Article

Collective Action and Constraints on Repression at the Endgame

Tiberiu Dragu[1] and Yonatan Lupu[2]

Abstract

How can human rights abuses be prevented or reduced? Using a simple game-theoretic model, we demonstrate that repression can become a coordination game when the potential for abuses is greatest: when dissent against a regime has grown sufficiently powerful. In such scenarios, repression depends on how the leader's agents coordinate on implementing a repression order. If and to the extent agents believe other agents will not comply with an order to repress, leaders can expect agents to disobey orders and will be less likely to order repression. This logic of expectations constitutes a third mechanism for constraining repression, in addition to sanctioning (i.e., the logic of consequences) and normative mechanisms (i.e., the logic of appropriateness). We formally explore how the logic of expectations can constrain the implementation of repression and also show that the logic of expectations has the greatest potential to constrain repression in middle regimes or "anocracies." In turn, this has broader implications for the strategies human rights advocates use in such regimes, how leaders structure their security forces, and for the study of why legal rules might be especially effective in such regimes.

[1]New York University, New York City, USA
[2]George Washington University, Washington, DC, USA

Corresponding Author:
Yonatan Lupu, Department of Political Science, George Washington University, 2100 G Street NW, Washington, DC 20052-0086, USA.
Email: ylupu@gwu.edu

Introduction

How can human rights abuses be prevented or reduced? Scholars have ana-
lyzed two mechanisms that can constrain state repression. Some focus on the
logic of consequences, or ways in which raising the cost of repression might
constrain leaders. Actors and institutions—such as governmental veto play-
ers, nongovernmental organizations (NGOs), and legal rules—can make
repression less likely, although these effects are conditional, probabilistic,
and unlikely to completely eradicate abusive practices (e.g., Conrad & Ritter,
2013; Dancy & Michel, 2015; DeMeritt, 2015; Fariss, 2014; Fariss &
Schnakenberg, 2014; Hill, 2010; Lupu, 2013b; Simmons, 2009). Others
focus on a set of related mechanisms that together are sometimes referred to
as the logic of appropriateness. Advocacy groups and other actors use persua-
sive and legitimizing power to change norms, identities, and preferences, cre-
ating societies in which human rights violations are deemed inappropriate or
even unthinkable (e.g., Finnemore & Sikkink, 1998; Sikkink, 2011). Yet
skepticism abounds over the effectiveness of these mechanisms. Critics
counter that institutional constraints often do not generate enough costs to
deter governments with incentives to repress, that advocacy groups lack
enforcement capabilities and therefore cannot impose such costs on govern-
ments, and that while social norms and preferences can change, they do so
slowly, often too slowly to prevent abuses (e.g., Posner, 2014).

We argue that a third mechanism, which we refer to as the logic of expecta-
tions, can constrain repression under certain conditions. Implicitly or explic-
itly, existing work tends to assume that the regime's repressive capacity—the
agents and other resources needed to conduct repression—is exogenous to
constraints on repression. Yet, when ordered to repress, members of the
regime's security apparatus do not always do so. To understand better when
leaders can be constrained from ordering repression, we also need to under-
stand when such leaders can expect their agents to follow orders to repress.
Scholars have begun to take this principal–agent problem into account
(DeMeritt, 2015; Svolik, 2013), but most existing work analyzes regime
agents as a unitary actor. Relaxing this assumption allows us to uncover situa-
tions in which the extent of repression can be thought of as a coordination
game among the regime's agents. In many situations, the extent to which
repression can be constrained crucially depends on how the leader's security
agents coordinate on implementing an order to repress. This is especially true

in scenarios in which there is the greatest potential for wide-scale abuses. In this context, repression depends in part on the extent to which security agents believe other security agents will follow orders. This is the logic of expectations.

To illustrate the logic of expectations mechanism and its relationship to other mechanisms by which repression can be constrained, we analyze a game-theoretic model of the relationship between leaders' orders to repress and the extent to which agents follow such orders.[1] The model suggests that repression can become a coordination game with multiple equilibria in situations in which the potential for human rights abuse is greatest: when dissent against a regime has grown sufficiently powerful such that a leader needs to resort to large-scale repression to keep power. We call these situations endgame scenarios (Pion-Berlin, Esparza, & Grisham, 2014). Especially in an endgame scenario, a leader cannot rely on a small number of highly trusted agents to keep her in power; if she orders repression, she needs to be able to rely on obedience by the common foot-soldiers. Yet agents may disobey orders in precisely these contexts. Sometimes they coordinate around firing on dissenters; at other times they coordinate on mutiny.

The primary contributions of this article are (1) to describe a theoretical mechanism for constraining repression, that is, the logic of expectations, and (2) to show that the logic of expectations has the greatest potential to constrain repression in middle regimes or "anocracies." While other work has analyzed leader–agent interactions and elite coordination problems in related contexts (Casper & Tyson, 2014; DeMeritt, 2015; Dragu & Polborn, 2013; Moore, 2000; Svolik, 2013), our article explores how the logic of expectations can constrain the implementation of repression and also describes the conditions in which it is most likely to do so. Our argument highlights the importance of agents' beliefs regarding what other agents are likely to do when ordered to repress, which has several important implications. The extent to which repression can be constrained depends not only on the extent to which actors can raise the cost of violations nor only on the extent to which actors can mobilize social action and effect long-term changes in social norms. A human rights complier, according to existing work, is either a compliant type or afraid of the possibility of being punished for repression. A human rights complier may also be one who believes he is surrounded by others who are either compliant types or who fear such punishments. Thus, government agents' beliefs regarding repression can be crucial in terms of constraining repression.

We analyze our formal model to explain the conditions under which the logic of expectation has the greatest potential to impose a constraint on repression. The first implication of this analysis is that the logic of expectations is

more likely to constrain leaders when the potential costs of repression are neither very low nor very high, for example, when norms and institutions for protecting human rights are relatively new. In other words, this suggests that focal factors are relatively most important in anocracies or middle regimes. Others have argued that structural factors, especially legal institutions, may be especially important constraints on repression in such regimes (Moravcsik, 2000; Simmons, 2009). One implication of our theoretical model is that the observed effects of legal constraints in such regimes may be due, in part, to the logic of expectations, a possibility that has not been explored in existing work. In addition, our theoretical model implies that, ex ante, actors in such countries will have the greatest incentives to use strategies to shape focal factors, a conjecture we explain more fully in the "Incentives for Human Rights Advocates" section.

The second implication is that the logic of expectations is more likely to constrain leaders in regimes that are neither too poor to reward agents for following orders nor too rich to provide benefits sufficiently lavish to outweigh concerns about what agents would do. Finally, we show that, under certain conditions, the constraining potential of this mechanism can be larger than those of other mechanisms. While empirical work has focused on the effects of consequentialist and normative constraints, our argument points toward expanding the empirical agenda to analyze whether and when the logic of expectations has constrained repression.

Our argument suggests that various actors will have incentives to attempt to shape the common knowledge environment that affects these beliefs to suit their interests. In the "Broader Ramifications" section, we discuss how this helps us better understand several observable phenomena regarding how leaders structure their security apparatus, how pro-human rights groups conduct advocacy, and the importance of framing in the context of social conflict. Explicating the logic of expectations thus allows us to uncover important links between the literatures on repression and dissent, ethnicity, civil–military relations, and human rights advocacy. We discuss in the "Broader Ramifications" and "Conclusion" sections the research agendas suggested and opened up by these connections. Finally, our model suggests that factors that shape agents' beliefs can be relatively more important than factors that shape the marginal costs of repression (e.g., institutional constraints and internalized norms) during periods in which institutions are new, weak, or unpredictable, or in which norms of respect for human rights have yet to be internalized.

The article proceeds as follows. In the "Leaders and Constraints on Repression" section, we briefly discuss the key insights and findings upon which we build. "The Logic of Expectations as a Constraint on Repression at

the Endgame" section describes our theory informally. In "The Coordination Game in the Endgame Scenario" section, we construct our formal model and derive several propositions. In the "Broader Ramifications" section, we discuss several implications of the logic of expectations for the incentives of leaders and other actors. In the "Conclusion" section, we conclude by offering the future research implications of our article.

Leaders and Constraints on Repression

Leaders prefer to stay in power, and domestic challengers threaten their grip on power. Among other risks, leaders face losing a reelection bid, being forced to resign after a loss of legitimacy, being deposed in a coup, and assassination (Rozenas, 2016). Leaders weigh these risks against the costs, appropriateness, and probabilities of success of their possible responses to dissent, which include repression, co-optation, accommodation, persuasion, and neglect (Gandhi & Przeworski, 2007; Moore, 1998, 2000). The benefits of repression can include reducing the opposition's mobilized resources, raising the cost of mobilization, and deterring potential challengers to the state (Conrad & Ritter, 2013; Davenport, 2007; Ritter, 2014; Tilly, 1978). Yet repression also comes with costs, including a reduction in the leader's coercive capacity, generating new grievances, a loss of domestic support or legitimacy, economic sanctions, loss of foreign aid and/or investment, and the possibility of legal sanctions (Dragu, 2017; Dragu & Polborn, 2014; Kaempfer, Lowenberg, & Mertens, 2004; Levitsky & Way, 2005; Moore, 2000; Sikkink, 2011; Simmons, 2009). Many leaders lose power after employing even the most brutal tactics.

