

# Rational Overreaction to Terrorism

William Spaniel\*

January 4, 2018

## Abstract

Some terrorist organizations provoke their targets into deploying massive countermeasures, allowing terrorists to mobilize a greater share of their audience. Why would a government pursue such a costly strategy if it only strengthens the opponent? I develop a signaling model of terrorism, counterterrorism, and recruitment. If a target government is unsure whether the terrorists' audience is sympathetic to the cause, weaker groups sometimes bluff strength by attacking. To check this bluff, governments sometimes respond to attacks with large-scale operations, even though they know they might be overreacting. Comparative statics reveal that overreaction regret is most likely when the target is wealthy and large operations are more effective. Thus, a selection effect creates the false impression that provocation is most effective against geopolitically privileged targets.

---

\*Assistant Professor, Department of Political Science, University of Pittsburgh. ([williamspaniel@gmail.com](mailto:williamspaniel@gmail.com), <http://williamspaniel.com>). I thank Vincent Bauer, Martha Crenshaw, Carl Gustafson, Morgan Kaplan, Kate Cronin-Furman, Bart Lemming, Iris Malone, David Siegel, and Lucas Theriault for their comments.

“[O]ne of those great challenges of fighting terrorism is to find the balance between being tempted to overreact but also being absolutely certain...that people understand that we aren’t going to take this sitting down.” –Paul Bremer, in an interview with NBC’s Tom Brokaw at 12:22 pm, September 11, 2001.<sup>1</sup>

## 1 Introduction

Terrorist organizations face an inherent tradeoff between acquiring support and suffering military defeat. If an organization commits too few attacks, it may lose support (Thornton 1964; Crenshaw 1981, 386-388; Bloom 2005). Yet too many attacks risks a crackdown that could break the organization. The strategy of provocation, however, negates this tradeoff entirely: an organization commits an attack and hopes for a disproportionate response from its target.<sup>2</sup> These strong-arm tactics polarize an audience sympathetic to the terrorist organization’s cause (Siqueira and Sandler 2006), leading to more recruits and increased funding. The target’s response thus backfires, costing the price of the operation yet only increasing the terrorist group’s strength.

Such a strategy is inherently clever but also raises questions about how it could succeed. If overreactions are self-defeating, why don’t governments pursue more moderate measures?<sup>3</sup> One possibility is that targets consistently make irrational mistakes (Berry 1987, 9). But this is unsatisfying. Governments have understood provocation since at least the days of Batista’s Cuba (Price 1977, 54; Gude 1969), and American policymakers expressed this precise concern even as the events of September 11 were unfolding. It therefore seems strange that they would systematically continue making the same mistake all the way through the United States’ war on terror today.

In this paper, I argue that the overreaction problem runs deeper than current re-

---

<sup>1</sup><https://youtu.be/89G749GrTBQ?t=5h22m10s>

<sup>2</sup>Both scholars and policymakers have paid great attention to the strategy of provocation. See Fromkin 1975, 962-964; Price 1977; Crenshaw 1981, 387; Berry 1987, 8-10; Laqueur 1987; Lake 2002; Bloom 2005, 107-110; Kydd and Walter 2006, 69-72.

<sup>3</sup>More specifically, I focus on offensive strategies to defeat terrorism. This contrasts with the over-provision of domestic counterterrorism, like enhanced security measures at airports (Mueller 2006). Bueno de Mesquita (2007) shows that such over-provision occurs when citizens cannot effectively monitor their government’s investment in clandestine counterterrorism activities. Since the literature on these other explanations for large-scale counterterrorism are relatively developed, I limit the model to analyzing provocation to keep the analysis tractable.

search suggests.<sup>4</sup> Governments face uncertainty regarding the sympathies of a terrorist group's audience and their resources (Schmid and Jongman 1988, 488; Lapan and Sandler 1993). Consequently, they cannot know how that audience will react to moderate versus large-scale counterterrorism operations. Meanwhile, terrorists can strategically select whether to attack based on their knowledge of how their audience will react. This gives rise to interesting signaling behaviors. Organizations robust to large-scale operations want to strike under these conditions. But this incentivizes weaker groups to bluff strength by attacking too, thereby obfuscating the target's optimal response. Consequently, overreaction may be the result of a rational decisionmaking.

To work through this logic, I develop a model of how governments, terrorist organizations, and their audiences behave facing such constraints. All told, the outcome of this strategic problem depends on the relative likelihood of how the audience will respond to various countermeasures. When the government believes that the group is likely to have an audience that will rally to the group *en masse*, the terrorist group always attacks. The government only issues a moderate response, deterred by the belief that large-scale assaults will probably backfire. This benefits groups without much support, and they can attack without fear of significant retribution.

Strategic incentives grow more complex when the government believes the terrorist group is unlikely to benefit from large-scale countermeasures. Here, the government rationally overreacts with positive probability in equilibrium. This is because terrorist groups in a relatively strong position always attack, while more vulnerable groups sometimes attack and sometimes do not. Due to the uncertainty this creates, the government sometimes undertakes a large-scale response and sometimes undertakes a moderate response. With positive probability, this interaction ends with the government overreacting, choosing a large response when a moderate response is preferable. However, uncertainty prohibits the government from always pursuing the optimal response, causing the overreaction.

The model provides useful comparative statics on the probability of overreaction and the risk of attack. First, conditional on a large-scale response, I show that overreactions are especially likely when such operations kill a greater share of terrorists. This may be surprising; after all, an overwhelmingly strong military would eliminate

---

<sup>4</sup>Kydd and Walter (2006, 72), for example, simply state that “[t]he best response to provocation is a discriminating strategy that inflicts as little collateral damage as possible.”

more terrorists than the countermeasures would create, thus insulating itself from the overreaction problem. Yet this logic fails in less extreme cases. Instead, stronger military capabilities protect a government from the negative consequences of overreaction. Given the counterbalance, the government is happier to engage in large-scale operations even though they may backfire.

Second, the model shows large-scale responses are more likely to be mistakes as they become cheaper. One might assume that being in such a strong position would lead to better counterterrorism operations, but this is not the case conditional on an attack. Rather, governments suffer less if they guess incorrectly when costs are low. As such, they have greater incentive to call what might be a bluff, leading to more overreactions. Taken together, these comparative statics help explain why rich and powerful countries frequently implement ineffective repression policies (e.g., Lafree, Dugan, and Korte 2009; Dugan and Chenoweth 2012; Benmelech, Berrebi, and Klor 2015).

Interestingly—and heretofore unmentioned in the provocation literature—is that this uncertainty can also cause suboptimal *underreaction*. That is, the government may select a moderate response because it falsely believes that the terrorist group could benefit from a large-scale response. The comparative statics flip here; underreaction is more likely to occur when large-scale responses kill a smaller share of terrorists and when they are expensive to conduct.

Finally, although the above logic may suggest that states directly benefit from cheaper and more powerful responses to terrorism, this is false when facing uncertainty. Counterintuitively, the model shows that the value of these qualities falls entirely in their second-order effects. Conditional on the use of large-scale responses, states with cheap and effective military options are as well off as states with expensive and ineffective options. Rather, *all* of the benefit of this additional coercive power comes from a deterrence effect, with weaker terrorist organizations selecting out of the interaction by not attacking. This implies that the standard recommended response to provocative attacks—discriminate counterterrorism (Kydd and Walter 2006, 72)—has a heretofore unappreciated downside risk.

The substantive direction of my model is most similar to Bueno de Mesquita (2005), Bueno de Mesquita and Dickson (2007), and Overgaard (1994) in that we all wish to analyze optimal responses to terrorism. However, we vary in the information structures.<sup>5</sup>

---

<sup>5</sup>In contrast, Lapan and Sandler features uncertainty over terrorists' resources, but those resources

Bueno de Mesquita's (2005) game, for example, has complete information. Therefore, although the counterterrorism's productivity is a random variable (leading to possible *ex post* regret), terrorist groups cannot exploit an informational advantage to induce the government into pursuing *ex ante* suboptimal behaviors.

Meanwhile, Bueno de Mesquita and Dickson suppose that the audience of sympathizers is uncertain of the target's underlying motives; citizens wish to support soft-line governments but resist hard-line governments. Soft-line governments can correspondingly separate by choosing lower levels of violence; hard-line governments meanwhile do not internalize citizens' suffering and select harsher countermeasures. Thus, they emphasize the signaling qualities of the *response* to terrorism whereas I explore the signaling qualities of the initial terrorist attack.

Correspondingly, my model shares Overgaard's emphasis the information problem governments face in not knowing how sympathizers will respond to counterterrorism.<sup>6</sup> Overgaard, however, explores a situation where terrorist groups use resources to coerce a bargaining good out of a rival. Thus, regardless of type, groups most prefer capitulation from the government. He correspondingly finds that weaker groups have an incentive to misrepresent, similar to the argument in Fearon 1995. My model is distinct in that, consistent with the provocation literature, terrorist groups with stronger military positions prefer large-scale responses. This leads to a more complex strategic scenario in which *both* types have incentives to misrepresent: weaker types wish to mimic strength to obtain a deterrence effect, whereas stronger types wish to mimic weakness to reap recruitment gains. As a result, the target government faces greater difficulty in interpreting the signals it receives.

