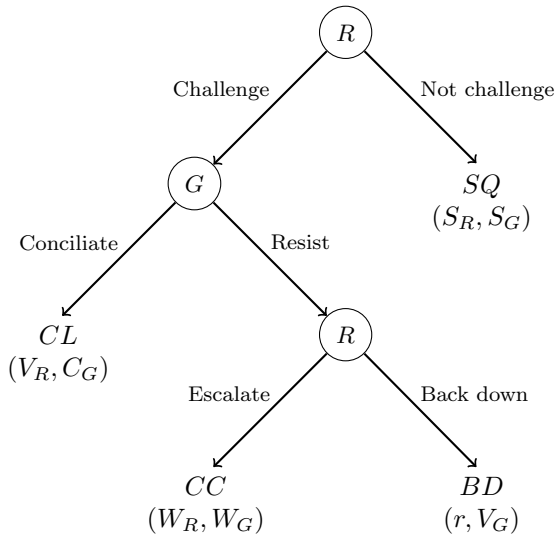
Following Lewis and Schultz (2003), I introduce private information in the form of three additively separable payoff shocks:

$$W_R = \bar{W}_R + \varepsilon_R$$
$$W_G = \bar{W}_G + \varepsilon_G$$
$$r = \bar{r} + \varepsilon_r,$$

where each ε term is drawn from a standard normal distribution and represents private information that actors have about the costs of civil conflict ($\varepsilon_G, \varepsilon_R$) and the expected punishment for backing down (ε_r). Substantively, this information setup reflects the fact that both sides know something about their own costs to entering a conflict, and the group has private information over the anticipated size of punishment. At the start of the game, Nature reveals ε_R and ε_r to the group and ε_G to the government. This information structure ensures that every outcome is reached with positive probability (i.e., there are no off-path beliefs to consider as every path is possible in equilibrium). This full-support over the outcomes guarantees that there are no zero-likelihood problems in fitting the model to data.

A perfect Bayesian equilibria (PBE) to the game can be represented using by a vector of choice probabilities. In particular, let p_g be the probability that G resists a demand, p_c is the probability

Figure 1: The dispute escalation game



that R challenges, and p_e is the probability that R escalates given that it made a demand. Jo (2011) shows that a PBE, $p^* = (p_c^*, p_g^*, p_e^*)$, can be characterized as a solution to a system of three equations $p^* = \Psi(p^*; \theta)$. The full expression of Ψ can be found in Appendix A and full derivation details can be found in Lewis and Schultz (2003) and Jo (2011).

With this equilibrium characterization, we can consider equilibrium analysis and comparative statistics. However, the model is flexible enough that $\bar{r}$ can have heterogeneous effects under reasonable parameter values. Additionally, there can be multiple equilibria and standard equilibrium refinements based on off-path beliefs (e.g., divinity) do not help with selection, as every outcome is on-path in every equilibrium. Empirically, this level of flexibility is a strength as it means I do not "bake in" a fixed relationship between punishments and outcomes before introducing the model to data. Theoretically, it means that any comparative statics considered at this time will reflect arbitrary parameter choices that may lead to very different or even indeterminate predictions. Fitting the model to data first and then analyzing it at the estimated parameter values and equilibrium enables me to study the version of the model that best matches the real world by considering the comparative statics that are most closely tethered to the observables.

Before fitting the model, however, it is worth considering how increased punishments ($\bar{r} \to -\infty$)

can lead to heterogeneous effects.[4] Specifically, making $\bar{r}$ more negative can make groups more or less likely to initiate a challenge depending on the other parameter values. The latter is a direct effect wherein increasing the punishment for backing down deters groups from challenging the state in the first place. Here, groups considering a dispute, weigh the risk that they may later find themselves having to choose between costly armed conflict or facing the punishment for backing down. As the value of this future choice decreases, groups may be less likely to initiate a dispute.

The former is a more subtle effect wherein increased punishments endow challengers with enhanced credibility. To see this logic, first note that as expected punishments for backing down get more severe, groups are more likely to stand firm after at this final node. As a result, governments recognize that as these expected punishments increase, resisting a challenge increases the risk of to armed conflict. To the extent that governments want to avoid an armed conflict, they become more likely to make concessions as punishments increase. Finally, groups, knowing that governments are more likely to back down in the face of a demand, become more likely to initiate a challenge. This is an emboldening effect: as the prospect of punishment increases, groups become more willing to challenge the state.

Fitting the model to data allows us to see which of these effects appear in the data. Note, that these effects are not exclusive. Indeed, it is entirely plausible that we can find some observations where the deterrence effect dominates, some observations where the emboldening effect emerges, and some observations where both appear (i.e., a U-shaped relationship between punishments and challenges). In such cases, as the expected punishments increase there is an initial deterrence effect wherein the change in government behavior (decreased p_g) is not large enough to make groups want to gamble on a challenge. This trend can reverse as expected punishments get even larger and the effect on p_g is large enough to make gambling attractive to groups. Depending on the estimated parameter values and each observation's data-profile, we may find any of these trends in any particular observation. As such, another advantage of this structural model is that it allows us to uncover the possibly heterogeneous effects that punishments have on dispute outcomes.

[4]Kurizaki and Whang's (2015) Figure 2 provides additional insight into how these heterogeneous effects may emerge in these types of models.