Scholars have extensively studied two sets of mechanisms that can constrain decisions to repress. Focusing on the logic of consequences, many explore the possibility that leaders can be deterred from repressing by raising the costs of repression. Domestic actors and institutions, such as independent judiciaries, constitutional protections, and veto players, can, under certain conditions, either create ex ante barriers to repression or raise its ex post costs by imposing sanctions (Davenport, 2007; Hill & Jones, 2014; Ritter, 2014; Simmons, 2009). These domestic mechanisms can also increase the extent to which international treaty commitments reduce repression, although these effects are conditional and probabilistic (Conrad, 2014; Dai, 2014; Fariss, 2014; Hill, 2010; Lupu, 2013a, 2013b, 2015; Simmons, 2009; Smith-Cannoy, 2012).

Others analyze how the logic of appropriateness can constrain repression. NGOs and other advocacy groups rely on normative arguments to legitimize and strengthen their pressure on governments to refrain from repression

(Bell, Clay, & Murdie, 2012; Carpenter, 2007; Clark, 2010; Keck & Sikkink, 1998; Murdie & Bhasin, 2011; Murdie & Davis, 2012; Wong, 2012). NGOs and media outlets focus on collecting and diffusing information about repression (Keck & Sikkink, 1998), which facilitates mobilization against governments (Dancy & Michel, 2016; Keck & Sikkink, 1998; Simmons, 2009), naming-and-shaming activities (DeMeritt, 2012; Krain, 2012; Murdie & Davis, 2012), and legislative opposition (Lupu, 2015). These activities can change social norms and preferences over time (Clark, 2010; Finnemore & Sikkink, 1998; Lutz & Sikkink, 2000; Simmons, 2009). Yet some are skeptical that advocacy can have meaningful effects on the extent to which governments use repression. For example, Posner (2014) claims that "this argument does not survive scrutiny. . . . Occasionally, boycotts and other forms of pressure follow from these efforts, but their overall effectiveness is clearly limited" (p. 82).

When ordering repression, leaders also consider whether they expect their orders to be obeyed. Leaders do not conduct repression themselves, but instead pursue such policies by using the government's repressive apparatus, which can include regular police, the military, and intelligence services. In discrete or small-scale situations, leaders can rely on a small number of highly trusted agents to repress dissidents. By contrast, when the opposition has grown sufficiently powerful and numerous citizens are protesting, leaders need to rely on large-scale coordinated repression to stay in power. In many such historical cases, police have shirked their duties, military leaders have refused to mobilize in support of the regime, soldiers have refused orders to fire on or otherwise repress opposition members, and individual agents of government coercion have aided or defected to the opposition (Johnson, 1982; Pion-Berlin et al., 2014; Pion-Berlin & Trinkunas, 2010). In endgame scenarios, the likelihood of the leader ordering repression in the first place may be reduced if a leader expects his agents will not follow his orders or not follow them fully.

The Logic of Expectations as a Constraint on Repression at the Endgame

This section develops our theory of how the logic of expectations can constrain human rights abuses when leaders need to rely on large-scale coordinated repression to keep power. In these endgame scenarios, several mechanisms can limit such repression. Repression can be deterred by increasing its costs (i.e., the logic of consequences) and by changing social norms (i.e., the logic of appropriateness). Yet repression can also be constrained by the logic of expectations, that is, the extent to which government agents

believe other agents are likely to comply with orders to repress. The extent to which the logic of consequences constrains repression depends on the strength of potential costs imposed on human rights abusers, while the extent to which the logic of appropriateness does so depends on the strength of norms against such abuses. By contrast, the extent to which the logic of expectations mechanism constrains repression depends on the extent to which government agents expect other government agents to disobey orders to repress; among other factors, such expectations can be shaped by the belief that other agents fear the consequences of repression or have internalized norms against it.

Before turning to our formal model, we lay out its basic structure and intuitions. We focus on situations in which an opposition movement has grown powerful enough that the leader's options are limited to either attempting to put down the opposition with large-scale repression or giving up power. In such scenarios, leaders who order repression must rely on the rapid obedience of many repressive agents. These are typically cases in which the potential for the most widespread abuses is large. Leaders sometimes decide against ordering repression. For example, Eduard Shevardnadze resigned as President of Georgia in 2003 following demonstrations against his rule, although some evidence suggests he or members of his government considered using force (Welt, 2010, pp. 175-178). Likewise, Mahinda Rajapaksa resigned as Prime Minister of Sri Lanka in 2015 following an electoral loss. Rajapaksa is alleged to have considered a coup attempt to stay in power, having stationed army troops outside Colombo, but ultimately decided to cede power peacefully ("Sri Lanka to Investigate 'Rajapaksa Coup Plot,'" 2015).

When leaders order repression, their agents must decide the extent to which they will obey the order, that is, the level of effort they will devote to repression or the amount and severity of repression. In some cases, repressive agents have famously obeyed such orders, such as the troops who fired on demonstrators in Austria and Prussia during the March Revolution of 1848. Yet history is also full of examples of repressive agents disobeying orders, abandoning their posts, shirking their responsibilities, and defecting to the opposition. During the Arab Spring, leaders in Tunisia and Egypt were brought down in part because militaries refused to crush opposition movements (Nepstad, 2013). In nine of the 13 cases between 1990 and 2010 in which militaries in Latin America were ordered to intervene against civilian opposition forces, the military refused to obey such orders (Pion-Berlin & Trinkunas, 2010, p. 395).

Table 1 provides examples of endgame scenarios since the year 2000. These are cases in which mass dissent was sufficiently potent so as to make it likely that mass repression would have been required for the leader to

Table 1. Examples of Endgame Scenarios Since 2000.

Leader orders mass repression and agents generally obey	Leader orders mass repression and agents generally disobey	Leader does not order mass repression
Bolivia (2003)	Argentina (2001)	Philippines (2001)
Malaysia (2007)	Ecuador (2005)	Georgia (2003)
Iran (2009)	Tunisia (2010)	Ukraine (2004)
Bahrain (2011)		Sri Lanka (2015)

remain in power. In some cases, such as Bolivia (2003) and Malaysia (2007), the leader ordered repression, the regime's agents generally obeyed such orders, and the leader was able to remain in power. In a second set of cases, such as Ecuador (2005) and Tunisia (2010), a sufficient number of regime agents generally disobeyed orders to repress, so the leader could not hold on to power. Finally, in a third set of cases, such as Sri Lanka (2015), the leader stepped down without ordering mass repression.[2]

When a leader considers whether or not to order repression, he or she must take into account the extent to which agents may follow the order. Leaders who order mass repression but nonetheless lose power can face high social and legal costs, including informal retribution and formal transitional justice mechanisms that could be implemented after a regime change. All else equal, the lower his expectation of obedience, the less likely the leader will be to order repression. The extent to which security and military forces respond to a potential order to repress can thus affect how likely it is that the leader would order repression. If soldiers refuse to use force against demonstrators, the leader is constrained in his policy options and his grip on power is likely to fade. On the contrary, if soldiers agree to intervene forcefully against the demonstrators, the leader is less constrained in his policy options.

Why might agents devote low levels of effort to repression or completely withhold from repression? Agents are concerned with several factors. The first is the benefit they expect to receive from repression. This can include rewards from the leader (e.g., promotion, increases in pay, and other private goods) if they follow orders.[3] Such promises will be fulfilled only if the leader maintains power. Some agents may also have deeply held ideologies against the relevant opposition movement and may therefore gain an internal reward from repressing it.

Second, agents are concerned with the possibility of costs they may incur if they do follow through with repression. These costs can come from many sources and manifest in different ways. Some agents may have a personal or normative aversion to conducting repression, which we can think of as

raising their cost of following orders to repress. Agents might also expect that if they follow orders to repress and the leader nonetheless loses power, they may be subject to some form of informal or formal retribution, such as legal sanctions. One example is a criminal penalty resulting from the prosecution of past illegalities, which may occur in the aftermath of the regime change, such as posttransition prosecutions in Chile and Argentina. Today, such agents also face the possibility of other forms of transitional justice, including domestic trials, truth commissions, ad hoc tribunals, prosecution by the International Criminal Court, prosecution by the national courts of other states under the doctrine of universal jurisdiction, and combinations of these institutions (Olsen, Payne, & Reiter, 2010; Sikkink & Walling, 2007; Wiebelhaus-Brahm, 2010). In recent history, the potential legal costs have played an important role in several decisions by national militaries to disobey orders to repress uprisings, including Serbia (2000), Argentina (2001), Ukraine (2004), and Tunisia (2010-2011) (Pion-Berlin et al., 2014). Other potential costs for following orders to repress include career costs such as losing a position or possible promotion.

The extent to which agents will reap the benefits and incur the costs of following the leader's orders depends on whether the leader remains in power. For example, if the leader has been deposed, he or she is unlikely to be able to reward the agents with promotions. Similarly, if a repressive leader remains in power, legal sanctions are less likely to be imposed on those who helped him or her do so. The possibility of the leader losing power is far from trivial. Svolik (2009) finds that 205 of 303 (68%) authoritarian leaders were deposed between 1945 and 2002.

