Torture in Counterterrorism: Agency Incentives and Slippery Slopes

Hugo M. Mialon, Sue H. Mialon, and Maxwell B. Stinchcombe[1]

July 22, 2011

Abstract

We develop a counterterrorism model to analyze the effects of allowing a government agency to torture suspects when evidence of terrorist involvement is strong. We find that legalizing torture in strong-evidence cases has offsetting effects on agency incentives to counter terrorism by means other than torture. It lowers these incentives because the agency may come to rely on torture to avert attacks. However, it also increases these incentives because other efforts may increase the probability of having strong enough evidence to warrant the use of torture. Parametric analysis indicates that it is more likely to reduce non-torture efforts if these efforts are more effective at stopping attacks and less effective at turning up strong evidence when the suspect is guilty. In general, if it reduces non-torture efforts, it can reduce security and is more likely to do so if the attack threat is higher. Moreover, if torture is used in strong-evidence cases even if torture is banned, legalizing torture in strong-evidence cases necessarily reduces security if it reduces non-torture efforts. Lastly, it can increase incentives to torture even in weak-evidence cases—a slippery slope. (JEL H1, K4)

[1] Hugo Mialon and Sue Mialon, Department of Economics, Emory University, Atlanta GA 30322-2240 (hmialon@emory.edu and smialon@emory.edu). Maxwell Stinchcombe, Department of Economics, University of Texas at Austin, Austin TX 78712-1173 (maxwell@eco.utexas.edu). We are extremely grateful to the Editor, two anonymous referees, Satyajit Chatterjee, Martin Dumav, Amy Farmer, Andrew Francis, Louis Kaplow, Tilman Klumpp, Preston McAfee, Tim Mathews, Erik Nesson, Josh Robinson, Paul Rubin, Louis Seidman, Steve Shavell, Jeroen Swinkels, Kevin Tsui, Adam Winship, Kathy Zeiler, and seminar participants at Clemson, Emory, Harvard, University of Texas, and the 2009 SEA Meeting for helpful comments.

1. Introduction

Since the September 11, 2001 terrorist attacks on the U.S., a number of legal authorities, political authorities, and poll results have favored the use of torture in counterterrorism. Many arguments in favor of the use of torture begin with, or prominently feature, some variant of a "ticking bomb" scenario in which torturing one suspect leads, with near certainty, to saving many lives. However, if a ticking bomb scenario arises, it is because other preventive efforts have failed. In this paper, we analyze the effects of legalizing torture that arise through substitution within the portfolio of counterterrorism efforts.

The Push for Torture. In 2002, as Assistant Attorney General in the U.S. Department of Justice, John Yoo recommended that the White House withdraw its recognition of the rules prohibiting torture imposed by the Geneva Conventions.[2] In an influential book, Dershowitz (2002) proposed controlling torture through a system of judicial warrants for the use of torture in limited circumstances. In 2004, Senator Charles Schumer publicly rejected the idea that torture should never be used.[3] In 2007, Attorney General Michael Mukasey refused to state that waterboarding is illegal.[4] In 1987, the Supreme Court of Israel consented to the use of torture to stop attacks by Palestinian terrorists (Imseis, 2001, and Strauss, 2003). In a 1999 decision, the Court returned to a ban on torture, but the ban is not absolute and has not consistently been enforced as discussed by Imseis (2001). The Economist (2006) reported results of public opinion polls about torture. In a BBC poll of 27,000 people in 25 countries, 33 percent of people polled, 36 percent of Americans polled, and 46 percent of Israelis polled said that "some degree of torture is permissible." In a poll

[2] Memorandum from John Yoo, Deputy Assistant Attorney General, and Robert J. Delahunty, Special Counsel, to William J. Haynes II, General Counsel, Department of Defense (Jan. 9, 2002).
[3] Federal Government's Counterterrorism Efforts: Hearing Before the Sen. Judiciary Subcommittee, 108th Cong. (2004) (statement of Sen. Charles Schumer, Member, S. Judiciary Committee).
[4] Executive Nomination: Hearing on the Nomination of Michael B. Mukasey To Be Attorney General of the United States Before the S. Comm. on the Judiciary, 110th Cong. (2007). See Egan (2007)

of 2,000 people living in the U.S., which was conducted by The Pew Research Foundation (2009) between October and November of 2009, 54 percent of people polled thought that "torturing terrorist suspects is often or sometimes justified."

The Ticking Bomb Scenario. Most arguments in favor of legalizing torture are presented as variants of the following mass terrorism scenario.[5] Suppose the government learns that terrorists are planning an attack in a populated area. If the attack succeeds, many people could die. The government has in custody a suspect who may know about the attack but refuses to cooperate. The government can perhaps force the suspect to reveal what he knows through torture. The suspect could survive the torture, and the information extracted could save many lives. The government does not, at that time, have other means of stopping the attack. In some versions of this scenario, cost-benefit analysis suggests that allowing the government to torture the suspect is the socially efficient policy.