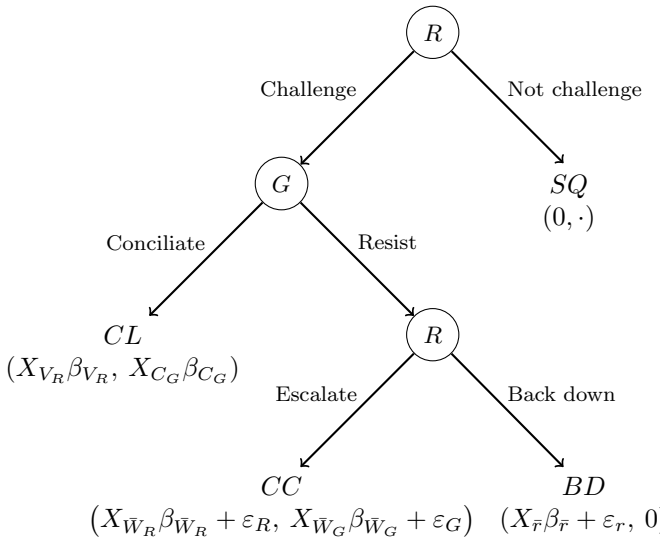
Additionally, it is worth asking: What does the parameter $\bar{r}$ mean in the civil conflict context? Depending on the estimated value of $\bar{r}$, it can be a cost or benefit that the group receives for backing down in an internal dispute relative to its other payoffs. Throughout, I focus on the relationship between the backing down payoff $\bar{r}$ and the status quo payoff S_R, which is to say that punishment is conceptualized as receiving a worse payoff for challenging and backing down than having just accepted the status quo. For instance, we might find $\bar{r} < S_R$, that is groups, on average, face a penalty for making a threat against the state and not following through on it. In this case, civil conflict emerges at this final decision node by groups who decide that such a punishment is worse than taking their group into conflict. However, it is not at all obvious that backing down will be costly. After all, both the state and internal group members may wish to actually encourage the group to not enter conflict. As such, it is entirely possible that groups may actually benefit from making demands/threats and not following through, in this case we could find that $\bar{r} > S_R$. A group that appreciates the media exposure, attention, and celebrity associated with just airing their demands might fall into this category. If this is case, then disputes would escalate to civil conflicts when there is also a high value to fighting, perhaps from the existence of oil, or some kind of exogenous shock makes conflict more attractive.

3.1 Empirical strategy and identification

Figure 2 reprints the theoretical game but parameterizes the utilities in terms of observed regressors. Note that the Figures 1 and 2 show the exact same model (i.e., there is no difference between the theoretical and statistical model) but the latter presents it in terms of covariates. Because the theoretical and empirical models are the same, there are no major differences between these figures other than to clarify some identification constraints. Specifically, S_G is omitted as it does not enter into any either player's strategic calculations, making it unidentifiable. Likewise, S_R and V_G are normalized to zero making them reference categories for the group's and government's other payoffs, respectively. This means that all the estimated utilities for the groups are interpreted as relative to their status quo, while all the estimated government utilities are relative to their victory payoff.

Direct methods of estimating β using maximum likelihood are problematic due to the possibility

Figure 2: Statistical specification of the dispute escalation game



that any particular guess at β may be associated with multiple equilibria (Jo 2011). Recall that there are no off-path beliefs in this model as every outcome is reached with positive probability in every equilibrium. As such, standard refinements (e.g., the intuitive criterion or divinity) do not alleviate the problem. Additionally, Jo (2011) and Crisman-Cox and Gibilisco (2021) show that these multiplicities can: (i) arise at plausible payoff values, (ii) that small changes in parameter values can change the number of equilibria, and (iii) that estimation problems persist even if there is a unique equilibrium at the true parameter values.[5] Given these concerns, I use the nested pseudo-likelihood estimator from Crisman-Cox and Gibilisco (2021), which they show works very well with this particular model and information structure. Additional details can be found in Appendix A.

In addition to the model's structure and the above normalizations, the identification of $\bar{r}$ is informed by the relative frequencies of the outcome variable. In particular, increasing the number of status quo or conflict outcomes, all else equal, will move the estimated $\bar{r}$ toward $-\infty$, while increasing the number of backing down outcomes shifts the estimate in the other direction. The effect of an increase in CD outcomes on the estimation of $\bar{r}$ is less direct and depends on other payoffs.

[5]To see the problem, note that while attempting to maximize the likelihood, an optimizer may find itself in an area of the parameter space where there are multiple equilibria, which can lead it astray.