---

are exogenous; counterterrorism cannot make the government worse off. Resources in my model instead depend on the government's decision, potentially motivating groups to commit attacks. Meanwhile, Rosendorff and Sandler (2004) allow for increased recruitment following a crackdown but only analyze proactive counterterrorism. This prevents them from analyzing the signaling properties of an attack, which is my focus.

<sup>6</sup>It comes at the cost in understanding why certain responses can decrease support for terrorist organizations, which Bueno de Mesquita and Dickson's model excels at. See Funes 1998 for a substantive description of this mechanism in the Basque separatist movement. See also Kydd and Walter 2002 for a similar model with uncertainty about a government's type.

## 2 Microfoundations: Provocation and the Calculus of Counterterrorism

Although the targets of terrorism are often random, terrorist organizations are deliberate in their actions.<sup>7</sup> Broadly, terrorists (and insurgent movements in general) face two competing strategic imperatives. Primarily, they must survive. One standard solution to this problem is to initiate wars of attrition, designed to avoid large confrontations (Mao 1937; Moloney 2007, 149-162). Under normal circumstances, terrorist organizations cannot survive overwhelming force. In turn, their best chance of victory requires forcing the enemy to continually engage it at manageable levels over the course of a long period until the opponent concedes.

Second, terrorist organizations must maintain support and impose coercive leverage against their targets, both of which usually require committing some amount of violence. Of course, violence can incite a militarized response, which directly conflicts with the primary goal of survival. The threat of reprisals is sometimes sufficient to deter organizations from committing attacks. Other times, though, terrorist groups can use this violence to their advantage (Kydd and Walter 2006, 69-72). Many organizations throughout history have hoped that large-scale responses would drive more supporters to their causes.<sup>8</sup> Indeed, military crackdowns can increase support for their targets by limiting economic opportunities, radicalizing moderates, and inspiring support for their victims (Bueno de Mesquita 2005; Bueno de Mesquita and Dickson 2007; Rosendorff and Sandler 2004).

This provocation strategy does not always succeed, however. Sometimes governments recognize the risks of large-scale operations and instead exercise moderation, limiting their interventions to careful and discriminate crackdowns (Thornton 1964). Other times the targets call the provoker's bluff and undertake broad operations. As

---

<sup>7</sup>The scholarly definition of terrorism is murky at best (Hoffman 2013, 1-42). I deliberately avoid staking a precise claim on this front. Rather, the model presented below simply assumes that the terrorist organization values an attack on its target and the target suffers from that an attack. Contingent on the other assumptions in the model about how counterterrorist operations function, this formal definition is broad and may apply to all sorts of "terrorist" attacks, be they targeting civilians or governments, originating from small groups or large, or part of a broad insurgency or not. I call these acts "terrorism" for substantive motivation and convenience.

<sup>8</sup>An incomplete list includes ETA (Woodworth 2001; Zirakzadeh 2002), Al-Qaeda (Lake 2002), the Zionist Irgun (Bell 1977, 71-72), Hamas (Mishal and Sela 2006), the National Liberation Front in Algeria (Hoffman 2013, 60-61), and Cuban resistance to Batista (Price 1977, 54; Gude 1969).

in Venezuela (Price 1977, 54; Gude 1969) and Bolivia (Weitz 1980, 410), doing so can put the organization in a worse position than had they committed no attack at all.

Policymakers recognize these tradeoffs. Terrorist organizations openly boast of provocation as a strategy. Indeed, the first three steps of al-Qaeda’s “Strategy to the Year 2020” document were to (Atwan 2008, 221-222):

1. Provoke the “ponderous American elephant” into invading Muslim countries
2. Mobilize *mujahedin* by awakening the Umma’s (Islamic community’s) latent anger toward the United States
3. Fight a war of attrition

Further, in *Management of Savagery*, Abu Bakr Naji states that one of the main goals for jihadists should be replenish lost terrorists, which will come in large part from

[a]nger over the obvious, direct American interference in the Islamic world, such that that anger compounds the previous anger against America’s support for the Zionist entity. It also transforms the suppressed anger toward the regimes of apostasy and tyranny into a positive anger. Human aid for the renewal movement will not dry up, especially when heedless people among the masses—and they are the majority—discover the truth of the collaboration of these regimes with the enemies of the Umma to such an extent that no deceptive veil will be of use and no pretext will remain for any claimant to the Islam of these regimes and their like. (Naji 2004, 10)

Similar statements recur throughout the book.<sup>9</sup>

Thus, policymakers face a difficult challenge. Terrorism is often the product of foreign occupations and expansive foreign policy positions (Pape 2005, 23; Savun and Phillips 2009). If large-scale operations will provoke mass mobilization to the terrorist group, such operations only increase the “stock” of attackers (Kaplan et al 2005). Target governments ought to then pursue moderate responses. Failing to do this will

---

<sup>9</sup>*Management of Savagery* is an Al-Qaeda treatise on strategy, originally distributed online in 2004. Similar themes appear in Osama Bin Laden’s post-September 11 comments and *Dabiq*, the Islamic State’s English-language online magazine. See Marighela 1971 (94-95) for an example predating the modern war on terror.

only make the problem worse, especially if military options would not be particularly effective anyway. Correspondingly, most terrorism countermeasures merely involve judicial approaches or police intervention (Duyvesteyn 2008; Crenshaw 2015), with only around 1% of terrorist attacks in post-war Western Europe triggering a military response (Carter 2016).<sup>10</sup> Yet if a terrorist group’s audience will tolerate large interventions, then the target government should not be afraid to engage fully. Choosing the wrong strategy here misses an opportunity to kill a greater share of terrorists (Kydd and Walter 2006, 71). Moreover, by not committing to a large-scale response, the target fails to enjoy the benefit of deterrence against weaker terrorist groups (Brophy-Baermann and Conybeare 1994).

If policymakers had a clear understanding of sympathy toward terrorist groups, the choice would be easy. However, the information problem runs deeper than that. This is, after all, the *terrorists’* audience. Correspondingly, terrorist groups can more readily identify how their audience will react to various interventions before they occur. Although governments can speculate about those support levels—and perhaps make inferences based on whether a terrorist organization committed an attack—they nevertheless face an asymmetric information problem. Put simply, estimating a terrorist group’s support is among the most difficult challenges a government faces (Schmid and Jongman 1988, 488).

The above discussion indicates that a formal model will be useful in sorting out expectations in such interactions.<sup>11</sup> For a government, the decision to undertake large-scale counterterrorism operations is not as simple as weighing the relative likelihood of blowback. Rather, a terrorist organization’s decision whether to commit an attack endogenously determines a target government’s beliefs. Yet, as explained above, a terrorist group’s decision to attack simultaneously depends on how it expects the target to react. Consequently, it is not immediately clear how terrorists ought to influence those beliefs and how governments ought to respond. The formal analysis below sorts out those issues.

---

<sup>10</sup>Small-scale military or intelligence operations might be systematically underreported because governments sometimes lack incentive to disclose such activities. This is likely not a problem for large-scale response, which generate enough activity to be readily identifiable.

<sup>11</sup>Formal theory acts as “accounting standards” to ensure that assumptions map to logically valid conclusions (Powell 1999, 30-33).

### 3 The Model

The interaction features three players: a terrorist organization, a target government, and the audience, which is a group of citizens with sympathies toward the terrorist group. Nature starts the game by drawing the citizens as “sympathetic” or “unsympathetic,” which I define in a moment. The citizens and the terrorist group observe the draw, but the government does not. Play begins with the terrorist group choosing whether to launch an attack against the target government. If no attack occurs, the game ends.<sup>12</sup> Meanwhile, if the terrorists attack, the government decides whether to initiate a moderate response or start a large-scale operation.<sup>13</sup>

I model the citizens as astrategic. Instead, following the substantive discussion in the previous section, I assume that citizens’ support depends on both the terrorist group’s and government’s actions. In particular, let  $v_{-A}^U > 0$  represent the total value of support citizens offer to the group if it commits no attack at all and the citizens are the unsympathetic type. Likewise,  $v_M^U > 0$  and  $v_L^U > 0$  represent support given that the citizens are unsympathetic following an attack and a moderate response or large response, respectively. Because an attack advertises the group and a large scale response creates grievances against the government, I assume  $v_L^U > v_M^U > v_{-A}^U$ .<sup>14</sup> I define  $v_{-A}^S$ ,  $v_L^S$ , and  $v_M^S$  analogously for when citizens are the sympathetic type.