In this paper, we examine this efficiency argument more carefully. There are unintended consequences to allowing torture, and those consequences vary with the degree of evidence required to make torture legal. Specifically, if a ticking bomb scenario arises, it reflects a failure of other preventive efforts.[6] Assuming that situations can arise in which the cost-benefit calculations described above favor torture, we develop a model of counterterrorism to evaluate the overall effects of legalizing torture. Throughout, we maintain the assumption that there is a fundamental agency problem: the counterterrorism agency puts, relative to society, more weight on preventing terrorist attacks than on protecting individual rights.

When The Agency Obeys Directives. In the first part of our analysis, we assume the agency obeys directives on torture. In this case, allowing the agency to torture when

[5] See e.g. Dershowitz (2002), Strauss (2003), Luban (2005), and Bagaric and Clarke (2006).

[6] For domestically based attacks, such efforts include tracking of materials used in bomb-making, restrictions on bomb-making activities, increased security at likely targets, and baggage and cargo screening at airports. For extraterritorial sources, such efforts include hiring, training and paying attention to analysts fluent in language and culture, cultivating allies, and bilateral or multilateral cooperative international policing.

evidence of guilt is strong has two opposing effects on the agency's incentives to counter terrorism by means other than torture; first, it tends to reduce these incentives because it ameliorates a situation in which other efforts have failed–a decommitment effect; second, it tends to increase these incentives because other efforts improve the chances of gaining enough evidence to warrant the use of torture if other efforts fail–a complementarity effect.

Parametric analysis indicates that the decommitment effect is more likely to dominate if the agency's non-torture efforts are more effective at stopping attacks and less effective at turning up strong evidence when the suspect is guilty. Furthermore, allowing torture in a broader range of cases lowers the complementarity effect. When the decommitment effect dominates, legalizing torture reduces the agency's other preventive efforts. In this case, we have a formalization of the observation in Rejali (2007) that reliance on torture typically makes an agency sloppier in its other preventive work and leads to agency "deskilling." [7] If legalizing torture reduces the agency's preventive effort, it can reduce security, and it is more likely to do so if the attack threat is higher. Moreover, it reduces welfare if the costs of torturing the innocent are sufficiently high.

The Enforcement Problem. The core of the Dershowitz (2002) argument is that agen-

[7] Rejali (*op. cit.* Chap 21 and 22) documents the deskilling effect across a large number of 19'th and 20'th century instances. For example, there is evidence that the Gestapo's suppression of the Resistance in World War II was far more effective when it relied on informers and careful interrogation before it turned extensively and "unprofessionally" to torture. In the French-Algerian war, French army units that tortured became insubordinate to central army authority and abandoned basic police techniques – in one instance, going directly to torture rather than checking the personal effects of an apprehended suspect allowed an Algerian resistance bomb factory to be safely relocated. The radical part of the Algerian resistance movement followed a policy of identifying members of the moderate opposition when tortured, and because the French army did not check the veracity of what was revealed under torture, it wiped out the moderate opposition.

For a more recent example, when Abdul Hakim Murad was arrested by Filipino police in 1995 with fake passports, bomb-making materials, and an encrypted computer, police tortured him instead of trying to decrypt the computer. He revealed little specific information under torture, but when the CIA decrypted his computer years later, it revealed detailed information about Al Qaeda plots to blow up planes in the US, down to specific procedures and flight schedules.

The CIA's unedited *Human Resources Exploitation Training* manual summarizes the deskilling effect of torture with "The routine use of torture ... corrupts those that rely on it as the quick and easy way out" (See the National Security Archives at website http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB122/).

cies do not obey directives, torture happens even though it is illegal, and an enforced system of judicial warrants could bring this under control, resulting in less torture of the innocent. In the second part of our analysis, we extend our model to encompass the possibility that the agency is willing to disobey directives on torture at the risk of legal sanction.[8] In the extension, the agency can choose whether to use torture even when torture is not allowed, and it faces potential punishment if it uses torture illegally. If torture is legal in strong-evidence cases, the agency only faces potential punishment if torture in weak-evidence cases.

In this extended context, there are conditions under which the agency's optimal torture policy is to use torture in strong-evidence cases whether or not torture is legal in strong-evidence cases. We find that, if these conditions hold, then legalizing torture in strong-evidence cases has the previous two effects, decommitment and complementarity, as well as a third, more subtle, decomplementarity effect, on the agency's non-torture efforts. First, it tends to reduce such efforts because it eliminates the agency's cost of using torture in strong-evidence cases (decommitment). Second, it tends to increase such efforts because the agency thereby increases the chances of obtaining strong evidence if the efforts fail to stop an attack, in which case the agency can torture with impunity (complementarity). Third, it tends to reduce such efforts because the agency thereby increases the chances that it has strong evidence even if there turns out to be no attack, in which case the agency can escape punishment for using torture on innocent individuals (decomplementarity).