4 Data

In the first three models, case selection comes primarily from the CONIAS data introduced by Bartusevičius and Gleditsch (2018), which records violent and non-violent disputes.[6] The data are built into dyad-decade observations, such that each observation records the interaction between a state and group. Decades before a challenge are recorded as status quo observations and decades with ongoing challenges are not included. Overall, this approach closely follows Whang, McLean and Kuberski (2013) and Crisman-Cox and Gibilisco (2021) who also use dyad decades when fitting an empirical crisis-escalation model. This aggregation mitigates measurement concerns in cases where any particular dispute can unfold over the course of years before reaching its terminal node.[7]

Using only the peace decades before a group challenges the state as status quo observations risks introducing selection bias by not including groups that never challenge the state. To address this concern, the second and third models include a set of never-challenge groups from the Minorities at Risk (MAR) project, with MAR groups that do not appear in CONIAS serving as additional status quo (non-dispute) outcomes. This expanded selection of status quo cases identifies a plausible set of could-be challengers that decide not to initiate a dispute over the time period considered (as in Walter 2006). Defining status quo outcomes in this way provides us with a set of politically relevant cases that includes all cases where internal disputes occur alongside a set of cases where a dispute was plausible. To summarize, this approach leads to two types of status quo cases: (i) MAR (ethnic) groups that never challenge the state and (ii) CONIAS (ethnic and non-ethnic) groups that do not challenge the state in the current decade, but do in a different decade. The combined MAR and CONIAS data deliberately omit any states without a politically active minority presence for both theoretical (these states are not playing this game) and technical reasons (there is no data on groups in countries without a recognizable group). While the ideal data would record both the

[6]The CONIAS data are produced by the Heidelburg Institute for International Conflict Research.

[7]As an alternative to the dyad-decade, I also consider an episodic approach where each observation is a government-group dyad with the final outcome as the dependent variable. The episodic approach has the advantage of having fewer aggregations but has the drawback of perhaps understating the number of status quo outcomes.

MAR groups that never challenge the state *and* a set of non-ethnic, could-be groups that never organize or challenge the state, the latter set are an inherently unobserved set of could-be groups that do not form. As such, these data are a second-best approach that allows for progress in the absence of ideal conditions.

For non-status quo outcomes, I use CONIAS's record of "political conflicts," which they describe as violent or non-violent disputes between a government and an opposition actor that are beyond the scope of ordinary politics (Trinn, Wencker and Schwank 2016). They code these cases on a five-point scale based on their final outcome. Values 3-5 on this scale denote a civil dispute that escalates to violence (outcome CC). In these cases, the dispute is treated as ending at civil conflict with neither side backing down during the escalation phase. Splitting the CONIAS cases in this way matches Bartusevičius and Gleditsch (2018), who demonstrate that this split maps into common understandings of civil conflict and reliably matches the violent conflicts recorded by the Uppsala Conflict Data Program (UCDP).

CONIAS values 1 and 2 are non-violent events, which means that they end with either the government offering concessions to the group (CL) or with groups backing down from their demands rather than escalating to conflict (BD). These cases were individually researched and hand-coded to determine which of these two outcomes best describes the situation.[8] The main distinction used to code these cases is whether the dispute ends with major changes to the status quo policy in favor of the challenger. Examples of cases that ended with a non-violent settlement in favor of the challenger include countries challenging and exiting the Soviet Union (without violence) at the end of the Cold War, non-violent democratic transitions, or the granting of elections or autonomy status. These are all situations where a demand was made, and a noticeable pro-group change occurred that ended the dispute without an escalation to violence.

Alternatively, backing down includes situations where groups initiate a challenge that neither escalates to violence nor gets any meaningful concessions from the state by the time the dispute ends. These cases are typically less well known as they neither escalate to civil conflict nor result

[8]The main sources for researching these cases are news coverage (local and global) and encyclopedia entries about the relevant groups.

in new policies or government concessions. Some examples include cases such as the Khmer people in Vietnam and various non-LTTE Tamil groups in Sri Lanka that threatened violence but never followed through. Again, the main coding depends on searching for pro-group changes to the status quo at the time COINAS records a dispute as over. Cases where the coding was ambiguous or otherwise could not be determined are excluded here but considered in an online appendix. The value added of the CONIAS data is that it tracks these nonviolent cases, even those that result in little-to-no policy changes.

4.1 Potential biases from coding

The above understanding of backing down is deliberately broad and potentially over counts the number of disputes where a group backs down as it includes groups that disbanded or merged without receiving any kind of concession. These cases are clearly neither concessions by the government nor an escalation to conflict, and as such they reflect groups that initiated disputes but exited without policy changes. This coding rule may not be perfect as it might reflect cases where groups challenge but back down because they lack the resources to escalate to conflict. However, it seems likely that internal audiences would still punish this outcome as it is still a failure and likely damages a group's reputation. Indeed, the broad understanding matches the continuum of outcomes from Fortna (2015) who argues that "fizzling out" is a clear sign of group failure; this kind of failure would be a prime candidate for either internal punishment as the group's reputation is surely tarnished. Additionally, this rule is designed to work *against* the main hypotheses by trying to over count BD cases. As mentioned above, over counting BD works against the hypothesis that groups expect to be punished for backing down. All else equal, as the number of BD outcomes increases, the estimate of $\bar{r}$ moves away from negative infinity (i.e., the costs of backing down are underestimated).

A similar concern may be in recording a conflict as ending at BD because a government crack down leads the group to back down rather than escalate (as in cases like the 2007 Khmer protests in Vietnam). In these cases, $\bar{r}$ is capturing the external punishment of the state for challenging without an ability to follow through. Failing in this way can be thought of as an extreme form of external punishment that groups will want to consider when challenging the state and should be

in their initial calculus of whether they want to issue that challenge. Likewise, and as mentioned above, over counting BD cases works against the main hypothesis as it will push $\bar{r}$ away from $-\infty$. As such, recoding these cases as backing down is the most conservative approach to considering them.