In addition, holding fixed how the attack/response interaction ends between the terrorist group and government, the terrorist group receives more volunteers and support when the citizens are sympathetic. That is,  $v_{-A}^S > v_{-A}^U$ ,  $v_M^S > v_M^U$ , and  $v_L^S > v_L^U$ .

---

<sup>12</sup>One might imagine that by not attacking, the government does not know whom to target and thus cannot adequately respond. Alternatively, including an explicit response in the model would not change the fundamental predictions as long as such responses result in sizable citizen backlash against the government. This may be because citizens would sympathize with a group being targeted without provocation. Under these conditions, the government would prefer restraint, which matches the assumed non-response in the model analyzed.

<sup>13</sup>Although the interpretation of these terms is context dependent, one may conceptualize the moderate response as asset freezing, targeted bombings, and limited raids. In contrast, large-scale operations might represent broad wars meant to eradicate greater portions of the terrorist organization.

<sup>14</sup>Many strategic behaviors from the citizens could result in this ordering. These orderings fit the notion that more onerous actions translate to increases in terrorism (Pape 2005, 23). The model therefore features a response that radicalizes the audience in the manner that Lake (2002) discusses. It is also consistent with the notion that even limited responses can still backfire to some degree (Linkekilde 2014). Thus, by modeling citizens as astrategic, I do not imply that citizens act irrationally here. Rather, I suppress this aspect of the interaction to instead focus on how the terrorist group and government interact given these outcomes.

This creates the central tension of the model: the terrorist group knows latent citizen sympathy but the government does not. The government must therefore consider the signal the terrorist group sends by attacking in formulating its response.

Meanwhile, the terrorist group’s payoff depends on three factors: the value of resources going to the organization (Cronin 2011, 40), the portion of resources that remain available after an intervention, and an expressive payoff for committing an attack. Let  $p_L \in [0, 1)$  represent the proportion of the value the terrorist group actually receives following a large-scale response;  $p_M \in (p_L, 1]$  reflects proportion of the value the terrorist group actually receives following a moderate response.

Let the value  $a > 0$  reflects the expressive value of the attack. Overall, then, the terrorist group with sympathetic citizens earns  $p_L v_L^S + a$  if it attacks and the government responds with a large-scale operation,  $p_M v_M^S + a$  if it attacks and the government responds with a moderate operation, and  $v_{-A}^S$  if it commits no attack.<sup>15</sup> Payoffs are analogous for the case when citizens are sympathetic, swapping the superscript S with U.

Finally, the government’s payoff depends on four factors: the value of resources going to the organization, the portion of those resources that remain available after an intervention, the damage suffered in an attack, and the cost of large-scale counterterrorist operations. Rather than gaining from more terrorist resources, the government incurs the negative. Let  $c > 0$  represent the cost of undertaking the larger action. Then the government earns  $-p_L n - a - c$  if the terrorist group attacks and it responds with a large-scale operation,  $-p_M n - a$  if the terrorist group attacks and it responds with a moderate operation, and  $-n$  if the group commits no attack.<sup>16</sup>

Because the terrorist group’s central goal in this provocation model is to mobilize support, it is worth describing the parameterizations in further detail. The various values of  $n$  reflect the *gross* value of support citizens offer to the terrorist group. This includes the number of volunteers who wish to join the group, monetary offerings from citizens who wish to remain civilians, any logistical support, and how well the group

---

<sup>15</sup>If the optimal response to terrorism is no response at all, this model covers that contingency by setting  $p_M = 1$ —that is, the moderate “non-response” leaves all terrorists alive. The equilibria described below exist under  $p_M = 1$  for non-knife edge combinations of the remaining parameters.

<sup>16</sup>Note that the government incurs no explicit cost for a moderate attack. Implicitly, however, this is irrelevant—the model could instead assign two separate costs for large-scale and moderate operations and obtain the same results. That said,  $c$  in the model is better understood as the additional cost of large-scale operations compared to a moderate response.

can exploit those offerings.

However, the *net* value of support is the gross value multiplied by the potential reductions through a moderate ( $p_M$ ) or large-scale ( $p_L$ ) intervention. The model black boxes the machinations of the intervention, but interventions in practice have two effects. First, their immediate impact is to kill existing terrorists and destroy the organization's infrastructure. This reduces the value of any citizen support given afterward, as the group must rebuild to effectively utilize it. Larger interventions intuitively do more damage here. Second, over the long term, interventions can stop assistance and kill new recruits. Larger interventions do more here as well. Indeed, even if especially effective interventions radicalize a large share of the population, massive responses may leave the group incapable of exploiting their new resources (Kydd and Walter 2006, 71).

The model captures both of these effects, but a central tension remains. Larger interventions may take a greater value away from terrorist groups than moderate interventions. However, larger interventions create a greater *gross* value of support, so the *net* value to the terrorist group may still be greater despite the larger reduction. Moreover, the government's uncertainty means that it may not know whether the larger gross support outweighs the usefulness of a large-scale intervention.

Given that, we can now obtain a formal definition of overreaction:

**Definition.** *A government **overreacts** if it chooses a large-scale response and has ex-post regret—that is, after payoffs are realized, it would have preferred selecting a moderate response.*

I will refer back to this definition in the process of discovering when overreaction is most likely to occur.

### 3.1 Complete Information Equilibria

Before delving into delving a substantively more important version of the model with incomplete information, it will first prove useful to understand how the interaction works with complete information. Depending on how much the terrorists value an attack, the responsiveness of citizens to counterterrorist operations, the effectiveness of said operations, and the cost of undertaking large-scale responses, many different outcomes may occur. However, some of these parameter spaces are of greater theoretical interest

than others. Consequently, I take a few conditions on the parameter space, which I describe below. All the main propositions focus on these parameter spaces, though I briefly describe the results outside their regions.

**Condition 1.** *If the audience is sympathetic, suppose the terrorist group is sufficiently robust to counterterrorism operations that it prefers incurring a large-scale response to a moderate response. That is,  $p_L v_L^S + a > p_M v_M^S + a$ .*

**Condition 2.** *Suppose the government prefers to initiate a large-scale response if the audience is unsympathetic. That is,  $-p_L v_L^U - c > -p_M v_M^U$ .*

Although these are only two conditions, they imply other important preferences from the players. To begin, Condition 1 implies that the government prefers a moderate response against a sympathetic audience, or  $-p_M v_M^S > -p_L v_L^S - c$ . This statement is similar to Condition 2. Note that the two values appearing on either side of the inequality in Condition 2 and the implication of Condition 1 are the government’s payoff for taking a particular action given an attack against the respective type of audience. I focus on these parameters because they imply that the different types of audiences act differently. If they failed to hold, the distinction between “sympathetic” and “unsympathetic” is meaningless—the government would apply the same response regardless of audience support. Thus, for the government to face a real strategic dilemma, the parameters must fall in the described region.

Condition 1 ensures that the terrorist group operates from a position of strength if the audience is sympathetic to its cause. Indeed, it is necessary to address the provocation logic at all; if groups never benefited from large-scale responses, provocation would never occur. Combined with Condition 3 below, such a group most prefers an attack with a large response, least prefers no attack at all, and places a moderate response to an attack in the middle. Once again, the purpose of the condition is to focus on the substantively interesting case; if Condition 1 did not hold, both the group with a sympathetic audience and the government would agree to a moderate response following an attack.<sup>17</sup>

Lastly, note that Condition 2 implies that  $p_M v_M^U + a > p_L v_L^U + a$ . In conjunction with Condition 1, this draws a distinction between the terrorist group’s preferences

---

<sup>17</sup>This also distinguishes my model from Overgaard’s (1994), in which all types of terrorist organizations prefer a “soft” government response.

depending on the sympathy of the audience. If that implication were not true, a group with an unsympathetic audience would have preferences that match the a group with a sympathetic audience. Thus, they would be behaviorally identical. With this ordering, however, the signal that the terrorist organization sends provides information and forces the government to deduce its meaning.

To ensure that the analysis focuses on motivated terrorist organizations, I add one more condition:

**Condition 3.** *Regardless of the audience’s sympathy, the terrorist group prefers committing an attack and incurring a moderate response to no attack at all. That is,  $p_M v_M^S + a > v_{-A}^S$  and  $p_M v_M^U + a > v_{-A}^U$ .*

This condition ensures that parameters exist such that the terrorist group would want to commit an attack. If this were not the case, the signaling problem becomes trivial—when updating its belief conditional on an attack, the government would quickly conclude that a group with an unsympathetic audience is certainly not responsible for the violence. Consequently, the government could infer which type of audience it is facing and take the appropriate countermeasures. Without this, the information problem would functionally not exist.