Agents understand this and make decisions accordingly. When the Volhynian regiment refused the order to fire on demonstrators in 1917 Petrograd, it did so in part because Russia's performance in World War I led soldiers to believe the Tsar was unlikely to hold on to power (Johnson, 1982, p. 105). Similarly, Bushnell (1985) documents 202 mutinies by Russian troops during the Revolution of 1905-1906, arguing that soldiers, responding to political developments during the revolution, mutinied based on changing beliefs about the Tsar's likelihood of staying in power:

> All that these events had in common was the construction that soldiers put on them: they contained, for soldiers, unambiguous messages about the authority of the Tsarist regime. When they believed the regime's writ had expired, soldiers mutinied. When they believed the regime's authority to be intact, they repressed civilians. (p. 226)

A more recent example occurred in Argentina in December 2001. A crippling economic crisis led to violent civil unrest and mass riots demanding the

ouster of President Fernando de la Rúa. When the police could not contain the demonstrators, the President asked the military to intervene, but the military refused. The military was wary of entering the fray because of the potential penalties that could be imposed on it (and individual members) under Argentine law, especially if the government fell. During the Dirty War of the 1970s and 1980s, the military had famously obeyed government orders to violently repress opposition movements. In the aftermath, many of its members faced legal penalties. In part, the military's sensitivity to these costs was driven by its experiences in the prior decades and in part by its expectation that de la Rúa would not be able to remain in power. Without the military's support, de la Rúa was forced to resign. As Pion-Berlin and Trinkunas (2010) note, "To have obeyed would have meant bloodshed, and a new administration taking over would be less sympathetic to the perpetrators of state violence. . . . The military calculated correctly. No harm would come to it with succeeding administrations" (pp. 403-404).

One of the key factors that affect whether the leader remains in power is the collective level of effort that agents use in repression. As Arendt (1972) notes, "Power belongs to a group and remains in existence only so long as the group keeps together" (p. 143). If agents are not likely to follow the order, the probability of losing power increases; if security agents effectively repress the opposition, the leader is more likely to hold on to power. Thus, while repressive agents *individually* make decisions based on an expected probability of the leader remaining in power, *collectively* their joint repressive effort can change this probability.

Individual agents, therefore, must try to gauge what their peers and colleagues are likely to do.[4] In this type of situation, repression can be thought of as a coordination game.[5] The agents may coordinate on following the leader's order to repress or they may coordinate on disobedience. At this point in the process—after the order for mass repression is given—they coordinate based on their beliefs about what other agents may do (for a similar point, see McLauchlin, 2010). Depending on security agents' beliefs about the potential actions of other security agents, multiple levels of repressive efforts are possible, resulting in a set of possible amounts and severity of repression. While this type of coordination problem can arise in scenarios of small-scale repression, we focus on endgame scenarios in which the problem is likely to be especially severe. In endgame scenarios, the leader cannot rely on a small number of trusted, easily rewarded, and easily monitored agents, but rather has to resort to large-scale, coordinated repression to keep power.

At this stage, the extent of agents' direct costs and benefits of repression are not the sole determinants of the equilibrium outcome. Agents also coordinate on an equilibrium based on the logic of expectations. They obey the

order to repress if they believe others will do so, or disobey if they believe others will do so. What matters in this context, therefore, is not only the extent to which an agent expects to face consequences for repression and the extent to which an agent has internalized norms of respect for human rights; instead, the key factor can be the extent to which an agent believes *other* agents expect to face such consequences and/or have internalized such norms. The logic of expectations, therefore, is based on common beliefs about the extent to which other actors are affected by mechanisms associated with the logics of consequences and appropriateness.

Agents' beliefs about what other agents may do in these scenarios can therefore have important effects on repression. This mechanism can constrain orders to repress in ways that are distinct from (but build upon) mechanisms that affect the direct costs and benefits of repression, but the effects of this mechanism depend on agents' beliefs. In turn, this argument implies that actors will have important incentives, ex ante, to shape regime agents' beliefs to suit their interests. Leaders who wish to keep the repression option open have incentives to try to lead their agents to believe other agents would comply with such orders. Likewise, human rights advocates and other forces opposed to the government will try to cause government agents to believe other agents are unlikely to abuse human rights. These incentives have important implications we discuss in more detail in the "Broader Ramifications" section.

The Coordination Game in the Endgame Scenario

In this section, we formalize the strategic interaction between a leader and his security apparatus when the leader needs to resort to repression to maintain power. In this endgame scenario, there are two types of players: a leader and (a continuum of) security agents. The leader makes a binary decision, $r \in \{0,1\}$, where 1 means that the leader orders repression and 0 means that the leader chooses not to order repression. If the leader orders repression, each agent chooses a level of repressive effort $e_i \in [e_l, e_h]$ to implement the leader's order.

The probability that the leader maintains power is given by a function $\Pi(\bar{e})$, which depends on the average level of repressive effort $\bar{e}$. Intuitively, the probability of the leader staying in power increases if the average level of repressive effort is higher, that is, $\Pi(\bar{e})$ increases in $\bar{e}$.[6]

If the leader maintains power, security agent i receives a benefit $B_A(e_i)$ for implementing the leader's repression decision, a benefit which is larger when the agent exerts more repressive effort. Therefore, $B_A(\cdot)$ is increasing in e_i and exhibits decreasing marginal returns in e_i (i.e., $B'_A > 0$ and

$B_A'' < 0$). If the leader loses power, agent i suffers a cost $C_A(e_i)$ for implementing the repression order because the leader cannot guarantee with certainty that there will be no future consequences for implementing a repressive policy. Therefore, the cost $C_A(\cdot)$ is increasing in e_i and is convex in e_i ($C_A' > 0$, $C_A'' > 0$).

As the previous discussion suggests, the cost $C_A(e_i)$ can have several sources and manifestations, including career costs, norm violation, or legal sanctions. We do not assume that agent i will definitely pay a cost for his repressive effort, but only that the agent can never be certain, at the time of executing the repression order, that there would be no future consequences for his actions. We make no assumption about the magnitude of the cost $C_A(e_i)$; it can be low or high, depending on the strength of the mechanisms for imposing such costs.

Given this, a security agent i's (expected) utility is the following:

$$U_A(e_i, \overline{e}) = \Pi(\overline{e}) B_A(e_i) - [1 - \Pi(\overline{e})] C_A(e_i). \tag{1}$$

Let $U_L(0)$ represent the leader's payoff if the leader decides not to resort to repression (i.e., $r = 0$). Without loss of generality, we normalize this payoff to 0, that is, $U_L(0) = 0$. If the leader chooses repression (i.e., $r = 1$), the leader receives a benefit $B_L > 0$ if the leader maintains power. However, even if the leader orders repression, he or she might not maintain power, in which case the leader might pay a cost $C_L > 0$ if power is lost. Also, the leader's payoff if he orders repression depends on the security agents' level of effort because the probability of keeping power is a function of the repressive effort. Therefore, the leader's (expected) payoff from ordering repression is

$$U_L(1, \overline{e}) = \Pi(\overline{e}) B_L - [1 - \Pi(\overline{e})] C_L. \tag{2}$$

To ensure that the leader's optimal choice depends on the security agents' level of repressive effort, we assume that $U_L(1; e_l) < 0$ and that $U_L(1; \hat{e}) > 0$ for some $\hat{e} \in (e_l, e_h]$. That is, the leader's payoff is lower if (a) he orders repression, and the security agents exert the lowest level of repressive effort than if (b) the leader does not order repression. Also, the leader's payoff is higher if (a) he orders repression, and the security agents choose some level of repressive effort $\hat{e} > e_l$ than if (b) the leader does not order repression.[7] Without these simple assumptions, there are no trade-offs for the leader— The leader would choose to order repression or not to do so regardless of the level of repressive effort security agents choose.

The timing of the game is as follows. First, the leader decides whether or not to order repression, $r \in \{0,1\}$. Second, if the leader orders repression (i.e., $r = 1$), the security agents simultaneously choose their respective level of repressive effort, $e_i \in [e_l, e_h]$.

The Repression Subgame

The security agents choose how much effort to exert if the leader orders them to repress, that is, $r = 1$. Maximizing agent i's objective function implies that her (interior) optimal action is the solution to the first-order condition[8]:

$$\Pi(\bar{e}) B'_A(e_i) - [1 - \Pi(\bar{e})] C'_A(e_i) = 0. \tag{3}$$

Because $U_A(e_i, \bar{e})$ is strictly concave in e_i, there is a unique optimal e_i for any given $\bar{e}$. As a consequence, a security agent i has a well-defined best response function to the average level of effort $\bar{e}$, which we denote by $\phi(\bar{e})$. Moreover, because Equation 3 is continuous in $\bar{e}$, the best response function $\phi(\bar{e})$ is continuous in $\bar{e}$, and we can apply the implicit function theorem to find the slope,

$$\phi'(\bar{e}) = -\frac{\Pi'(\bar{e}) \left[B'_A(e_i) + C'_A(e_i) \right]}{\Pi(\bar{e}) B''_A(e_i) - \left[1 - \Pi(\bar{e}) \right] C''_A(e_i)} > 0,$$

when agent i's optimal action is interior. The above expression indicates that the marginal payoff of agent i's effort is increasing in $\bar{e}$ for an (interior) optimal action, which implies that the interaction among the security agents exhibits strategic complementarities.[9] The solution of $\phi(\bar{e}) = \bar{e}$ for $\bar{e} \in [e_l, e_h]$ is a pure-strategy equilibrium, and given that $\phi(\bar{e})$ is nondecreasing in $\bar{e}$ (and strictly increasing if agent i's optimal action is interior), the existence of a pure-strategy equilibrium follows from applying Tarski's fixed-point theorem in our context (Topkis, 1979; Vives, 1990).[10] Moreover, because the repression subgame is symmetric (i.e., exchangeable against permutations of players) and each security agent has a unique best response to any given $\bar{e}$, all pure-strategy equilibria are symmetric.[11] We have the following result:

> **Proposition 1:** The repression subgame has a (symmetric) pure-strategy Nash equilibrium.