4.2 Alternatives

As alternatives to this hand-coded approach, I also consider two alternative models (Models 3 & 4, below). The first continues to use the CONIAS and MAR cases, but codes the non-violent dispute outcomes using the Ethnic Power Relations (EPR) dataset. Here, I check every non-violent dispute against the cases listed in the EPR. If the case is in the EPR, I code the outcome based on whether the group's EPR power status improves following the dispute (outcome CD) or not (outcome BD). If the case is not in EPR, I use my original coding from above.

The second alternative uses the Revolutionary and Militant Organizations dataset (REVMOD) in place of the CONIAS and MAR data (Acosta 2019). The REVMOD data codes whether the group is strictly nonviolent and whether their goals are completely, partially, or not at all obtained. Here, I code the concession outcome (CD) as occurring when a group obtains at least partial success without ever using violence, while backing down (BD) is recorded when a group does not achieve any success and never uses violence. REVMOD defines a partial success as when "organization reaches its ... goal in a limited way," and examples include power-sharing agreements, partial territorial control, autonomy, or similar changes to the status that are in the group's favor (Acosta 2019, 728). This definition is very much inline with the hand-coding approach.

The REVMOD data has additional benefits, too. Notably, the data include some information on organizational aspects of groups. Of particular interest may whether the group has a political wing, as this could be an important institution for imposing internal punishments for backing down. The MAR cases are not included here as we do not have the organization-level covariates for MAR groups. As such, the downside of fitting this model with REVMOD data is that we do not have data for groups that do not challenge the state, so the set of status quo cases is restricted to just the decades when group is active, but before it challenges the state. Overall, these approaches give us several different datasets and codings. While all of them have flaws, together they may present

a more complete analysis than any one alone.

4.3 Independent variables and specification

The main parameters of interest are the expected punishments that groups receive for backing down ($\bar{r}$). Ideally, $\bar{r}$ would be specified as group-specific fixed effect that allows each group to have their own level of expected punishment (i.e., by having a dummy variable for each group). A fixed-effects specification has the advantage of capturing all the time-invariant, group-level (and thus state-level) differences, including organizational capacity and goals that affect the punishment payoff. I consider this approach in the Online Appendix, but there are two drawbacks that preclude its use as the main model: I cannot include any group that never challenges the state (i.e., MAR groups), and the estimates exhibit signs of numerical instability. As such, I instead turn to a control variables strategy to look at how expected punishments deviate from an overall average represented by a constant term. To achieve this interpretation all non-binary covariates are standardized to mean zero. In specifying this payoff, I consider two avenues for punishment: external (from the state) and internal (from within the group/supporters).

For external punishment, I include polity score as a proxy for the institutional ability/constraints on punishment and I use logged military personnel per capita as a measure of how capable the state is at inflicting this punishment. The understanding here is that more democratic regimes may have more constraints on their ability/willingness to punish groups, while military personnel proxies for an ability to inflict punishment. Models using alternative measures of state ability/willingness to inflict punishment (tax revenue, political terror score) are considered in the Online Appendix.

For internal punishment, I use the number of active groups within the state-decade, including the MAR status quo cases. With more groups there will be enhanced internal pressure for a group to not back down so that it can stand out within the competitive environment, which is a standard argument in outbidding theories (e.g., Conrad and Greene 2015). This outbidding logic suggests that supporters will punish groups who back down by switching allegiance to another group. The REVMOD data allows for additional group-level variables as well. Here I consider a dummy for whether the group has a centralized political leadership; I expect that groups with centralized

political leadership have a greater ability to punish leaders who back down. Likewise, I include dummies for whether the group's ideology is Islamic or leftist to allow for different types of groups to impose different levels of expected punishment.

Two important points to reiterate at this time. First, the ideal data to specify deviations from the average expected punishment would record the organizational abilities and capacity of groups that do and do not escalate to violence. To the extent that such data exist, it does not consider the never-challenge groups. The group fixed-effects specification in the Online Appendix provides an important check that accommodates any group-specific heterogeneity. Second, the main hypotheses focuses on $\bar{r}$. As such the individual covariates provide interesting insight into how punishments change across dyad-decades, but the linear combination $\bar{r} = X_{\bar{r}}\beta_{\bar{r}}$ are the main parameters of interest.

The remaining payoffs reflect various pre-conflict escalation decisions and as such, I base their specification on standard control variables in the civil conflict literature. Specifically, I focus on controls that are known to both parties prior to onset of a challenge and plausibly affect the decisions to both make a demand and follow through on it. As in Lewis and Schultz (2003), identification depends on there being no regressor (including the constant) that appears in all the payoffs for any given actor. To satisfy this condition, the group's status quo payoff, S_R, and the government's victory payoff when the group backs down, V_G, are normalized to zero. Following previous work, I only allow individual variables to enter one of the governments remaining payoffs (either C_G or $\bar{W}_G$, as in Crisman-Cox and Gibilisco 2021; Kurizaki and Whang 2015; Whang, McLean and Kuberski 2013). This restriction does not follow from any formal identification result, but it tends to improve numeric stability.