With these conditions set, the subgame perfect equilibrium of the game is straightforward. If the government knows it is facing the sympathetic type, it chooses a moderate response by Condition 1. The group with a sympathetic audience thus faces a choice between suffering the moderate response following an attack and not attacking. Then also by Condition 1, the group with a sympathetic audience attacks. In essence, the strength of public support gives the terrorist organization the confidence to attack and scares the government into the weaker countermeasure. If the government knows it is facing an unsympathetic audience, it chooses a large-scale response by Condition 2. The terrorist group then chooses not to attack rather than attack and face the large-scale response. Here, the lack of backlash against great countermeasures deters the terrorist organization from initiating a conflict.

Note that the probability of overreaction here is 0. This is because complete information allows the government to always choose an optimal response and thereby never experience *ex-post* regret. By relaxing this assumption below, I show that incomplete information causes overreaction.

### 3.2 Perfect Bayesian Equilibria

Consider the following incomplete information extension. Suppose at the beginning of the game, Nature draws the distribution of citizens as sympathetic with probability  $q$  and unsympathetic with probability  $1 - q$ . Citizens observe their own draws and the terrorist organization knows the drawn distribution, but the target government only has the prior belief. Thus, in this richer strategic environment, the terrorist organization moves knowing how the audience will react to an eventual attack. The government does not but can make an inference about the terrorist group’s strength based on the implicit signal the group sends by choosing whether to attack (Lapan and Sandler 1993). This reflects the idea that the organizations—which are on the ground and are naturally better connected to their possible constituencies—better understand their audience’s inclinations, ideological preferences, psychological weaknesses, and willingness to contribute to collective action.<sup>18</sup>

Going forward, I refer to the terrorist organization as the “sympathetic type” or the “unsympathetic type” throughout this section. This is for convenience of discussion. In fact, both types of terrorist organizations have identical preferences. However, following Conditions 1 and 2, they prefer different responses from the government. As a result, they effectively play as different types in the traditional signaling game manner.

In any case, because this is now a sequential game with incomplete information, I solve for its perfect Bayesian equilibrium (PBE). Although PBE sometimes admit multiple equilibria depending on off-the-path beliefs, the government must reach its information set in every equilibrium under these parameters. Consequently, the equilibrium is essentially unique here. As Figure 1 illustrates, equilibrium strategies depend on the relative likelihood the citizens are sympathetic or unsympathetic. I begin with the case where they are sufficiently likely to be sympathetic:

**Proposition 1.** *If the probability that the citizens are sympathetic is sufficiently high (i.e.,  $q > r^* \equiv \frac{p_M v_M^U - p_L v_L^U - c}{p_L v_L^S - p_M v_M^S + p_M v_M^U - p_L v_L^U}$ ), the terrorist groups pool on attacking. The government’s posterior belief equals its prior, and it undertakes a moderate response.*

See the appendix for all proofs. The intuition for Proposition 1 is straightforward. When  $q$  is high, the government believes that the citizens are likely sympathetic. In

---

<sup>18</sup>The substantive results only require that the terrorist organization have *better* information than the target government.

# Equilibrium Responses

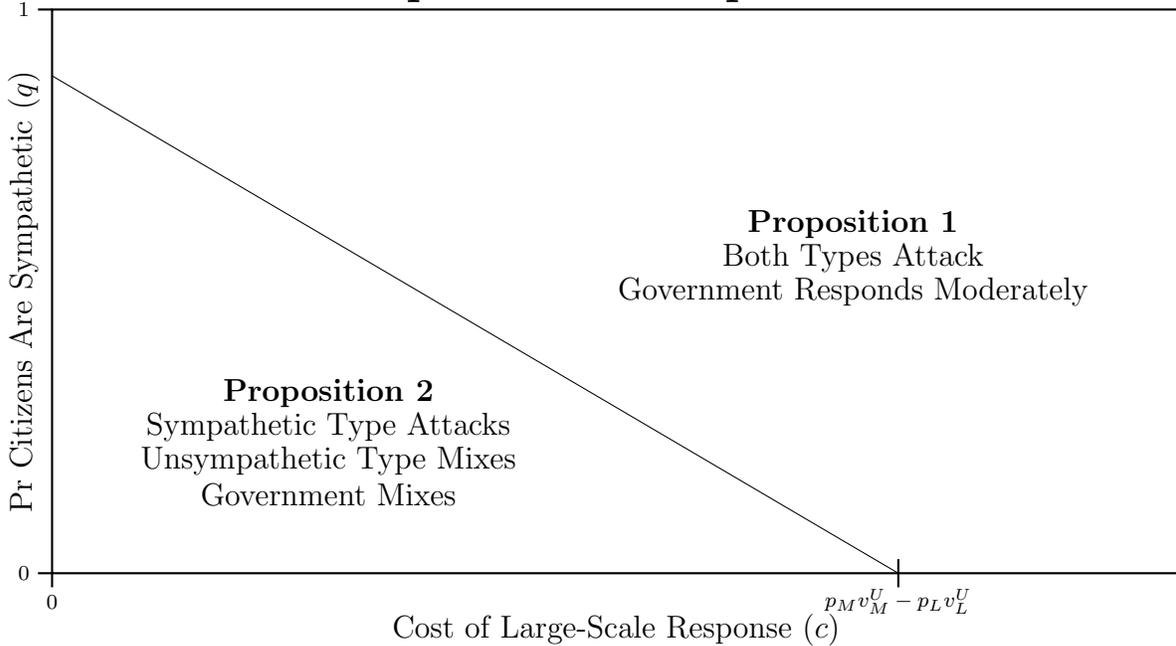


Figure 1: Equilibrium actions as a function of  $q$  and  $c$ . If the cost of a large-scale response is big relative to the probability the citizens are sympathetic, deterrence fails and the terrorist group always attacks. If the cost of a large-scale response is small relative to the probability the citizens are sympathetic, terrorist groups with an unsympathetic audience are sometimes deterred from attacking. The line represents  $r^*$ .

turn, the government worries that a large-scale response will do more harm than good. For the sympathetic type, the decision is simple: the least desirable response to an attack from the government (moderation) is preferable to not attacking. Thus, attacking strictly dominates not attacking for the sympathetic type, so it attacks. The unsympathetic type knows this and knows that the government will infer this as well. In turn, even if the unsympathetic type attacks with certainty, the government will still respond with moderation. This outcome is optimal for the unsympathetic type, so it attacks.

Given that this paper's focus is on rational overreactions, it is interesting to note that such an outcome never occurs under these parameters. Rather, the government's belief that large-scale operations will result in blowback deters it from undertaking a major response. In this light, it is perhaps unsurprising that most responses to terrorism are minimal, mostly involving legal procedures (Duyvesteyn 2008). But this is not

inherently good for the target. Moderation also inspires the weaker, unsympathetic types to commit attacks they would otherwise not get away with. Consequently, more terrorism occurs here than if the target would initiate a large-scale response. The government can still make the wrong move—but it is a matter of *underreaction* here.

In contrast, overreactions occur in the remaining parameter space:

**Proposition 2.** *If the probability that the citizens are sympathetic is sufficiently low (i.e.,  $q < r^*$ ), the terrorist groups semi-separate. The sympathetic type attacks with certainty; the unsympathetic type attacks with probability  $\sigma_A^* \equiv \frac{qr^*}{(1-q)(1-r^*)}$ . The government's posterior belief that it is facing the sympathetic type equals  $r^*$ , and it undertakes a large-scale response with probability  $\sigma_L^* \equiv \frac{p_M v_M^U + a - v_A^U}{p_M v_M^U - p_L U(w(L))}$ .*

Previously, if all the unsympathetic types committed attacks, the relatively large presence of sympathetic types deterred the government from undertaking a large response. For Proposition 2's parameters, however, the unsympathetic type is frequent enough that the government prefers the large response if all terrorist groups always attack. This plays out poorly for the unsympathetic type, which prefers no attack to decimation via a large-scale operation. As such, the terrorist groups cannot pool on attacking in equilibrium.

Full separation cannot occur either. Per the discussion of Proposition 1, the sympathetic type always attacks. Thus, the only separating strategy possible requires the unsympathetic type to not attack. If this were an equilibrium, conditional on an attack, the government would know that the terrorist group is the sympathetic type. It would correspondingly select a moderate response to avoid certain blowback. But this implies that the unsympathetic type has a profitable deviation: it could attack and suffer a moderate response, which is preferable to no attack. Therefore, separation cannot occur in equilibrium.

The remaining possibility is that the unsympathetic type semi-separates. As Proposition 2 describes, the unsympathetic type sometimes commits an attack and sometimes does not. The precise mixed strategy ensures that the government is indifferent between a moderate and a large-scale response. Such indifference permits the government to mix between the two. This mixing guarantees that the unsympathetic type is indifferent between attacking and not attacking, thus ensuring that further attacks are not profitable.

### 3.3 Generality of the Equilibrium

As is common with formal analysis, this paper has investigated only one extensive form of the strategic interaction between a terrorist group and a government. It is therefore worthwhile to think about how the main result regarding overreaction would extend to other game forms.