Given that the equilibrium level of effort is also the average level of effort, that is, $e^* = \bar{e}$, substituting this in Equation 3, an (interior) equilibrium level of repressive effort is the solution to

$$\Pi\left(e^*\right)B_A'\left(e^*\right)-[1-\Pi\left(e^*\right)]C_A'\left(e^*\right)=0.$$

In this context, multiple equilibrium levels of effort e^* exist in the repression subgame if the strategic complementarities among the security agents' actions are strong—a sufficient condition for multiple equilibria to exist is that the slope of the best response function $\phi'(e^*)>1$ for some candidate equilibrium $e^* \in (e_l, e_h)$.[12] When multiple equilibria exist, the equilibrium set has "extremal" (i.e., smallest and largest) pure-strategy equilibria, and these equilibria can be ranked in the sense that the equilibrium with higher effort Pareto dominates the lower effort equilibrium (Milgrom & Roberts, 1990).

The multiplicity of equilibria if the leader orders repression is the result of a coordination game among security agents predicated upon the fact that the leader's grip on power depends on how much repressive effort the agents collectively put into executing a potential order to repress. If all security agents exert high repressive effort, the leader is more likely to maintain power, that is, $\Pi(e^*)$ is larger. Thus, the security agents' potential cost for repressive effort may decrease substantially while their potential benefits increase. This implies that the individually optimal level of repressive effort is also larger. In contrast, if security agents coordinate on a lower repressive effort level, then the probability that the leader is out of power is larger, that is, $1-\Pi(e^*)$ is larger. Thus, security agents' potential costs may increase substantially while their potential benefits decrease. This implies that the individually optimal level of repressive effort is also lower.

The Leader's Repression Decision

Given the optimal behavior of agents, we next analyze the leader's optimal decision. If the leader does not order repression, that is, $r=0$, then the leader's payoff is 0, and if the leader orders repression, his payoff is $\Pi(e^*)B_L -[1-\Pi(e^*)]C_L$, where e^* is the agents' equilibrium level of effort in the repression subgame. Therefore, the leader's optimal choice is $r=1$ if

$$\Pi\left(e^*\right)B_L \geq \left[1-\Pi\left(e^*\right)\right]C_L, \tag{4}$$

and $r=0$ otherwise. Equation 4 has an intuitive interpretation: The leader orders repression if the expected benefit exceeds the expected cost where the expectation depends on the probability that the leader maintains power (which in turn depends on the equilibrium level of agents' repressive effort). We have the following result:

Proposition 2: The leader's equilibrium choice is

$$
r^* = \begin{cases} 1 & \text{if } \dfrac{C_L}{B_L} \leq \dfrac{\Pi\left(e^*\right)}{1-\Pi\left(e^*\right)} \\[4mm] 0 & \text{if } \dfrac{C_L}{B_L} > \dfrac{\Pi\left(e^*\right)}{1-\Pi\left(e^*\right)} \end{cases}.
$$

Taken together, Propositions 1 and 2 complete the characterization of the game. Next, we discuss some implications of our analysis. In particular, the importance of coordination among security agents when the leader orders repression has novel implications for understanding various mechanisms for constraining repression.

Mechanisms for Constraining Repression

The level of effort agents choose can constrain the leader's initial order to repress. When there are multiple equilibria in the repression subgame, the leader's decision of whether or not to order repression crucially depends on what level of effort the security agents coordinate on. For example, suppose that if the leader orders repression, there are two possible equilibrium levels of repressive effort, e_L^* and e_H^*, which implies that the probability that the leader stays in power can be either $\Pi(e_L^*)$ or $\Pi(e_H^*)$. Consequently, given identical underlying fundamentals of the strategic interaction (i.e., the same potential costs and benefits for ordering repression, B_L and C_L), the leader may rationally choose different strategies (i.e., repression or no repression) depending on whether security agents coordinate on the low or on the high equilibrium level of repressive effort.

The foregoing also demonstrates the relationship between multiple mechanisms for constraining repression. As others have argued, repression can sometimes be constrained by raising its costs or by changing social norms. Yet a security agent's choice of a specific level of repressive effort is not solely determined by the costs/benefits of such a course of action but also determined by her expectation of what other agents might do. To illustrate this point, consider a scenario in which the leader orders mass repression. Engaging in such actions might have potential costs that could be due both to ex post sanctions or to the fact that security agents have internalized (to some extent) norms against repression or to a combination of both. Regardless of the sources of these costs, for any given $C_A(\cdot)$, costs alone cannot fully explain the level of

repressive effort a security agent chooses. In the coordination problem faced by agents ordered to conduct mass repression, the equilibrium outcome (as is often the case in coordination problems) will also depend on agents' beliefs about what other agents will do. Agents are more likely to disobey orders to repress when they believe other agents will do the same, which constitutes a third mechanism, a logic of expectations, that can constrain repression.[13]

The previous discussion suggests that a security agent's beliefs about what others might do can become critical. In what follows, we explore the conditions under which this logic of expectations may be at work and also its relationship with other mechanisms of constraining repression. First, we investigate the necessary conditions for the logic of expectations to be a relevant mechanism for constraining repression. We have the following result:

> **Proposition 3:** The repression subgame has a unique equilibrium level of effort (i.e., $e^* = e_h$) if the (marginal) cost of repression is sufficiently low, $C'_A(e_h) < \underline{C}$. The repression subgame has a unique equilibrium level of effort (i.e., $e^* = e_l$), if the (marginal) cost of repression is sufficiently high, $C'_A(e_l) > \overline{C}$.

Proposition 3 indicates that a necessary condition for the logic of expectations to be at work is that the (marginal) cost of repression is not too low and not too high (i.e., $C'_A(e_h) > \underline{C}$ and $C'_A(e_l) < \overline{C}$).[14] First, this result indicates that the logic of expectations is more likely to be relevant when the security agents' cost for repressive effort is not so large as to ensure that they choose the lowest level of repressive effort regardless of what other agents do. In turn, this suggests that agents' beliefs about how other agents may react when ordered to repress may be especially important when institutions are new or when the process of norm internalization is in its early stages. This can also be interpreted as indicating that such beliefs may be relatively more important in polities that lack fully credible institutions to punish repressive state actors. If and to the extent institutions have succeeded in imposing sanctions and/or actors have internalized norms of compliance (which might be the case in fully democratic countries), these factors may be sufficient to ensure that orders to abuse human rights are disobeyed (i.e., the cost of repression is too high). Yet when institutions are relatively new and efforts to change norms have not begun to cascade, the logic of expectations still has the potential to constrain repression to the extent agents believe other agents will disobey orders to repress.

Second, Proposition 3 suggests that the logic of expectations is more likely to be relevant when the security agent's cost for repressive effort is not so low that a security agent will choose the highest level of repressive effort

regardless of what other security agents do. In turn, this suggests that agents' beliefs about how other agents may react when ordered to repress may not be important in polities in which institutions for punishing repressive and illegal behavior are nonexistent or not functioning at all. Putting together the two necessary conditions for multiple equilibria to exist, Proposition 3 suggests that the logic of expectations is more likely to be relevant in anocracies or "middle regimes," polities in which the institutions for imposing sanctions for illegal behavior are not too weak (so they are not completely irrelevant) but, at the same time, are relatively new and not too strong; thus, the extent to which they will actually be effective is uncertain. This proposition is particularly important in light of the influence of existing arguments that legal institutions may have the greatest effect on repression in such regimes (Moravcsik, 2000; Simmons, 2009). Existing work, however, focuses on the logics of consequences and appropriateness to explain the relationship between regime type and repression; our argument suggests that the logic of expectations mechanism may help explain why legal rules constrain repression in these regimes.

We also derive a set of necessary conditions on the benefits of repression, that is, $B_A(\cdot)$, to further delineate when the logic of expectations is more likely to be relevant. We have the following result:

> **Proposition 4:** The repression subgame has a unique equilibrium level of effort ($e^* = e_l$) if the (marginal) benefit of repression is sufficiently low, that is, $B'_A(e_l) < \underline{B}$. The repression subgame has a unique equilibrium level of effort ($e^* = e_h$) if the (marginal) benefit of repression is sufficiently high, that is, $B'_A(e_h) > \overline{B}$.