For the group's victory-without-violence payoff (V_R), I include a constant, an indicator for whether they're goal is territorial separation, and logged GDP per capita. The group's goal are hand-coded by graduate student research assistants for the CONIAS conflicts and based on the MAR territorial concentration variable for remaining cases.[9] The logic behind these controls are

[9]The logic here being that territorial concentration makes it easy for groups to coordinate on territory as a goal and as such this becomes the closest proxy for goals among MAR cases.

that groups may value different goals differently, and concessions from wealthier states may be more beneficial to groups.

For the government, this outcome results in their concession payoff (C_G). Here I specify it with whether the group has territorial separation has a goal, with the understanding that different types of disputes may be particularly bitter or difficult to resolve. Additionally, I include the state's polity2 score. Finally, I control for the number of other groups within that state-decade to account for possible reputation concerns among governments. Specifically, Walter (2006) finds that states are less willing to offer concessions when there are other potential challengers. In the REVMOD model, I also include a dummy for whether the group has a political organization as this may make negotiations less costly.

For both sides' civil conflict payoff ($\bar{W}_R$, $\bar{W}_G$), I include the logged standard deviation of the terrain at the conflict location (from Shaver, Carter and Shawa 2019), along with the logged state's per capita GDP (from the Penn World Table) and logged military personnel per capita (from the NMC data). These three variables reflect the economic and military capabilities within the conflict, with rugged terrain serving as a common pre-conflict proxy for a group's conflict ability. Additionally, for the group's armed conflict payoff, I include an indicator for whether the state is an oil exporter (as in Fearon and Laitin 2003) to proxy for the level of natural resources that may be available to a potential rebel group. Finally, in the REVMOD model, I also include the logged size of the group at the earliest estimate listed in the data. More details on the model specification and summary statistics of all included covariates can be found in Appendix B.

5 Results

The first set of models are presented in Table 1. Here we consider the CONIAS cases with hand-coded outcomes. The Models 1 & 2 consider country-level covariates along with an indicator for whether the groups has a territorial claim. The difference between these models is whether they include the MAR status quo cases. The other two models consider the alternative codings of the outcome variables based on EPR and REVMOD, respectively. Recall that in the EPR case, only the outcome variable is recoded, whereas the group-level REVMOD data is used in place of the CONIAS data. The REVMOD model also contains more group-level covariates, but at the cost of

losing status quo observations where groups never challenge the state. In the interest of space, I only present the estimates associated with the backing down payoff $\bar{r}$. The other estimates can be found in Appendix C.[10]

For each model, I consider a few measures of fit. First, a χ^2 test for each model against an all-constants version of the model; larger values of the test statistic provide support for this model over the null model to test for whether the covariates offer any explanatory power. Second, I also consider the two "sensible" payoff restrictions described in Schultz and Lewis (2005): the government prefers receiving the good without a fight to conceding the good ($V_G = 0 > \hat{C}_G = X_{C_G}\hat{\beta}_{C_G}$), and groups prefer receiving the good without resistance to the status quo ($\hat{V}_R = X_{V_R}\hat{\beta}_{V_R} > \hat{S}_R = 0$).[11] The second restriction hold in every observation across the four models, and the first holds in most observations in the CONIAS based models. It is concerning that the first restriction only holds in about 7% of observations in the REVMOD model. This may lead us to be weary of these results and suggests that the much smaller sample and loss of never-challenge cases negatively impacts model fit. The fact that most of the data satisfy these sensible payoff orderings in the first three models does provide some validity.

Turning our attention to the estimates, the first thing to note is that these are the direct effects that each variable has on a group's utility for backing down. As a result, the direct interpretation of any one estimate is limited as the individual choices and outcomes are generated by the highly nonlinear choice probabilities listed in Appendix A. The main point of interest here, however, is on the estimated punishment parameter $\bar{r}$ and their relation to the status quo payoff $S_R = 0$. Specifically, we want to know if the fitted values of $\bar{r}$ are less than S_R for each observation. Across the four models, we see that for every observation in the data the payoff for backing down is negative. Additionally, we can reject the null hypothesis that $\bar{r} = 0$ in favor of the alternative that $\bar{r} < 0$ at the 5% level for each observation in Models 1-3 and in 83% of observations in Model 4.

[10]There are a few points of interest in the estimates presented in Appendix C. Notably, concessions are costly for governments; there is a large negative constant term in C_G, but these concessions tend to be less costly to democracies.

[11]Note that the estimated payoffs are all N by 1 column vectors where N is the number of observations used to fit the model. They are best thought of as vectors fitted values which is why I conduct these comparisons for each observation individually.