First, if the presence of mixed strategies is bothersome, one should recall the purification theorem. Rather than think about a single unsympathetic type mixing and a single government mixing, one could imagine that each of these actors is drawn from a continuum with a slight perturbation in their payoffs. For example, the unsympathetic group might have a value for committing an attack drawn from the continuous interval  $[a - \epsilon, a + \epsilon]$  and the government might have a large-scale cost drawn from the interval  $[c - \epsilon, c + \epsilon]$ , with  $\epsilon$  sufficiently small. The purification theorem says that the unsympathetic types with highest values of terrorism commit the act as a *pure* strategy and that governments with smallest costs of large-scale responses engage in that form of counterterrorism as a *pure* strategy. The other types take the opposite action, with overall population parameters reflecting the mixing probabilities above. This means that the equilibrium results and corresponding comparative statics can still hold even if actors only play pure strategies.

Second, as mentioned earlier, the game ends if the terrorist group does not commit an attack. Supposing that the government could identify the group in the absence of an attack, one may wonder whether Proposition 2 holds up if the government could engage in counterterrorism anyway. This may be of particular interest because only the unsympathetic type does not attack in that equilibrium. In turn, the government would update its belief and know with certainty that the non-attacking type is unsympathetic.

To explore this, let  $v_{-A,L}^U$  represent the quantity of volunteers and support conditional on no attack and a large-scale counterterrorist campaign. Naturally, in a provocation framework, support for the group would be greater following such sizable countermeasures given that the terrorist group did not initiate the conflict than compared to an attack followed by a large-scale response. Given this, the government prefers a non-response to a large-scale response if  $-v_{-A}^U > -p_L v_{-A,L}^U - c$ . Although the unsympathetic type gets its label because it is vulnerable to large-scale responses, this inequality would hold if the unprovoked large-scale measures create sympathy for the organization such that  $v_{-A,L}^U$  is sufficiently large. If that is the case, then the government would not

engage in counterterrorism, and the results presented carry over identically.

Finally, recall that the model simplifies the government’s response as a binary choice. One may imagine an alternative model that allows the government to choose an effort level from a continuum of choices, with greater effort costing more, upsetting greater portions of the audience, and killing a greater share of terrorists. Depending on the convexity of this function, the government’s decision may ultimately look like the binary choice presented here, where it would want to exert maximal effort if it believed it was facing an unsympathetic audience but would want to exert minimal effort if it believed it was facing a sympathetic audience.

Alternatively, the government’s maximization problem could produce an interior solution that falls in between its optimal effort against an unsympathetic audience and its optimal effort against a sympathetic audience. Separating equilibria in which the unsympathetic type does not attack and the sympathetic type attacks still fail to exist if the optimal effort versus the sympathetic type is so low that the unsympathetic type would prefer to deviate. Faced with a posterior belief that places positive probability on both types, the government’s optimal response implies an overreaction as defined above. That is, because the government intervenes at a level between the optimal response to each specific type, the government must have *ex post* regret whenever it was actually facing the sympathetic type.

## 4 Empirical Implications

Upon first inspection, it may seem that the above results lack predictive power. After all, for Proposition 2’s parameter range, five different observable outcomes occur: (1) no attack, (2) an attack and a moderate response against a sympathetic audience, (3) an attack and a moderate response against an unsympathetic audience, (4) an attack and a large-scale response against a sympathetic audience, and (5) an attack and a large-scale response against an unsympathetic audience. Given that this paper investigates the plausibility of the overreaction outcomes that (4) describes, one might think that the only further analysis that can be made is that “overreactions sometimes happen.”

However, the observational frequency of each outcome depends on the various parameters of the game. Some outcomes may be rare, while others may be common. I now investigate the circumstances most likely to cause an overreaction.

## 4.1 The Effect of Uncertainty

As the central premise of the paper anticipated, uncertainty plays a major role in determining whether the government overreacts. With complete information, the government has nothing to miscalculate and chooses the optimal countermeasure. If uncertain over how the terrorist group’s audience will react, the government sometimes selects a large-scale response when it would *ex-post* prefer a moderate response.

The rise of Peru’s Shining Path in the 1980s illustrates the mechanism. The model indicates that high levels of latent support and government uncertainty about that support creates a perfect storm for overreaction. Indeed, Shining Path—at least initially—had latent support among its audience. The group originated in Ayacucho, a highlands region about halfway between Lima and Cusco. Citizens in Ayacucho were mostly poor peasants. The central government had created grievances among this population. In the years leading to Shining Path’s ascension, leaders had attempted to collectivize farm lands. Ordinarily, this would help the poorer peasantry, as they would directly benefit from the redistribution. However, incompetent management doomed the collective farms, leaving peasants worse off than they were before (Cronin 2011, 114).

Shining Path capitalized on the peasants’ anger. Its initial assassinations targeted Ayacucho’s wealthier class. “Sympathy for the guerillas, then, was based on notions of economic justice...and there [was] resentment against those who accumulate[d] wealth” (Berg 1994, 115). Following Maoist philosophy, this was part of a broader plan. Shining Path’s leader, Manuel Ruben Abimael Guzmán Reynoso, believed that establishing a base of support was just the initial step (McCormick 1988, 114). From there, they could establish “liberated zones” of territory before beginning a larger uprising across the countryside.

The government in Lima rested blissfully unaware of Shining Path’s latent support in Ayacucho. Part of the problem was geographic distance between Lima and Shining Path’s homeland. Cultural barriers were also a problem. Whereas Spanish descendants dominated Lima, Ayacucho’s population was mostly of native descent. Consequently, the government struggled to comprehend Shining Path’s popularity despite openness about support in Ayacucho (de Wit and Gianotten 1994, 71).

Predictably, the government’s lack of information caused trouble. Shining Path explicitly adopted a provocation strategy during its fourth plenary meeting in May

1981 (Gorriti 2000, 104).<sup>19</sup> Correspondingly, it ramped up the level of violence. After a tepid initial response, the government declared a state of emergency in 1982 and gave its security force free reign in the region. These countermeasures were broad, resulting in mass arrests and disappearances. Security forces killed thousands extrajudicially, resulting in Amnesty International giving Peru the world’s worst human rights record in 1987 (Cronin 2011, 126).

Given the latent sympathy for Shining Path, the population’s response was predictable. The government’s approach only fed the Shining Path’s narrative; whereas Shining Path targeted individuals who had committed financial “crimes,” many taken by the government had done no clear wrongdoing (Berg 1994, 116). The loss of state legitimacy built a large insurgency, leading to 90 provinces to fall under a state of emergency (Cronin 2011, 127). Peasants went from referring to Shining Path as “terrorists” in 1982 to “comrades” in 1985 (Berg 1994, 116). Security forces killed a lot of Shining Path’s lieutenants, but more kept rising from the ashes (Cronin 2011, 19). The government eventually realized its folly and switched to more targeted operations (Palmer 2007, 297-300). It also began arming locals to use their geographic advantage to reduce the threat without needing the government to interfere in local affairs.

Moving back to general results from the model, strong intelligence can mitigate these issues. One can conceptualize that intelligence by considering equilibrium behavior as the government’s prior belief  $q$  approaches 0 or 1—that is, as the government becomes increasingly sure that it is facing an unsympathetic or sympathetic audience. The case when the government increasingly suspects that it is facing a sympathetic type (i.e., as  $q$  goes to 1) is straightforward—at sufficiently high levels of  $q$ , the parameters assuredly fall under Proposition 1, and the government opts for a moderate response with certainty.

When the government believes it is sufficiently likely to be facing an unsympathetic type, the interaction is more complex. Here, the government sometimes opts for a large-scale response. Conditional on that, the probability it is actually facing the sympathetic type is its posterior belief  $r^*$ . Thus,  $r^*$  is also the probability of overreaction conditional on a large-scale response. Note that  $r^*$  is *not* a function of  $q$ . This means that the

---

<sup>19</sup>Shining Path would later release a propaganda pamphlet entitled *Let Us Develop the Guerrilla War!* The second and third sections respectively discuss how they intend the counterrevolution to incite anger and how this will lead to a reactionary popular revolution (Gorriti 2000, 174). These steps mirror the second stage of al-Qaeda’s “Strategy to the Year 2020.”

probability of overreaction conditional on a large-scale response remains fixed. The benefit, however, is that the probability that the government faces an attack at all declines and therefore so does the overall probability of overreaction.

These results on uncertainty indicate that terrorism research ought to further investigate when governments face deeper asymmetric information problems. One obvious empirical implication is that countries undertaking foreign counterterrorism operations are at a major disadvantage compared to governments facing domestic political violence. After all, the marginal cost of information gathering for foreign operations is substantially higher due to geographic, cultural, and linguistic barriers—all problems Peru faced against Shining Path. Correspondingly, the model suggests that internal threats with culturally and linguistically similar to governments are less likely to lead to overreaction.