Proposition 4 implies that a necessary condition for the logic of expectations to be a constraining mechanism is that the benefits of engaging in repression for security agents are neither too small nor too large, that is, $B'_A(e_l) > \underline{B}$ and $B'_A(e_h) < \overline{B}$.[15] If the benefits of repression are too small, then a security agent may not have sufficient incentive to exert a repressive effort higher than e_l regardless of what other security agents do. This can occur, for example, when a regime does not have the resources to offer its agents sufficient rewards. Likewise, if the benefits of repression are too large, the coordination incentive does not enter a security agent's calculus because he or she will have the incentive to repress to reap those benefits regardless of what others do. This can be the case when the regime has sufficient resources at its disposal to provide lavish rewards to agents who follow orders. Thus, an interpretation of Proposition 4 is that, all else equal, the logic of expectations is likely to be relevant in authoritarian regimes that are in the "middle range" of wealth distribution. Agents in the poorest regimes

may have too little incentive to follow orders; while the potential benefits to agents in the richest regimes may be large enough to outweigh concerns about what others would do.

The previous analysis illustrates the necessary conditions for the logic of expectations to be a relevant mechanism for constraining the leader's decision to repress. Furthermore, our model allows us to assess the relative impact of the logic of expectations on the leader's repression decision. That is, the model allows us a simple way to analyze the extent to which the constraints on the ruler's decision to (not) order repression derive from factors that shape agents' beliefs versus other factors (including norms and institutions that raise the cost of repression).

Recall that the leader's optimal strategy is to order repression when $(C_L / B_L) \leq (\Pi(e^*)/1 - \Pi(e^*))$. Note that the expression $(\Pi(e^*))/(1 - \Pi(e^*))$ increases in e^*, which implies that the range of parameters (i.e., C_L / B_L) for which the leader will choose repression is larger when security agents coordinate on a higher equilibrium level of effort. As such, $(\Pi(e^*))/(1 - \Pi(e^*))$ can be taken as a measure of how constrained a leader is in his decision to order repression; a larger value means that the leader is less constrained in ordering repression because the leader's optimal choice is to repress. To illustrate this, consider a situation in which there are two possible equilibrium levels of effort in the repression subgame: e_L^* and e_H^*. In this context, we can think of $(\Pi(e_L^*))/(1 - \Pi(e_L^*))$ as the measure of the constraint on the leader's decision to repress that would occur with certainty given structural factors (i.e., the costs and benefits of repressive effort), and we can think of $((\Pi(e_H^*))/(1 - \Pi(e_H^*))) - ((\Pi(e_L^*))/(1 - \Pi(e_L^*)))$ as the measure for the constraint on the leader's repression decision due to the logic of expectations. As the example below shows, the former can be larger than the latter, which implies that factors that affect how agents coordinate beliefs can have larger effects in constraining the leader's decision to repress. Figure 1 shows such a situation.

To further illustrate the argument that the logic of expectations can, in some situations, be the most important factor in inducing a leader to not order repression, consider the following parametric example. Let a security agent's benefit for repressive effort be given by $B_A(e_i) = (3/8)e_i$ and the cost for repressive effort be given by $C_A(e_i) = (1/2)e_i^2$. Also, let the probability that the leader maintains power be $\Pi(\bar{e}) = \bar{e} + (1/8)$, and each agent i's choice of effort be $e_i \in \left[\dfrac{1}{8}, \dfrac{1}{2} \right]$.

Given these specifications, there are two equilibria levels of effort in the repression subgame: $e_L^* = 1/8$ and $e_H^* = 1/2$.[16] If the equilibrium level of repressive effort is $e_L^* = 1/8$, then $(\Pi(1/8))/(1 - \Pi(1/8)) = 1/3$, and if the equilibrium
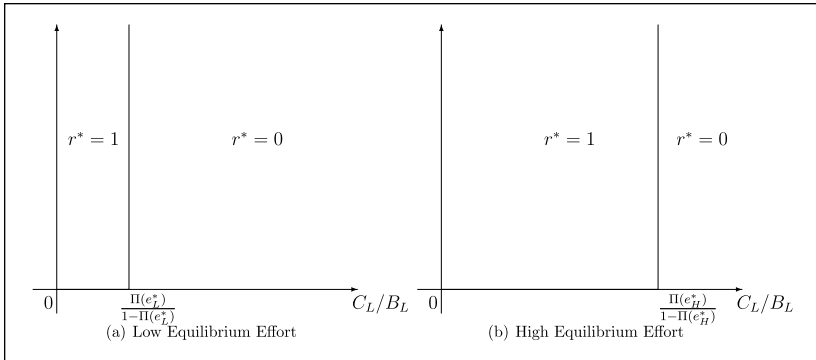
Figure 1. Multiple equilibria and repression.

level of repressive effort is $e_H^* = 1/2$, then $(\Pi(1/2))/(1-\Pi(1/2)) = 5/3$. Thus, $((\Pi(e_H^*))/(1-\Pi(e_H^*)))-((\Pi(e_L^*))/(1-\Pi(e_L^*))) = 4/3$, which suggests that factors that affect how agents coordinate their beliefs can be more important in constraining the leader's decision to repress in this example.

Broader Ramifications

Our model indicates that repression in endgame scenarios can depend on agents' beliefs regarding what other agents are likely to do. In turn, this means that activities that can shape such beliefs can have important effects on the choice of equilibria and level of repression. The potential importance of the logic of expectations thus creates incentives for actors both within and outside the regime to, ex ante, try to shape the common knowledge environment that affects agents' beliefs in endgame scenarios. That is, because an endgame scenario may happen in the future, actors have incentives well before such a scenario arises to either try to make compliance with orders to repress focal or to try to make respect for human rights focal. This has several observable implications which we discuss below. In so doing, we draw on our argument to make potentially important links between the literatures on repression and dissent, ethnicity, civil–military relations, and human rights advocacy.

Incentives for Human Rights Advocates

Our argument suggests that human rights advocates would rationally seek to shape security agents' beliefs such that agents coordinate around disobeying orders to repress. The logic of expectations has two implications with respect

to human rights advocacy. The first is that the effects of human rights advocacy on repression may not be limited to the ability of advocates to impose costs on violators or shape social norms; advocacy can also have an effect by making respect for human rights focal. Second, our argument leads us to conjecture that human rights advocates are more likely to adopt strategies intended to influence focal, rather than structural, factors in regimes in which focal factors are relatively more important, that is, middle regimes.

The effects of advocacy are not limited to persuasion, norm internalization, and shaming. Our argument implies that advocacy may be able to constrain repression by shaping agents' beliefs about what other agents are likely to do when ordered to repress. Where do these beliefs come from? Beliefs are a self-enforcing equilibrium in the context of obeying orders (Myerson, 2004). These beliefs can come from several focal factors in the common knowledge environment or culture. Institutions and third-party messages can be effective at leading expectations to converge around a set of beliefs (Hadfield & Weingast, 2012; McAdams & Nadler, 2005).

Human rights advocacy may affect respect for human rights by serving as third-party messaging that can lead actors to expect other actors to respect human rights. Advocates have a repertoire of tactics to choose from, and some activities are aimed more at influencing structural factors, while others are aimed more at focal factors. While a logic-of-appropriateness advocacy campaign may have a message along the lines of "Abusing human rights is wrong," the message of a logic of expectations campaign would be more along the lines of "Your peers believe abusing human rights is wrong."

There is already evidence that advocacy groups use strategies aimed at focal factors. As Wong (2012) shows, advocacy can make compliance with human rights norms into focal points in some societies. Such focal points can be especially important in endgame scenarios in which security agents need to coordinate to implement a leader's order to repress public protests and demonstrations. Likewise, as Welch (2001) argues, NGOs often focus on creating a culture of respect for human rights (p. 50). Educating members of the regime's security apparatus about human rights norms may be especially important, and many National Human Rights Institutions (NHRIs)—such as those in India, Nepal, and Afghanistan—have adopted this strategy. These activities focus not only on persuading individuals that repression is wrong (which would be consistent with the logic of appropriateness) but also on collective activities designed to lead individual agents to believe that other agents have internalized such norms (which is consistent with the logic of expectations). Likewise, advocacy groups often attempt to make security agents doubt whether other agents would comply with future orders to

repress, such as publicizing past disobedience and examples of security agents who espouse human rights norms.

Our formal model indicates that, in relative terms, focal factors have the potential to constrain repression most in middle regimes, that is, regimes that are neither fully authoritarian nor democratic. Assuming that human rights advocates behave strategically, this implies that they would use strategies aimed at focal factors relatively more in situations in which such factors may be more important. While testing this conjecture is beyond the scope of this article, we hope it will lead scholars to analyze in depth the relationship between focal factors, regime characteristics, and strategies for human rights advocacy.

Government Incentives

Like human rights advocates, governments have incentives to shape their agents' beliefs, that is, to lead agents to believe their peers would likely follow orders to repress. This has several implications. First, these incentives may contribute to how governments structure their security apparatus, including strategies already identified by work in other literatures. Second, the incentive to shape such beliefs suggests other strategies that have been underexplored in existing work. Finally, these incentives create a dilemma for leaders: Leaders seek to maximize coordination around following orders while minimizing coordination around disobedience.