Table 1: Structural estimates for $\bar{r}$

	Model 1	Model 2	Model 3	Model 4
Case selection	CONIAS	CONIAS & MAR	CONIAS & MAR	REVMOD
Outcome coding	Researcher	Researcher	EPR	REVMOD
Constant	−2.46	−2.63	−2.62	−0.47
	(0.24)	(0.19)	(0.16)	(0.25)
Democracy	0.06	0.10	0.15	−0.02
	(0.18)	(0.09)	(0.08)	(0.17)
Military size	−0.01	0.06	0.11	0.18
	(0.19)	(0.11)	(0.09)	(0.19)
Other groups	−0.29	−0.21	−0.10	−0.14
	(0.19)	(0.13)	(0.11)	(0.34)
Separatist Dispute	0.31	0.16	0.33	−0.76
	(0.27)	(0.18)	(0.18)	(0.45)
Political wing				−0.83
				(0.39)
Islamic				−1.28
				(0.55)
Leftist				−0.69
				(0.31)
Log Likelihood	–717.13	–897.69	–921.87	–365.66
Observations	1207	2352	2357	533
χ^2	71.84	56.81	54.47	226.16
% $V_R > S_R$	100%	100%	100%	100%
% $V_G > C_G$	83%	100%	100%	7%
% $S_R > \bar{r}$	100%	100%	100%	100%

Standard errors in parentheses. All non-binary covariates are standardized. Constants represent the payoff when non-binary covariates are at their average and dummy variables are 0. Payoffs presented here are for $\bar{r}$ only. Other payoff estimates suppressed for space but presented in Table C.1.

To be precise, this means comparing the total value of $\hat{\bar{r}} = X_{\bar{r}}\hat{\beta}_{\bar{r}}$ to $0 = \hat{S}_R$ for each observation. Across these four models, I find evidence that groups expect to be punished if they back down.

This finding highlights the strategic tension that groups face in deciding whether to challenge the state. If the state stands firm against them, groups need to choose between costly fighting or accepting these punishments. Additionally, this result confirms and extends findings by Prorok (2016, 2018), who argues that groups are aware of and want to avoid punishments for backing down during armed conflicts with the state. The results in Table 1 provide new and direct evidence that these concerns exist in the intrastate conflict process and are present at this earlier stage

of the process. This finding at pre-armed conflict stage is new result for literature and using a structural model provides additional evidence to the idea that non-state groups are also concerned and act to avoid punishments for backing down. The advantage of the structural approach here is that it allow for the direct estimation of the punishment parameters, and we can use to consider comparative statics on the prospect of punishment. Previous work tends to rely on indirect tests based on implications derived from theories of punishment. The triangulation of those reduced-form tests with the structural approach used here is important for making progress on questions with unobservable quantities of interest, as it allows us to build a base of knowledge using different sets of assumptions, data, and methodological approaches.

Additionally, this finding that backing down is costly for groups provides some insight into the dispute escalation process and offers a new explanation for why some intrastate disputes escalate to armed conflict while others do not. Specifically, the existence of these punishments suggests that groups engage in risky behavior when they initiate a dispute. As such, when the government does not back down, groups may find themselves escalating to costly civil conflict because they might expect conflict to be less costly than the expected internal and external punishments for backing down. While this escalation process is familiar to interstate conflict scholars, the above results provide some of the first direct empirical evidence for these types of punishment concerns exist within the intrastate conflict process.

In Table 1, the number of other groups is the only predictor that has the same sign across the four specifications. On average, as the number of other actual or potential challengers to the state increases, punishments for backing down become more severe (more negative). This result suggests that outbidding dynamics influence punishment concerns. According to outbidding scholars, groups know that they are competing against other groups for resources, recruits, or exposure (Conrad and Greene 2015; Polo and González 2020). Backing down from a challenge hurts a group's ability to do these things as they develop a reputation for making non-credible challenges. This logic appears in the results as an increase in the number of other groups is associated with stronger expected punishments for backing down.

Turning to Model 4, we can see that the group-level features provide some insight into what

Table 2: The marginal effects of punishment (Model 2)

	Marginal effect of $\bar{r} \to -\infty$	
	Observations	Average Marginal Effect
$\uparrow$ Pr(Challenge)	69%	18%
$\uparrow$ Pr(Gov't resists | Challenge)	0%	−14%
$\uparrow$ Pr(Civil Conflict)	100%	31%

In the first column, percentages denote the proportion of observations where increasing the expected punishment increases the probability of interest. The second column reports the average percentage change in the probability when punishments increase

these punishments look like. Specifically, we see that groups with a political organization expect to face more severe punishments. This finding is intuitive as political organizations are a natural way to constrain leaders and to hold them accountable. Additionally, we see that leftist and Islamic terrorist groups are both associated with higher levels of punishment than groups with other ideologies. Whether this is due to harsher internal or external punishments (or both) is unclear from this analysis, although it is worth noting that leaders of Islamic groups are killed and assassinated at higher rates compared to other groups in the data. As mentioned, however, Model 4 is limited by not having covariates for groups that never challenge the state. Future work should focus on collecting more data to expand on the number of covariates we have on violent challengers, non-violent challengers, and non-challengers, allowing for a more thorough exploration of the predictors and mechanisms of punishments.

5.1 The prospect of punishment and civil dispute escalation

Having established that groups, on average, expect to be punished for backing down in disputes with the state, we are now interested in finding out how the prospect of punishment affects both group and government strategies. Specifically, we are interested in how changes in punishments affect the decision to challenge the state and how governments respond to these challenges. The statistical model allows each observation to have different equilibrium choice probabilities and marginal effects. Table 2 reports the percentage of government-group decades where the marginal effect of a more severe punishment has a positive effect on a quantity of interest.