## 4.2 Probability of Rational Overreaction (and Underreaction)

The model also has implications for material considerations. To begin, take the effects of lower costs on large-scale intervention:

**Proposition 3.** *Conditional on a large-scale response, the probability of overreaction decreases as the cost of large-scale intervention  $c$  increases.*

The formal intuition is as follows. For Proposition 3, conditional on a large-scale response, the probability of overreaction equals the posterior belief that the audience is sympathetic. The unsympathetic type creates this particular belief ( $r^*$ ) by mixing between attacking and not attacking. Thus, the probability of overreaction conditional on a large-scale response is  $r^*$ .

Consider how increasing  $c$  affects this indifference condition. When the cost of large-scale operations is greater, such a strategy looks less attractive to the government. Thus, for the government to be willing to continue mixing, the probability it is facing the unsympathetic type must increase—if not, the government would strictly prefer a moderate response. To fulfill this, the unsympathetic type must attack more often. Yet this means that, conditional on the government choosing a large-scale response, the probability of overreaction *decreases*.

At a more basic level, higher costs of larger operations deter the government from pursuing the more extreme response. Thus, when costs are high, it will only take that

option when it is *very* sure that it is facing the unsympathetic type; if not, the high cost would just backfire. In contrast, when the cost of large-scale responses is low, the government can better absorb mistakes. Consequently, it makes more of them.

Meanwhile, further increasing the cost shifts the parameters from those of Proposition 2 to those of Proposition 1. The government goes from sometimes undertaking the large-scale response to exclusively opting for moderation. Intuitively, this is because an arbitrarily large cost of large-scale interventions guarantees that smaller responses are preferable no matter how many terrorists the large-scale choice can kill.

There are three ways to interpret this finding substantively. First, states that expect to pay low physical costs to engage in large-scale interventions are more likely to see their responses backfire. However, the  $c$  term implicitly also measures “resolve” as the bargaining model of war literature often conceives of it (Fearon 1995). States that naturally find military measures less costly compared to the threat of terrorist violence find their  $c$  values to be lower than states willing to incur terrorist attacks. Similarly, states better capable of absorbing the costs of intervention economically implicitly have lower  $c$  values. States that fit either case are more likely to suffer from the overreaction problem.

Now consider how the effectiveness of large-scale interventions changes the probability of overreaction. One might initially believe that increasing  $p_L$  makes it impossible for the government to overreact. After all, for values of  $p_L$  close to 0, large-scale military operations kill almost all of the recruits, thus rendering the additional recruits irrelevant. And to some extent, this is true—sufficiently small  $p_L$  forces Condition 1 to fail, which then implies that even a group with a sympathetic audience would not attack. But this is not true in less extreme cases:

**Proposition 4.** *Conditional on a large-scale response, the probability of overreaction increases as the proportion of terrorists that large-scale interventions kill increases (i.e., as  $p_L$  decreases).*

The formal and substantive intuition follows similarly to that of Proposition 3. Within the confines of the equilibrium play from Proposition 2, the government’s indifference condition must hold as  $p_L$  fluctuates. When  $p_L$  decreases, large-scale operations are more attractive because they kill a greater percentage of terrorist operatives. Thus, to maintain the indifference, the sympathetic type—the type for which large-scale op-

erations is the wrong choice—must be more common. Because the sympathetic type always wishes to attack regardless of the government’s response, the unsympathetic type must attack less frequently to maintain the indifference. As such, conditional on choosing a large-scale operation, the government is more likely to be wrong as  $p_L$  decreases.

On a more basic level, recall that low values of  $p_L$  imply that the punishment for overreaction is unimportant; it may still be preferable to select a moderate response to the sympathetic type, but large-scale interventions kill so many terrorists that the difference becomes increasingly irrelevant. This dynamic, in turn, makes the government more willing to pursue large-scale interventions. Anticipating the more frequent large-scale response, unsympathetic types become less inclined to attack. However, the government knows this and therefore understands that a large-scale action is more likely to backfire. Yet it does not care—large-scale operations produce more terrorists in expectation, but the effectiveness of such a response still kills a significant proportion of them.

Nevertheless, making interventions sufficiently ineffective causes the parameters shift to that of Proposition 1. Now  $p_L$  is so large that the government prefers a moderate response even if it knows that the unsympathetic type always attacks. Overreaction never occurs because the poor prospects for a large-scale response deter it from ever going down that route.

Taken together, Propositions 3 and 4 predict that rich, militarily powerful countries are most likely to observe their large-scale responses backfire. Making causal claims about provocation and repression is therefore difficult due to selection into large-scale responses. Adler (2010, 209), for example, argues that Western democracies like the United States are “hard-pressed to withstand the political and emotional pressure to retaliate.” And examples of Western democracies making such mistakes are plentiful (e.g., Lafree, Dugan, and Korte 2009; Dugan and Chenoweth 2012; Benmelech, Berrebi, and Klor 2015). But these countries are among the richest and most militarily powerful in the world. Consequently, optimal play results in them overreacting more often than other countries. This also suggests the presence of omitted variable bias in regressions on counterterrorism effectiveness that model democracy but do not control for military strength and wealth (e.g., Daxecker and Hess 2013). Without those controls, democracy may appear to correlate with ineffective repression when it is actually capturing the

effect of military strength and wealth permitting the country to be more willing to make mistakes.

Although overreaction drives much of the policy discussion, it is worth noting that *underreaction* is also possible here—i.e., the government chooses a moderate response when it would have *ex-post* preferred a large-scale operation. A corollary to these comparative statics reveals an inexorable tradeoff:

**Corollary 1.** *Conditional on a moderate response, the probability of underreaction is decreasing in the portion of terrorists that large-scale interventions kill and increasing in the cost of large-scale operations.*

Like the overreaction calculations, the posterior probability the government is facing a sympathetic audience determines the likelihood of underreaction. Any change to the parameters that increases the probability of facing a sympathetic audience decreases the probability of facing an unsympathetic type. Thus, Corollary 1 flips Propositions 3 and 4. It also indicates that raising costs and enfeebling the power of large-scale responses does not resolve the government’s issue—it merely substitutes one problem for another.

### 4.3 Value of Deterrence

If lower costs and greater military effectiveness increase the probability of overreaction conditional on a large-scale response, does having seemingly better outside options provide any benefits? One may initially suspect that the lower costs and greater effectiveness outstrip any lost value from overreacting more frequently. However, this intuition is wrong. The logic is straightforward for the parameters of Proposition 1: the government never picks a large-scale response there, so altering  $c$  and  $p_L$  does not change its payoff.

The explanation is more complicated for Proposition 2’s parameters. Note that the indifference conditions require the government to have identical utilities for moderate and large-scale responses. Altering  $c$  and  $p_L$  does not change the payoff for a moderate response. Thus, conditional on an attack and being in the parameters of Proposition 2, manipulating  $c$  and  $p_L$  maintains the government’s payoff. Whatever gains the government makes from superior large-scale responses are exactly offset by the less frequent attacks from the unsympathetic type.

Instead, all the extra value comes from a lower chance of suffering an attack:

**Proposition 5.** *As the cost of large-scale intervention  $c$  increases, the probability the terrorist group attacks weakly increases.*

**Proposition 6.** *As the proportion of terrorists that large-scale interventions kill increases (i.e., as  $p_L$  decreases), the probability the terrorist group attacks weakly decreases.*

The intuition comes from above. As  $c$  and  $p_L$  decrease, the parameters eventually shift from those of Proposition 1 (where attacks occur with certainty) to those of Proposition 2 (where attacks occur with probability less than 1). Within the confines of Proposition 2's parameters, the probability that the unsympathetic type attacks diminishes as  $c$  and  $p_L$  decrease. This is because—as stressed before—the government must remain indifferent between moderate and large-scale responses even as the large-scale choice appears more attractive. But that can only happen if the government faces the sympathetic type more frequently, which requires the unsympathetic type to attack less often. Intuitively, the better outside option deters the unsympathetic type from initiating a crisis.

Quantitative research supports Proposition 5's result. Li and Schaub (2004), for example, find that a state's gross domestic product negatively correlates with transnational terrorist attacks occurring on that state's soil. Nevertheless, finding statistical evidence for Proposition 6 will be comparatively difficult. This is because terror attacks are a “weapon of the weak” that substitute for traditional military confrontations against powerful states (Fortna 2015). Thus, standard measures of military strength understandably positively correlate with terrorism. Future empirical research therefore ought to disaggregate the substitution effect from the vulnerability that groups with unsympathetic audiences face. Crenshaw (2015) provides one possibility: disaggregate provocative terrorist attacks from the rest and test whether provocative attacks are relatively less likely.