The literatures on civil–military relations and coup-proofing have long recognized that leaders structure their security apparatus with an eye toward securing obedience with potential future orders (Feaver, 2009; Huntington, 1957; Janowitz, 1960; Staniland, 2008). Our model may help enrich our understanding of the underlying mechanisms that explain phenomena already identified by those literatures. For example, leaders often create multiple layers of security forces with differing command structures to reduce the danger that might arise from one group not obeying orders (Quinlivan, 1999). With multiple forces present, agents in these differing forces may be embedded in different common knowledge environments, may come from differing ethnolinguistic groups, or may otherwise have little contact with each other, all of which can make it less likely that they can coordinate on disobedience if an endgame scenario were to arise. An example is Saddam Hussein's Republican Guard, an elite force of Sunni troops, on whom Hussein could count to obey the order to repress the Shia rebellion in 1991 to 1992. Likewise, leaders have an incentive to employ agents who are closely identified with the regime, potentially along ethnic, tribal, or religious lines (Quinlivan, 1999). Such individuals are especially likely to suffer costs if the regime falls. They are

also more likely to believe that their fellow agents would themselves face such consequences and, therefore, to believe that their fellow agents would follow orders to repress. As McLauchlin (2010) argues, when regime agents have been recruited from an "in-group," they are more likely to believe that other agents around them will be loyal to the regime, a point that is consistent with the logic of expectations mechanism.

Yet the logic of expectations also implies that leaders will, ex ante, use other strategies to make compliance with orders focal. First, rational leaders should attempt to shield their agents, to the extent possible, from advocacy and education activities because leaders have the incentive to focalize agents' beliefs around obedience. Leaders often go to great efforts to control the information their security agents can access (Talmadge, 2015), and our model helps to explain these efforts. In a recent example, Syria's Bashar al-Assad created a housing complex in a suburb of Damascus specifically for military officers, a strategy that may have been useful not only to create camaraderie but also to shield these officers from information contrary to Assad's interests (Khaddour, 2015).

Second, leaders who want to keep the repression option open rationally choose to recruit agents who are less likely to disobey orders to repress and, thus, agents who are less likely to believe their fellow agents are likely to disobey such orders. One implication of this is that leaders may prefer to use security agents who are not part of the same common knowledge environment as the individuals they may be ordered to repress. A small set of leaders solve this problem by employing agents from outside their country. Bahrain, for example, relies on a foreign mercenary Sunni army to prevent and put down potential Shia uprisings, as they did in 2011.

Finally, leaders know that history can be a source of beliefs and expectations. Agents who know that, in similar contexts, the leader's orders have been disobeyed in the past may believe others will disobey the present order—and vice versa. Symbols, breaking points, and other dramatic and highly publicized events can change such beliefs. To make compliance with orders to repress focal, leaders would rationally glorify individuals who have complied with such orders in the past. Such individuals are made into heroes or symbols, or promoted to command other agents, the latter of whom, knowing their commander's famous history, are more likely to believe their peers will repress when ordered to.

Yet leaders must be cautious when using these strategies. Coup risk increases when individuals within the security apparatus are better able to coordinate (De Mesquita & Smith, 2010; Little, 2017; Singh, 2014). Thus, some strategies for structuring the security apparatus that may allow regime agents to coordinate around repression in support of the existing leader may

also allow them to better coordinate on deposing the same leader. Cohesion within the security apparatus, as Pion-Berlin et al. (2014) note, can improve the prospects for obedience, but cohesion can also increase coup risk. Thus, leaders must take into account that strategies that can focalize agents' beliefs around obeying orders may also facilitate coordination to undertake activities that are harmful to the leader. This dilemma is related to the "civil-military problematique" identified by Feaver (1996), although his focus is on the strength of the military. We hope to explore this regime cohesion dilemma, and its relationship to debates in the civil–military relations literature, in future work.

Framing

The potential constraining effect of the logic of expectations also helps us understand the importance of framing in endgame scenarios. Leaders often attempt to provoke nonviolent protesters to use violence; when they succeed in doing so, they can better frame orders to repress as essential to the maintenance of internal security. Likewise, when confronted with large-scale peaceful demonstrations, governments often engage in an antihuman rights discourse and try to frame the opposition as "terrorists" to justify and induce their agents to repress opposition groups. Such strategies of framing the situation to one of "law and order" do not necessarily work, as exemplified by the failures of incumbent leaders to successfully deploy such strategies in Ukraine during the 2004 Orange Revolution and in Egypt during the 2011 uprising. However, if the government's message is the only message to which security agents are exposed, it may become an important factor in determining how security agents coordinate their actions in endgame scenarios. The more control such governments have over information, the more difficult it may be for respect for human rights to remain a salient factor. By extension, it may be the case that some level of freedom of speech or of the press is a precondition for the mechanism we have highlighted to have an effect.

Conclusion

Leaders are often desperate to cling to power and may order mass abuses of human rights to do so. Especially in such endgame scenarios, it is not clear whether institutional and normative mechanisms can sufficiently deter abuses. We have described a third mechanism for constraining repression that can be especially relevant in this type of endgame scenario: Repression can depend on the beliefs of regime agents regarding the expected actions of other regime agents. In this section, we describe the broader implications of

this theory, including its implications for several areas of research and the research agendas suggested by the logic of expectations.

The logic of expectations has broad implications for our understanding of the importance of human rights culture. Much of the literature discusses the importance of culture in terms of shaping and reflecting norms and identities. Yet common knowledge is also a key component of culture, and in the context of repression in endgame scenarios, it may prove vital in constraining repression by shaping agents' beliefs. If and to the extent respect for human rights becomes salient across the globe, our model implies this can have the effect of reducing levels of repression even if actors' expectations regarding potential sanctions do not increase and even if human rights norms do not change. This may potentially illuminate long-term mechanisms underlying important empirical findings. For example, Chenoweth and Stephan (2011) find that nonviolent resistance movements are more likely to be effective today than in past eras. While there are likely several reasons for this, our argument implies that one reason may be the salience in the present era of respect for human rights (which thus constrains governments from repressing nonviolent dissent based on the logic of expectations). Likewise, Fariss (2014) finds that, globally, respect for human rights has increased over time, which is interesting in part because the human rights treaty regime has likely only moderately improved human rights conditions. One reason for Fariss's finding may be that respect for human rights is more salient today than it was in the past.

Our argument also has an important implication for scholarly research on human rights law. Most countries have legal protections against mass repression (at least to some extent). Generally, legal institutions can themselves be important sources of beliefs. Rules can affect incentives, and rules can change preferences, but rules can also coordinate beliefs about what others might do (McAdams, 2015). Law can create focal points, around which actors' beliefs can converge in coordination problems (McAdams & Nadler, 2005; Morrow, 2014; Schelling, 1960). Compliance with law can often emerge from solutions to coordination problems (McAdams, 2015). Yet existing work on the effects of human rights law focuses on how it may reduce repression via the logics of consequences and appropriateness, but does not focus on whether and under what conditions it might also change beliefs and create focal points. Our theory implies that *if and to the extent* human rights law can affect government agents' beliefs about the extent to which other agents are willing to conduct repression, the law may reduce repression via the logic of expectations.

The extent to which agents coordinate around disobedience, and thus the extent to which the logic of expectations constrains repression, may depend

in large part on the extent to which advocacy can succeed in making respect for human rights focal. In turn, this implies that human rights advocacy may be significantly more important than some scholars perceive it to be. While scholars have focused on the ways in which advocacy can pressure, persuade, shame, and possibly even deter governments from using repression, critics such as Posner (2014) note that the effects of these mechanisms on powerful governments may be limited. Our arguments indicate that to the extent advocacy can make respect for human rights salient, advocacy can, in turn, affect repression especially in the type of endgame scenario in which leaders have the most powerful incentives to order repression. We also highlighted above the potential importance of this mechanism during the early periods of norm internalization and institutionalization. In such contexts, advocacy can be especially effective to the extent it can change actors' beliefs about what other actors are likely to do. This suggests that scholars of human rights advocacy may fruitfully turn their attention to better understanding how and when advocacy can have these effects.

The logic of expectations also has important implications for empirical research. Scholars often use aggregate observational data about government abuses to test hypotheses about the effects of various structural factors. Yet, if, as our model implies, these data are drawn from a multiple equilibrium context, the use of such data to test such hypotheses may be more complex than it initially appears. Let us suppose, for example, that there are two stable equilibria of agents' level of repressive effort in a particular country or set of countries. If so, then even if structural variables do not change, observational data on repression are drawn from two underlying distributions rather than one underlying distribution. There are at least two strategies for addressing this type of situation. One is to examine the data while relaxing the assumption that they are drawn from a single distribution, for example, by allowing for multidimensionality in the underlying distribution. Second, if the data are indeed drawn from a mixture of distributions, tests for the effects of structural factors can still be conducted, but theories about the comparative statics effects of structural factors may need to be precisely specified. In such contexts, a strategy suggested by Echenique and Komunjer (2009) is to conduct empirical testing of formal comparative statics by using quantile methods that focus on the bounds of the dependent variable in which the monotonicity of equilibria is theorized to hold, rather than on the full sample. This argument also implies that, to better empirically test the causes and consequences of repression, we need to collect better data on several other contexts with respect to which our argument has empirical implications. These include various aspects of social conflict over the shaping of security agents' beliefs and with respect to the control of information.