In the first row, we see a notable split in how the prospect of punishment affects the probability that a group challenges the state. In most observations, increasing the expected punishment *increases* the probability that groups initiates a dispute and the average effect of raising punish-

ments is a 18% increase (about 1.6 percentage points) in the probability that a group challenges the state. However, this average effect masks some notable heterogeneity in the effects of punishment on whether a group initiates a challenge. In roughly 40% of observations, a U-shaped relationship between expected punishment size and the probability of a dispute appears.[12]

This U-shape is an interesting and novel finding in the civil conflict literature, but it has some intuitive appeal. When punishments are small, it makes little difference to groups if they follow through. In these cases, groups can make as many non-credible challenges because these demands are relatively costless. At the other end of the U, however, groups may face large costs for making challenges and backing down. Here, their hands are tied. When they make a challenge, they are more likely to follow through because of these costs. This enhanced credibility can make challenges potentially beneficial to groups.

In particular, note that in every single observation, we see that an increase in expected punishments makes the government more likely to back down. This result suggests that the enhanced credibility from punishments makes governments more likely to give in and, as such, makes challenges attractive to groups despite the risk that they may have to pay it. To maximize the likelihood of extracting concessions from a government, groups should take lengths to have well-structured punishment mechanisms for when they back down from challenges. While deterrence logic suggests that such a structure should discourage prospective challengers from initiating a dispute in the first place, the results here clearly demonstrate that there are also hand-tying benefits: as expected punishments increase, governments are more likely to give into demands, making groups more willing to risk a challenge. Put another way, the coercive benefit of punishments can frequently outweigh the deterrence effect.

In the final row of Table 2, we see that the overall risk of armed conflict increases as expected punishments increase. Overall, this tells us that the encouraging effect that punishments have on challenges dominates the increased probability that governments back down. Without the structural approach, it would be difficult to identify these competing effects. Indeed, it would be

[12]To identify U-shaped relationships I look at the direction of the marginal effects at plus and minus two standard errors around the estimated constant in $\bar{r}$.

understandable to attribute the reduced-form trend (larger punishments increase the likelihood of conflict) to the government standing more firm and forcing groups to choose between standing firm and fighting. However, we see that this mechanism is not supported by the data. Instead, thanks to the model fitting enterprise, we can note that governments are not standing more firm against groups with larger punishments, rather these punishments make threats to fight more credible and governments give in more. The increase in civil conflict is driven by an increase in challenges against the state as groups are more incentivized to gamble with this credibility by making more challenges.

6 Conclusion

In this paper, I asked do rebel groups face punishment for backing down from demands, and how do these punishments effect the onset of both intrastate disputes and conflict? To address this question I used newly publicized data from the CONIAS and REVMOD projects that record instances of internal disputes that do and do not escalate to violence. Using these data, I fit a model of dispute escalation that tracks the bargaining process from the decision to initiate a dispute through the decision to either back down or escalate to armed conflict.

Several results emerged from this study. First, groups expect to be punished if they challenge the state and then back down. This prospect of punishment is an important part of their strategic choice to initiate a dispute against the state. Second, depending on how strong these punishments are, they can have disparate effects on the probability of dispute onset. In many cases, I uncover a U-shaped relationship between the size of punishments and the probability that a group initiates an intrastate dispute.

When punishments are larger than average, factors that increase them can make demands and threats more credible. Groups exploit this enhanced credibility by initiating more disputes. However, when punishments are weaker than average, increases to them make groups less likely to initiate a dispute. In this range, increasing punishments makes onset riskier for groups as the credibility benefits are less pronounced. Governments are very likely to resist demands from groups with small punishments, and groups are more likely to have to pay them.

Despite these competing effects at the initiation stage, there is clear evidence that a coercive

benefit emerges when we consider the government's decision. Increases in expected punishments lead to a decrease in the probability that the government resists demands. Governments understand that enhanced punishments tend to decrease the likelihood that a group will back down at its final node, and so governments become more likely to back down instead rather than risk armed conflict. Here, governments are making a calculation that the risk of armed conflict is too high and are more willing to give in to group demands.

The primary contribution of this work is to provide direct evidence that intrastate groups face the prospect of punishment when deciding whether to initiate a dispute and that the effects of these punishments are frequently non-monotonic. While recent work has identified that these punishments appear during armed intrastate conflicts (Prorok 2016, 2018) using indirect reduced-form methods. The results presented here show that punishment concerns also appear at the earliest stages of the conflict process and provide a more direct approach that connects theory and data. Additionally, while there are good reasons to suspect that the threat of punishment might deter potential rebel groups, the evidence suggests that such punishments tend to impart a coercive advantage that strengthens a prospective rebel's bargaining position and makes governments more likely to back down before the armed conflict stage. Indeed, these results provide a new take on why governments fight some rebels rather than others, wherein the size of the punishment faced by groups for backing down greatly effects the government's decision to either concede or risk a fight.

Future work should focus on expanding the set of observables associated with non-violent CONIAS and other status quo cases. Notably, while recent data collection efforts have been very good at recording more observable characteristics of armed rebel groups (including their leader selection mechanisms), more should be done to record the observable characteristic for the non-violent groups in order to understand the escalation process. While such an undertaking will be undoubtedly costly, it will likely result in improved estimates of both the punishment parameters and in modeling the strategic selection into armed conflict. Another avenue of future work could also look to further disentangle the differences between external and internal punishment. While the control variable approach used offers some headway on this problem, a more complex model that further separates the two paths of punishment may be of interest. Additionally, future scholars may also

want to consider the benefits of a fitting a fully dynamic game of intrastate dispute escalation and de-escalation to data.