Meanwhile, Proposition 5's comparative static suggests that a terrorist attack from one organization begets terrorist attacks from other organizations. States have limited military resources that they can use to coerce opponents (Triesman 2004). Further, when confrontations last long periods of time, domestic factions can develop war exhaustion, causing them to find additional conflicts progressively more unattractive

(Richardson 1960). With that in mind, suppose that a state has implemented a large-scale response to a terrorist attack. The limited resources and war exhaustion factors manifest themselves in a new interaction with a different terrorist group by increasing the cost of a second large-scale intervention. But as  $c$  increases, the probability the unsympathetic type attacks increases. Counterterrorism can create a vicious cycle.

That said, there is a countervailing factor to consider. Although the costs of another large-scale intervention may increase, the experience of past fighting should make the next fight more effective. Proposition 6 states that decreasing  $p_L$  (that is, making the large-scale response more effective) decreases the probability of the unsympathetic type attacking. Thus, the exact effect depends on how much  $c$  increases versus how much  $p_L$  decreases.

## 5 Conclusion

This paper explored a rationalist explanation for overreaction to terrorism. Terrorist organizations have asymmetric knowledge of how their constituencies will respond to military interventions. When sympathies are strong, the group can attack without hesitation. This gives terrorist groups with weaker sympathies an incentive to bluff strength by attacking. Consequently, governments cannot be sure how to respond to a terrorist attack. They correspondingly often undertake large-scale responses to call potential bluffs, knowing that this strategy will backfire some portion of the time.

This approach breaks from the standard recommendations to provocation, which is to simply resist temptation to respond heavily. Indeed, this is easy to see by considering how terrorists with unsympathetic audiences would respond to a blanket policy of moderation. The threat of large-scale responses deters these groups. Whatever gains the government earns by assuredly not overreacting to groups with sympathetic constituencies is more than offset by the increased frequency of attacks. Thus, the provocation literature has missed out on these hidden costs of moderate action—some terrorist attackers may in fact find large-scale responses devastating to their causes.

It also indicates that victims of terrorism should not limit their intelligence gathering efforts to the terrorist groups that target them. Rather, a full understanding of how the audience will respond to counterterrorist operations is critical to successful responses. The absence of this information leads to provocation strategies, entrapping targets into

wasteful countermeasures or missed opportunities to dismantle terrorist organizations. Either way, the cost of uncertainty can be high.

## 6 Appendix

This appendix gives proofs for all of the formal claims made in the body of the paper.

### 6.1 Proof of Propositions 1 and 2

Let  $r$  represent the government's posterior belief that the terrorist group is sympathetic. Then the government prefers a large-scale operation if:

$$r(-p_L v_L^S - a - c) + (1 - r)(-p_L v_L^U - a - c) > r(-p_M v_M^S - a) + (1 - r)(-p_M v_M^U - a)$$

$$r < r^* \equiv \frac{p_M v_M^U - p_L v_L^U - c}{p_L v_L^S - p_M v_M^S + p_M v_M^U - p_L v_L^U}$$

Analogously, the government prefers a moderate response if  $r > r^*$  and is indifferent between its options if  $r = r^*$ .

Now for the terrorist organization's decision. As described above, via Condition 1, the sympathetic type has a strictly dominant strategy to attack. Condition 2, however, shows that the unsympathetic type prefers attacking if the government responds with a moderate operation but wishes to not attack if the government follows with a large-scale assault.

So consider separating, pooling, and semi-separating equilibria. Because the sympathetic type must attack, the only separating possibility requires the unsympathetic type to not attack. The government's posterior equals 1 here, so it would select a moderate response. But the unsympathetic type could then profitably deviate to attacking. Thus, no separating equilibria exist.

The only pooling possibility requires both types to attack. Here, the government's posterior equals its prior. For the unsympathetic type to not want to deviate, the government must choose a moderate response. Per above, the government is willing to do this if  $r > r^*$ . Given that the unsympathetic type is pooling, this would then require  $q > r^*$ . This proves Proposition 1.

If  $q < r^*$ , however, pooling equilibria fail to exist. The remaining possibility is a semi-separating equilibrium. Because the sympathetic type must attack, checking for a semi-separating equilibrium entails looking for a mixed strategy  $\sigma_A$  that represents the unsympathetic type's probability of attacking. The indifference condition correspondingly mandates that the unsympathetic type receive equal payoffs for attacking and not attacking. But because the unsympathetic type receives a strictly greater payoff for a moderate response than inaction and a strictly greater payoff for inaction than a large-scale response, the government must also mix to maintain the unsympathetic type's indifference.

Now to calculate those values. Recall that the government is indifferent between moderate and large-scale responses when  $r = r^*$ . Using that indifference condition, the unsympathetic type attacks with probability:

$$\frac{\sigma_A(1-q)}{\sigma_A(1-q) + q(1)} = r^*$$

$$\sigma_A^* \equiv \frac{qr^*}{(1-q)(1-r^*)} = \frac{q \left( \frac{p_M v_M^U - p_L v_L^U - c}{p_L v_L^S - p_M v_M^S + p_M v_M^U - p_L v_L^U} \right)}{(1-q) \left( 1 - \frac{p_M v_M^U - p_L v_L^U - c}{p_L v_L^S - p_M v_M^S + p_M v_M^U - p_L v_L^U} \right)}$$

Meanwhile, let  $\sigma_L$  represent the probability that the government selects a large-scale response. The indifference condition for the unsympathetic type allows us to solve for  $\sigma_L$ :

$$v_{-A}^U = \sigma_L [p_L v_L^U + a] + (1 - \sigma_L) [p_M v_M^U + a]$$

$$\sigma_L^* = \frac{a + p_M v_M^U - v_{-A}^U}{p_M v_M^U - p_L v_L^U}$$

These are the parameters, strategies, and beliefs described in Proposition 2.  $\square$

## 6.2 Proof of Proposition 3

Proposition 3 refers to the probability that the government is facing the sympathetic type given that it selects a large-scale response. Large-scale responses only occur under Proposition 2's parameters. Thus, if the government pursues that option, its belief equals  $r^* = \frac{p_M v_M^U - p_L v_L^U - c}{p_L v_L^S - p_M v_M^S + p_M v_M^U - p_L v_L^U}$ . The first derivative with respect to  $c$  is:

$$-\frac{1}{p_L v_L^S - p_M v_M^S - p_M v_M^U + p_L v_L^U}$$

This is negative. Therefore, conditional on a large-scale response, increasing cost of large-scale operations decreases the probability of an overreaction.  $\square$

### 6.3 Proof of Proposition 4

Following the proof for Proposition 3, the critical piece of information is how  $r^*$  changes as a function of  $p_L$ . The derivative of the full value with respect to  $p_L$  is complicated, so the clearer proof involves looking at how the numerator and denominator individually change as  $p_L$  increases.

Recall that the numerator equals  $p_M v_M^U - p_L v_L^U - c$ . Since  $v_L^U$  is positive, increasing the always positive  $p_L$  decreases the numerator. Meanwhile, we can write the denominator as  $p_L[v_L^S - v_L^U] - p_M[v_M^S - v_M^U]$ . Because  $v_L^S > v_L^U$ , increasing  $p_L$  always increases the denominator. With both the numerator and denominator always positive, increasing  $p_L$  makes  $r^*$  smaller. Therefore, increasing the proportion of terrorists who survive a large-scale response decreases  $r^*$ , the probability of overreaction conditional on the large response.  $\square$

### 6.4 Proof of Propositions 5 and 6

Note that the probability of an attack is 1 for Proposition 1's parameters and  $q + (1 - q)\sigma_A^*$  for Proposition 2's parameters. Consider how the probability of attack changes in the second case. Substituting for  $\sigma_A^*$ , the probability of attack equals  $\frac{q}{1-r^*}$ . Recall from the proof for Proposition 3 that  $r^*$  is decreasing in  $c$ . As such,  $\frac{q}{1-r^*}$  is increasing in  $c$ .

Meanwhile, recall from the proof for Proposition 4 that  $r^*$  is decreasing in  $p_L$ . In turn,  $\frac{q}{1-r^*}$  is decreasing in  $p_L$ .  $\square$

## 7 References

Adler, Emanuel. 2010. "Damned If You Do, Damned If You Don't: Performative Power and the Strategy of Conventional and Nuclear Defusing." *Security Studies* 19 (2): 199-229.