Finally, our article has important implications for the study of the relationship between dissent and repression. Governments often respond to dissent with repression, and such repression has important effects on the activities of dissenting groups, yet theories and empirical findings regarding how this relationship works are mixed (Davenport, 2007; Moore, 1998; Ritter, 2014; Ritter & Conrad, 2016). When deciding whether or not to mobilize against the state, individuals face a coordination problem; they do not want to face armed security agents alone (Hollyer, Rosendorff, & Vreeland, 2015). Existing work tends to analyze individuals' coordination problem by treating the security apparatus as a unitary actor, focusing on the extent to which dissenters expect others to mobilize. Our model implies that an additional consideration for dissenters should be the extent to which they expect the security agents to follow orders to repress. We hope to explore the implications of this relationship in future work.

Supplemental Material

Supplementary material for this article is available online at the *CPS* website http://journals.sagepub.com/doi/suppl/10.1177/0010414017730077.

Notes

1. The model is related to and contributes to a broader literature that investigates repression and the various aspects of the interaction between leaders and their

    security agents (DeMeritt, 2012, 2015; Dragu & Polborn, 2013; Hollyer & Rosendorff, 2011; Moore, 2000; Rundlett & Svolik, 2016; Svolik, 2013).

2. In many such cases, leaders do order more limited amounts of repression in the buildup of the crisis. Our focus is on whether or not the leader ordered repression on a massive enough scale to allow the leader to remain in power.

3. This implies that the leader might simply offer agents extremely lavish rewards to induce obedience, but leaders have finite resources—and their agents know this.

4. In some situations, the decision to disobey may come from the top down, such as in Argentina in 2001, thus diminishing the coordination problem. When those at the top of the repressive apparatus confirm the leader's order, the coordination problem can become relevant for the agents on the street.

5. In a seminal article, Carlsson and Van Damme (1993) showed that, in coordination games, if each player observes a noisy signal of the true payoffs and if the ex ante feasible payoffs include payoffs that make each action strictly dominant, then players coordinate on a unique equilibrium, which is the risk-dominant equilibrium. Building on Carlsson and Van Damme, an extensive literature on global games explores coordination in situations of uncertainty and investigates the importance of "higher order beliefs" in various settings (Angeletos, Hellwig, & Pavan, 2006; Chassang & Miquel, 2010; Morris & Shin, 2003; Tyson & Smith, 2014).

6. We assume that $\Pi(e_l) > 0$ and $\Pi(e_h) < 1$.

7. Note that we make no assumptions on $\hat{e}$ other than that $\hat{e} > e_l$.

8. Note that because there is a continuum of agents, each agent's action has a zero effect on the average effort level.

9. The repression subgame is a supermodular game because $U_A(e_i, \overline{e})$ has increasing differences in e_i and $\overline{e}$.

10. Note that the left-hand side of Equation 3 is nondecreasing in $\overline{e}$. Therefore, if $\phi(\overline{e}') = e_l$ for some $\overline{e}' \in [e_l, e_h]$, then $\phi(\overline{e}'') = e_l$ for all $\overline{e}'' < \overline{e}'$, and if $\phi(\overline{e}') = e_h$ for some $\overline{e}' \in [e_l, e_h]$, then $\phi(\overline{e}'') = e_h$ for all $\overline{e}'' > \overline{e}'$. As $\phi'(\overline{e}) > 0$ when agent i's optimal action is interior, the best response function $\phi(\overline{e})$ is nondecreasing for $\overline{e} \in [e_l, e_h]$.

11. Intuitively, let $\overline{e} = e^*$ be an equilibrium level of average effort in the repression subgame, and suppose there are two security agents i and j whose levels of effort in this equilibrium are e_i^* and e_j^*. We have $\phi_i(e^*) = e_i^*$ and $\phi_j(e^*) = e_j^*$. By the symmetry of the security agents' payoff function, we have $\phi_i(\cdot) = \phi_j(\cdot)$ for any i and j, and given that any security agent has a unique optimal level of effort for any given $\overline{e}$, then we have $e_i^* = e_j^*$.

12. Recall that an equilibrium is the solution of $\phi(\overline{e}) = \overline{e}$ for $\overline{e} \in [e_l, e_h]$. Given that $\phi(e_l) \geq e_l$ and $\phi(e_h) \leq e_h$, if $\phi'(e^*) > 1$ at some equilibrium $e^* \in (e_l, e_h)$, then $\phi(\overline{e})$ intersects the 45° line multiple times, each corresponding to an equilibrium of the repression subgame.

13. This logic may be more precisely understood as the logic of strategic expectations. We use the shorter nomenclature to parallel the names of existing theoretical

mechanisms: the logic of consequences and the logic of appropriateness.

14. $\underline{C}$ and $\bar{C}$ are functions of exogenous parameters and are derived in the proof of Proposition 3 in the online appendix.

15. $\underline{B}$ and $\bar{B}$ are functions of exogenous parameters and are derived in the proof of Proposition 4 in the online appendix.

16. These equilibria can be obtained by noting that a security agent i's equilibrium level of effort is the solution to the following equation: $\left(\bar{e} + (1/8)\right)3/8 - \left(1 - \bar{e} - (1/8)\right)\bar{e} = 0$. The equilibria are found by solving the (quadratic) equation and by noting that the equation attains a maximum at the boundary $1/2$. Notice that there are three equilibria of repressive effort, $e_L^* = 1/8$, $e_M^* = 3/8$, and $e_H^* = 1/2$. However, the middle equilibrium level of effort is unstable in the sense that a small perturbation on the exogenous parameters induces the agents to move away from that equilibrium. In our analysis, we focus our discussion on stable equilibria. For a discussion on equilibrium stability in games with multiple equilibria, see Echenique (2004).

References

Angeletos, G.-M., Hellwig, C., & Pavan, A. (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy*, *114*, 452-484.

Arendt, H. (1972). *Crises of the republic: Lying in politics, civil disobedience on violence, thoughts on politics, and revolution* (Vol. 219). New York, NY: Houghton Mifflin Harcourt.

Bell, S. R., Clay, K. C., & Murdie, A. (2012). Neighborhood watch: Spatial effects of human rights INGOs. *The Journal of Politics*, *74*, 354-368.

Bushnell, J. (1985). *Mutiny amid Repression: Russian Soldiers in the Revolution of 1905-1906* (Vol. 1). Bloomington: Indiana University Press.

Carlsson, H., & Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica: Journal of the Econometric Society*, *61*, 989-1018.

Carpenter, R. C. (2007). Setting the advocacy agenda: Theorizing issue emergence and nonemergence in transnational advocacy networks. *International Studies Quarterly*, *51*, 99-120.

Casper, B. A., & Tyson, S. A. (2014). Popular protest and elite coordination in a coup d'état. *The Journal of Politics*, *76*, 548-564.

Chassang, S., & Miquel, G. P. I. (2010). Conflict and deterrence under strategic risk. *The Quarterly Journal of Economics*, *125*, 1821-1858.

Chenoweth, E., & Stephan, M. J. (2011). *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

Clark, A. M. (2010). *Diplomacy of conscience: Amnesty international and changing human rights norms*. Princeton, NJ: Princeton University Press.

Conrad, C. R. (2014). Divergent incentives for dictators domestic institutions and (international promises not to) torture. *Journal of Conflict Resolution*, *58*, 34-67.

Conrad, C. R., & Ritter, E. H. (2013). Tenure, treaties, and torture: The conflicting domestic effects of international law. *Journal of Politics*, *75*, 397-409.

Dai, X. (2014). The conditional effects of international human rights institutions. *Human Rights Quarterly*, *36*, 569-589.

Dancy, G., & Michel, V. (2016). Human rights enforcement from below: Private actors and prosecutorial momentum in Latin America and Europe. *International Studies Quarterly*, *60*, 173-188.

Davenport, C. (2007). *State repression and the domestic democratic peace*. New York, NY: Cambridge University Press.

DeMeritt, J. H. R. (2012). International organizations and government killing: Does naming and shaming save lives? *International Interactions*, *38*, 597-621.

DeMeritt, J. H. R. (2015). Delegating death military intervention and government killing. *Journal of Conflict Resolution*, *59*, 428-454.

De Mesquita, B. B., & Smith, A. (2010). Leader survival, revolutions, and the nature of government finance. *American Journal of Political Science*, *54*, 936-950.

Dragu, T. (2017). On repression and its effectiveness. *Journal of Theoretical Politics*, *29*(4), 599-622.

Dragu, T., & Polborn, M. (2013). The administrative foundation of the rule of law. *The Journal of Politics*, *75*, 1038-1050.

Dragu, T., & Polborn, M. (2014). The rule of law in the fight against terrorism. *American Journal of Political Science*, *58*, 511-525.

Echenique, F. (2004). A weak correspondence principle for models with complementarities. *Journal of Mathematical Economics*, *40*, 145-152.

Echenique, F., & Komunjer, I. (2009). Testing models with multiple equilibria by quantile methods. *Econometrica*, *77*, 1281-1297.

Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, *108*, 1-22.

Fariss, C. J., & Schnakenberg, K. E. (2014). Measuring mutual dependence between state repressive actions. *Journal of Conflict Resolution*, *58*, 1003-1032.

Feaver, P. D. (1996). The civil-military problematique: Huntington, Janowitz, and the question of civilian control. *Armed Forces & Society*, *23*, 149-178.

Feaver, P. D. (2009). *Armed servants: Agency, oversight, and civil-military relations*. Cambridge, MA: Harvard University Press.