References

Acosta, Benjamin. 2019. "Reconceptualizing resistance organizations and outcomes: Introducing the Revolutionary and Militant Organizations dataset (REVMOD)." *Journal of Peace Research* 56(5):724–734.

Arena, Philip and Brian Hardt. 2014. "Incentives to Rebel, Bargaining, and Civil War." *International Interactions* 40(1):127–141.

Bartusevičius, Henrikas and Kristian Skrede Gleditsch. 2018. "A Two-Stage Approach to Civil Conflict: Contested Incompatibilities and Armed Conflict." *International Organization* 73:225–248.

Chatagnier, J. Tyson and Emanuele Castelli. 2019. "The Arc of Modernization: Economic Structure, Materialism, and the Onset of Civil Conflict." *Political Science Research and Methods* 7(2):233–252.

Conrad, Justin and Kevin Greene. 2015. "Competition, Differentiation, and the Severity of Terrorist Attacks." *Journal of Politics* 77(2):546–561.

Crisman-Cox, Casey. 2022. "Democracy, Reputation for Resolve, and Civil Conflict." *Journal of Peace Research* 59(3):382–394.

Crisman-Cox, Casey and Michael Gibilisco. 2018. "Audience Costs and the Dynamics of War and Peace." *American Journal of Political Science* 62(3):566–580.

Crisman-Cox, Casey and Michael Gibilisco. 2021. "Estimating Crisis Signaling Games in International Relations: Problems and Solutions." *Political Science Research and Methods* 9(3):565–582.

Cunningham, Kathleen Gallagher. 2013. "Understanding strategic choice: The determinants of civil war and nonviolent campaign in self-determination disputes." *Journal of Peace Research* 50(3):291–304.

Cunningham, Kathleen Gallagher and Katherine Sawyer. 2019. "Conflict negotiations and rebel leader selection." *Journal of Peace Research* 65(5):619–634.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, insurgency, and civil war." *American Political Science Review* 97(1):75–90.

Fortna, Virginia Page. 2015. "Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes." *International Organization* 69(3):519–556.

Jo, Jinhee. 2011. "Nonuniqueness of the Equilibrium in Lewis and Schultz's Model." *Political Analysis* 19(3):351–362.

Kurizaki, Shuhei and Taehee Whang. 2015. "Detecting Audience Costs in International Disputes." *International Organization* 69(4):949–980.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11(4):345–367.

Nieman, Mark David. 2015. "Statistical Analysis of Strategic Interaction with Unobserved Player Actions:Introducing a Strategic Probit with Partial Observability." *Political Analysis* 23(3):429–448.

Partell, Peter J. and Glenn Palmer. 1999. "Audience Costs and Interstate Crises: An Emprical Assessment of Fearon's Model of Dispute Outcomes." *International Studies Quarterly* 43(2):389–405.

Polo, Sara MT and Belén González. 2020. "The Power to Resist: Mobilization and the Logic of Terrorist Attacks in Civil War." *Comparative Political Studies* 53(13):2029–2060.

Prorok, Alyssa K. 2016. "Leader Incentives and Civil War Outcomes." *American Journal of Political Science* 60(1):70–84.

Prorok, Alyssa K. 2018. "Led Astray: Leaders and the Duration of Civil War." *Journal of Conflict Resolution* 62(6):1179–1204.

Reed, William. 2000. "A unified statistical model of conflict onset and escalation." *American Journal of Political Science* pp. 84–93.

Schultz, Kenneth A. 2001. "Looking for Audience Costs." *Journal of Conflict Resolution* 45(1):32–60.

Schultz, Kenneth A. and Jeffrey B. Lewis. 2005. "Learning About Learning." *Political Analysis* 14(1):121–129.

Shaver, Andrew, David B. Carter and Tsering Wangyal Shawa. 2019. "Terrain ruggedness and land cover: Improved data for most research designs." *Conflict Management and Peace Science* 36(2):191–218.

Signorino, Curtis S. and Ahmer Tarar. 2006. "A Unified Theory and Test of Extended Immediate Deterrence." *American Journal of Political Science* 50(3):586–605.

Trinn, Christoph, Thomas Wencker and Nicolas Schwank. 2016. "Introducing the Heidelberg Approach to Conflict Research." Available at https://www.uni-heidelberg.de/md/politik/personal/croissant/personal/trinn_et_al_-_introducing_the_heidelberg_approach.pdf. Last accessed: 13 February 2020.

Walsh, Oonagh. 2002. *Ireland's Independence, 1880-1923*. New York: Routledge.

Walter, Barbara F. 2006. "Building Reputation: Why Governments Fight Some Separatists but Not Others." *American Journal of Political Science* 50(2):313–330.

Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1):35–64.

Whang, Taehee. 2010. "Empirical Implications of Signaling Models." *Political Analysis* 18(3):381–402.

Whang, Taehee, Elena V. McLean and Douglas W. Kuberski. 2013. "Coercion, Information, and the Success of Sanction Threats." *American Journal of Political Science* 57(1):65–81.