- Atwan, Abdel Bari. 2008. *The Secret History of al Qaeda*. Berkeley: University of California Press.
- Bell, J. Bowyer. 1977. *Terror out of Zion: Irgun Zvai Leumi, LEHI, and the Palestine Underground, 1929-1949*. New York: St. Martin's Press.
- Benmelech, Efraim, Claude Berrebi, and Esteban F. Klor. 2015. "Counter-Suicide-Terrorism: Evidence from House Demolitions." *Journal of Politics* 77 (1): 27-43.
- Berg, Ronald H. 1994. "Peasant Responses to Shining Path in Andahuaylas." In *Shining Path of Peru, Second Edition* Ed. David Scott Palmer. New York: St. Martin's Press.
- Berry, Nicholas O. 1987. "Theories on the Efficacy of Terrorism." *Journal of Conflict Studies* 7 (1): 7-20.
- Bloom, Mia. 2005. *Dying to Kill: The Allure of Suicide Terror*. New York: Columbia University Press.
- Brophy-Baermann, Bryan and John A.C. Conybeare. 1994. "Retaliating against Terrorism: Rational Expectations and the Optimality of Rules versus Discretion." *American Journal of Political Science* 38 (1): 196-210.
- Bueno de Mesquita, Ethan. 2005. "The Quality of Terror." *American Journal of Political Science* 49 (3): 515-530.
- Bueno de Mesquita, Ethan. 2007. "Politics and the Suboptimal Provision of Counterterror." *International Organization* 61 (1): 9-36.
- Bueno de Mesquita, Ethan, and Eric S. Dickson. 2007. "The Propaganda of the Deed: Terrorism, Counterterrorism, and Mobilization." *American Journal of Political Science* 51 (2): 364-381.
- Carter, David. 2016. "Provocation and the Strategy of Terrorist and Guerilla Attacks." *International Organization*, Forthcoming.
- Crenshaw, Martha. 1981. "The Causes of Terrorism." *Comparative Politics* 13 (4): 379-399.
- Crenshaw, Martha. 2015. "Terrorism Provocations, Retaliation and Deterrence." Manuscript, Stanford University.
- Cronin, Audrey Kurth. 2011. *How Terrorism Ends: Understanding the Decline and Demise of Terrorist Campaigns*. Princeton: Princeton University Press.
- Daxecker, Ursula E. and Michael L. Hess. 2013. "Repression Hurts: Coercive Government Responses and the Demise of Terrorist Campaigns." *British Journal of Political Science* 43 (3): 559-577.
- de Wit, Ton and Vera Gianotten. 1994. "The Center's Multiple Failures." In *Shining Path of Peru, Second Edition* Ed. David Scott Palmer. New York: St. Martin's Press.
- Dugan, Laura and Erica Chenoweth. 2012. "Moving Beyond Deterrence: The Effectiveness of Raising the Expected Utility of Abstaining from Terrorism in Israel." *American Sociological Review* 77 (4): 597-624.

- Duyvesteyn, Isabelle. 2008. "Great Expectations: The Use of Armed Force to Combat Terrorism." *Small Wars and Insurgencies* 19 (3): 328-351.
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379-414.
- Fortna, Virginia Page. 2015. "Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes." *International Organization*, Forthcoming.
- Fromkin, David. 1975. "The Strategy of Terrorism." *Foreign Affairs* 53 (4): 683-698.
- Funes, Maria J. 1998. "Social Responses to Political Violence in the Basque Country: Peace Movements and Their Audience." *Journal of Conflict Resolution* 42 (4): 493-510.
- Gorriti, Gustavo. 2000. *The Shining Path: A History of the Millenarian War in Peru*. Chapel Hill: University of North Carolina Press.
- Gude, Edward W. 1969. "Batista and Betancourt: Alternative Responses to Violence." In *The History of Violence in America* Ed. Ted Robert Gurr. New York: Bantam Books.
- Hoffman, Bruce. 2013. *Inside Terrorism: Revised and Expanded*. New York: Columbia University Press.
- Kaplan, Edward H., Alex Mintz, Shaul Mishal, and Claudio Samban. 2005. "What Happened to Suicide Bombings in Israel? Insights from a Terror Stock Model." *Studies in Conflict and Terrorism* 28 (3): 225-235.
- Kydd, Andrew H. and Barbara F. Walter. 2002. "Sabotaging the Peace: The Politics of Extremist Violence." *International Organization* 56 (2): 263-296.
- Kydd, Andrew H. and Barbara F. Walter. 2006. "The Strategies of Terrorism." *International Security* 31 (1): 49-80.
- LaFree, Gary, Laura Dugan, and Raven Korte. 2009. "The Impact of British Counterterrorist Strategies on Political Violence in Northern Ireland: Comparing Deterrence and Backlash models." *Criminology* 47 (1): 17-45.
- Lake, David A. 2002. "Rational Extremism: Understanding Terrorism in the Twenty-First Century." *Dialogue IO* 1 (1): 15-29.
- Lapan, Harvey E. and Todd Sandler. 1993. "Terrorism and Signalling." *European Journal of Political Economy* 9 (3): 383-397.
- Laqueur, Walter. 1987. *The Age of Terrorism*. Boston: Little, Brown.
- Li, Quan and Drew Schaub. 2004. "Economic Globalization and Transnational Terrorism: A Pooled Time-Series Analysis." 48 (2): 230-258.
- Lindekilde, Lasse. 2014. "A Typology of Backfire Mechanisms." In *Dynamics of Political Violence: A Process-Oriented Perspective on Radicalization and the Escalation of Political Conflict*. Eds. Lorenzo Bosi, Charles Demetriou, and Stefan Malthaner. Farnham: Ashgate.
- Mao Tse-Tung. 1937. *On Guerilla Warfare*. Translated by Samuel B. Griffith. Campaign: University of Illinois Press.

- Mishal, Shaul, and Avraham Sela. 2006. *The Palestinian Hamas: Vision, Violence, and Coexistence*. New York: Columbia University Press.
- Marighella, Carlos. 1971. *For the Liberation of Brazil*. New York: Penguin.
- McCormick, Gordon H. 1988. "The Shining Path and Peruvian Terrorism." In *Inside Terrorist Organizations* Ed. David C. Rapoport. New York: Columbia University Press.
- Moloney, Ed. 2003. *A Secret History of the IRA*. New York: Norton.
- Mueller, John E. 2006. *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them*. New York: Simon and Schuster.
- Muller, Edward N. and Karl-Dieter Opp. 1986. "Rational Choice and Rebellious Collective Action." *American Political Science Review* 80 (2): 471-487.
- Naji, Abu Bakr. 2004. *The Management of Savagery*. Translated by William McCants. John M. Olin Institute for Strategic Studies, Harvard University.
- Overgaard, Per Baltzer. 1994. "The Scale of Terrorist Attacks as a Signal of Resources." *Journal of Conflict Resolution* 38 (3): 452-478.
- Palmer, David Scott. 2007. "Countering Terrorism in Latin America: The Case of Shining Path in Peru." In *Countering Terrorism and Insurgency in the 21st Century, Volume 3*. Ed. James J.F. Forest. Westport: Praeger.
- Pape, Robert. 2005. *Dying to Win: The Strategic Logic of Suicide Terrorism*. New York: Random House.
- Post, Jerrold M. 1998. "Terrorist Psycho-Logic: Terrorist Behavior as a Product of Psychological Forces." In *Origins of Terrorism: Psychologies, Ideologies, Theologies, States of Mind* Ed. Walter Reich. Baltimore: Johns Hopkins University Press.
- Powell, Robert. 1999. *In the Shadow of Power: States and Strategies in International Politics*. Princeton: Princeton University Press.
- Price, Jr., H. Edward. 1977. "The Strategy and Tactics of Revolutionary Terrorism." *Comparative Studies in Society and History* 19 (1): 52-66.
- Richardson, Lewis F. 1960. *Arms and Insecurity*. Pittsburgh: Boxwood Press.
- Rosendorff, B. Peter and Todd Sandler. 2004. "Too Much of a Good Thing? The Proactive Response Dilemma." *Journal of Conflict Resolution* 48 (5): 657-671.
- Savun, Burcu and Brian J. Phillips. 2009. "Democracy, Foreign Policy, and Terrorism." *Journal of Conflict Resolution* 53 (6): 878-904.
- Schmid, Alex P. and Albert J. Jongman. 1988. *Political Terrorism: A New Guide to Actors, Authors, Concepts, Data Bases, Theories, and Literature*. North-Holland: Transaction Publishers.
- Siqueira, Kevin and Todd Sandler. 2006. "Terrorists versus the Government: Strategic Interaction, Support, and Sponsorship." *Journal of Conflict Resolution* 50 (6) 878-898.
- Thornton, Thomas Perry. 1964. "Terror as a Weapon of Political Agitation." In *Terrorism: Critical Concepts in Political Science* Ed. David Rapoport. New York:

Routledge.

- Treisman, Daniel. 2004. "Rational Appeasement." *International Organization* 58 (2): 345-373.
- Weitz, Richard. 1986. "Insurgency and Counterinsurgency in Latin America, 1960-1980." *Political Science Quarterly* 101 (3): 397-413.
- Woodworth, Paddy. 2001. "Why Do They Kill? The Basque Conflict in Spain." *World Policy Journal* 18 (1): 1-12.
- Zirakzadeh, Cyrus Ernesto. 2002. "From Revolutionary Dreams to Organizational Fragmentation: Disputes over Violence within ETA and Sendero Luminoso." *Terrorism and Political Violence* 14 (4): 66-92.