Finnemore, M., & Sikkink, K. (1998). International norm dynamics and political change. *International Organization*, *52*, 887-917.

Gandhi, J., & Przeworski, A. (2007). Authoritarian institutions and the survival of autocrats. *Comparative Political Studies*, *40*, 1279-1301.

Hadfield, G. K., & Weingast, B. R. (2012). What is law? A coordination model of the characteristics of legal order. *Journal of Legal Analysis*, *4*, 471-514.

Hill, D. W. (2010). Estimating the effects of human rights treaties on state behavior. *Journal of Politics*, *72*, 1161-1174.

Hill, D. W., & Jones, Z. M. (2014). An empirical evaluation of explanations for state repression. *American Political Science Review*, *108*, 661-687.

Hollyer, J. R., & Rosendorff, B. P. (2011). Why do authoritarian regimes sign the convention against torture? Signaling, domestic politics and non-compliance. *Quarterly Journal of Political Science*, *6*, 275-327.

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2015). Transparency, protest, and autocratic instability. *American Political Science Review*, *109*, 764-784.

Huntington, S. P. (1957). *The soldier and the state: The theory and politics of civil-military relations*. Cambridge, MA: Harvard University Press.

Janowitz, M. (1960). *The professional soldier: A social and political portrait* (Vol. 91618). Glencoe, IL: Free Press.

Johnson, C. A. (1982). *Revolutionary change*. Stanford, CA: Stanford University Press.

Kaempfer, W. H., Lowenberg, A. D., & Mertens, W. (2004). International economic sanctions against a dictator. *Economics & Politics*, *16*, 29-51.

Keck, M. E., & Sikkink, K. (1998). *Activists beyond borders: Advocacy networks in international politics*. Ithaca, NY: Cornell University Press.

Khaddour, K. (2015). *Assad's officer ghetto: Why the Syrian army remains loyal*. Beirut, Lebanon: Carnegie Middle East Center.

Krain, M. (2012). J'accuse! Does naming and shaming perpetrators reduce the severity of genocides or politicides? *International Studies Quarterly*, *56*, 574-589.

Levitsky, S., & Way, L. (2005). International linkage and democratization. *Journal of Democracy*, *16*, 20-34.

Little, A. T. (2017). Coordination, learning, and coups. *Journal of Conflict Resolution*, *61*, 204-234.

Lupu, Y. (2013a). Best evidence: The role of information in domestic judicial enforcement of international human rights agreements. *International Organization*, *67*, 469-503.

Lupu, Y. (2013b). The informative power of treaty commitment: Using the spatial model to address selection effects. *American Journal of Political Science*, *57*, 912-925.

Lupu, Y. (2015). Legislative veto players and the effects of international human rights agreements. *American Journal of Political Science*, *59*, 578-594.

Lutz, E. L., & Sikkink, K. (2000). International human rights law and practice in Latin America. *International Organization*, *54*, 633-659.

McAdams, R. H. (2015). *The expressive powers of law: Theories and limits*. Cambridge, MA: Harvard University Press.

McAdams, R. H., & Nadler, J. (2005). Testing the focal point theory of legal compliance: The effect of third-party expression in an experimental hawk/dove game. *Journal of Empirical Legal Studies*, *2*, 87-123.

McLauchlin, T. (2010). Loyalty strategies and military defection in rebellion. *Comparative Politics*, *42*, 333-350.

Milgrom, P., & Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, *58*, 1255-1277.

Moore, W. H. (1998). Repression and dissent: Substitution, context, and timing. *American Journal of Political Science*, *42*, 851-873.

Moore, W. H. (2000). The repression of dissent: A substitution model of government coercion. *Journal of Conflict Resolution*, *44*, 107-127.

Moravcsik, A. (2000). The origins of human rights regimes: Democratic delegation in postwar Europe. *International Organization*, *54*, 217-252.

Morris, S., & Shin, H. S. (2003). Global games: Theory and applications. In M. Dewatripont, L. Hansen, & S Turnovsky (Eds.), *Advances in economics and econometrics* (pp. 56-114). New York, NY: Cambridge University Press.

Morrow, J. D. (2014). *Order within anarchy: The laws of war as an international institution*. New York, NY: Cambridge University Press.

Murdie, A., & Bhasin, T. (2011). Aiding and abetting: Human rights INGOs and domestic protest. *Journal of Conflict Resolution*, *55*, 163-191.

Murdie, A., & Davis, D. R. (2012). Shaming and blaming: Using events data to assess the impact of human rights INGOs. *International Studies Quarterly*, *56*, 1-16.

Myerson, R. B. (2004). The empirical and theoretical underpinnings of the law merchant: Justice, institutions, and multiple equilibria. *Chicago Journal of International Law*, *5*, 91-763.

Nepstad, S. E. (2013). Mutiny and nonviolence in the Arab Spring Exploring military defections and loyalty in Egypt, Bahrain, and Syria. *Journal of Peace Research*, *50*, 337-349.

Olsen, T. D., Payne, L. A., & Reiter, A. G. (2010). The justice balance: When transitional justice improves human rights and democracy. *Human Rights Quarterly*, *32*, 980-1007.

Pion-Berlin, D., Esparza, D., & Grisham, K. (2014). Staying quartered: Civilian uprisings and military disobedience in the twenty-first century. *Comparative Political Studies*, *47*, 230-259.

Pion-Berlin, D., & Trinkunas, H. (2010). Civilian praetorianism and military shirking during constitutional crises in Latin America. *Comparative Politics*, *42*, 395-411.

Posner, E. A. (2014). *The twilight of human rights law*. Oxford, UK: Oxford University Press.

Quinlivan, J. T. (1999). Coup-proofing: Its practice and consequences in the Middle East. *International Security*, *24*, 131-165.

Ritter, E. H. (2014). Policy disputes, political survival, and the onset and severity of state repression. *Journal of Conflict Resolution*, *58*, 143-168.

Ritter, E. H., & Conrad, C. R. (2016). Preventing and responding to dissent: The observational challenges of explaining strategic repression. *American Political Science Review*, *110*, 85-99.

Rozenas, A. (2016). Office insecurity and electoral manipulation. *The Journal of Politics*, *78*, 232-248.

Rundlett, A., & Svolik, M. W. (2016). Deliver the vote! Micromotives and macrobehavior in flectoral fraud. *American Political Science Review*, *110*, 180-197.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Sikkink, K. (2011). *The Norton Series in World Politics: The justice cascade: How human rights prosecutions are changing world politics*. W.W. Norton.

Sikkink, K., & Walling, C. B. (2007). The impact of human rights trials in Latin America. *Journal of Peace Research*, *44*, 427-445.

Simmons, B. A. (2009). *Mobilizing for human rights: International law in domestic politics*. New York, NY: Cambridge University Press.

Singh, N. (2014). *Seizing power: The strategic logic of military coups*. Johns Hopkins University Press.

Smith-Cannoy, H. (2012). *Insincere commitments: Human rights treaties, abusive states, and citizen activism*. Washington, DC: Georgetown University Press.

Sri Lanka to Investigate "Rajapaksa Coup Plot." (2015, January). *BBC News*. Retrieved from http://www.bbc.com/news/world-asia-30769188

Staniland, P. (2008). Explaining civil-military relations in complex political environments: India and Pakistan in comparative perspective. *Security Studies*, *17*, 322-362.

Svolik, M. W. (2009). Power sharing and leadership dynamics in authoritarian regimes. *American Journal of Political Science*, *53*, 477-494.

Svolik, M. W. (2013). Contracting on violence: The moral hazard in authoritarian repression and military intervention in politics. *Journal of Conflict Resolution*, *57*, 765-794.

Talmadge, C. (2015). *The dictator's army: Battlefield effectiveness in authoritarian regimes*. Ithaca, NY: Cornell University Press.

Tilly, C. (1978). *From mobilization to revolution*. New York, NY: McGraw-Hill.

Topkis, D. M. (1979). Equilibrium points in nonzero-sum n-person submodular games. *Siam Journal on Control and Optimization*, *17*, 773-787.

Tyson, S. A., & Smith, A. (2014). Dual-layered coordination: Conflict and the role of information. *Journal of Politics*, *76*(2), 548-564.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics*, *19*, 305-321.

Welch, C. E. (2001). *Protecting human rights in Africa: Roles and strategies of non-governmental organizations*. Philadelphia: University of Pennsylvania Press.

Welt, C. (2010). Georgia's rose revolution: From regime weakness to regime collapse. In V. Bunce, M. A. McFaul, & K. Stoner-Weiss (Eds.), *Democracy and authoritarianism in the post-communist world* (pp. 155-188). New York, NY: Cambridge University Press.

Wiebelhaus-Brahm, E. (2010). *Truth commissions and transitional societies: The impact on human rights and democracy*. New York, NY: Routledge.

Wong, W. H. (2012). *Internal affairs: How the structure of NGOs transforms human rights*. Ithaca, NY: Cornell University Press.

Author Biographies

Tiberiu Dragu is Assistant Professor of Politics at New York University. His research focuses on positive political theory and political institutions, particularly in relation to questions of how to structure and constrain government power in both democratic and non-democratic polities.

Yonatan Lupu is Assistant Professor of Political Science and International Affairs at George Washington University. His research focuses on the repression of human rights and international institutions.