If the complementarity effect dominates, legalizing torture in strong-evidence cases increases the agency's other efforts, and thereby increases safety and reduces the probability of torturing the innocent. This case supports the Dershowitz argument that an open warrant system would actually increase incentives to obey the law and reduce torture of the inno-

[8] In wartime and on foreign grounds, the risk that torture will be used illegally may be high, as evidenced by the documented reports of sadistic torture at the Abu Ghraib and Guantanamo Bay detention facilities of the U.S. Military (Fay, 2004).

cent. However, legalizing torture in strong-evidence cases reduces the agency's other efforts, and thereby reduces safety and increases torture of the innocent, if the decommitment and decomplementarity effects dominate. Parametric analysis indicates that these two effects dominate if the agency's non-torture efforts are sufficiently effective at stopping attacks.

Slippery Slopes. We also find that legalizing torture in strong-evidence cases can lead to an increase in its use in other cases—a slippery slope. Intuitively, this involves the endogeneity of the quality of information. If the decommitment and decomplementarity effects dominate, legalizing torture in strong-evidence cases reduces the agency's efforts to counter terrorism by means other than torture, which in turn reduces the quality of the information on which the agency bases its decision to use torture if the other efforts fail, increasing agency incentives to use torture in other cases.

This mechanism differs from the three basic variants of the slippery slope arguments that we have found in the literature: utility change, cost change, and bureaucratic structure arguments. Volokh (2003, p. 1077) elegantly summarizes the utility change arguments as "the normative power of the actual." In more pedestrian language, a society that allows torture will perhaps come to see nothing wrong with it. Volokh (2003) and Rizzo and Whitman (2004) provide a detailed examination of cost-based slippery slope mechanisms in legal policymaking. These involve one decision lowering the cost to make another linked decision. In our context, if society pays the cost of training and supporting professional torturers, then the lower marginal cost of torture can lead to more frequent torture. Sobel (2000) provides a model of declining standards that may bear on the worry that any chosen evidence standard for torture may be prone to slip over time, perhaps by the accretion of precedents set by judges who are more sympathetic to agency arguments for torture.

Nonetheless, it is possible that a legal standard for torture would not slip if torture were

legalized, just as the legal standard for capital punishment does not seem to have slipped after capital punishment was legalized (Bagaric and Clarke, 2006). However, according to the slippery slope mechanism that we identify here, even if the legal standard for torture were not to slip, legalizing torture in certain circumstances could still entail a slippery slope in that it could increase illegal torture in other circumstances.

Models of Torture in the Literature. There is a growing literature on the economics of terrorism (see Enders and Sandler, 2004 and 2005, Siqueira and Sandler, 2007, Garoupa, Klick, and Parisi, 2006, and Berman and Laitin, 2008). However, this literature has not considered the use of torture in counterterrorism. If one considers avoidance of torture an individual right, the emerging literature on the economics of individual rights is relevant (see Mialon and Rubin, 2008, for a synthesis). Seidmann and Stein (2000), Mialon (2005), Leshem (2010), and Wickelgren (2010) examine the right to silence. These papers analyze the effects of preventing adverse inferences from a suspect's silence during interrogation, but they do not analyze the more fundamental right against torture in interrogation.

The only formal analyses of torture that we have found in the literature are Wantchekon and Healy (1999), Chen, Tsai, and Leung (2009), Chen, Chou, and Tsai (2009), and Baliga and Ely (2010). Wantchekon and Healy analyze torture as a game of incomplete information between a torturer and a victim. They show that torture is carried out with positive probability in equilibrium because even a weak victim might hold out to test whether the torturer is professional or sadistic and even a professional might torture to test whether the victim is weak. Chen, Chou, and Tsai model torture as a war of attrition. They show that torture may occur because of the torturer's uncertainty about the suspect's limit of pain endurance and the suspect's uncertainty about the torturer's limit on pain infliction. In their model, torture can be effective because the innocent cannot verifiably confess while

the guilty can. Thus, if torture is applied long enough, guilty suspects will confess and only innocent suspects will be tortured. Baliga and Ely show that torture may be largely ineffective if the torturer cannot commit to continue torturing a suspect he knows to be uninformed and to stop torturing a suspect he knows to be informed. In our paper, we do not focus on the strategic interaction between torturer and suspect but instead focus on the effects of legalizing and regulating torture on the behavior of the counterterrorism agency. Moreover, unlike the above papers, we consider the implications for security and torture of the innocent of agency problems and problems enforcing directives on torture.

Organization of the Paper. Section 2 presents our basic model of counterterrorism when the agency obeys directives on torture. Section 3 extends the model to consider the enforcement problems that arise from the agency being willing to run the risk of legal sanction. Section 4 summarizes.

2. When the Agency Obeys Directives

We begin with a description of the basic model of a government agency (e.g., the CIA, FBI, or DOD) and an individual who may have initiated a terrorist action. Under the assumption that the agency obeys directives about torture, we compare outcomes, in terms of agency behavior and social welfare, under three scenarios: torture is illegal; torture is illegal except in the face of strong evidence of suspect guilt; and torture is legal. The next section extends the model to study the enforcement problem that arises if the agency may choose, at the cost of legal sanctions, to undertake torture.

2.1 The Basic Model

At time 1, an individual or group of individuals in a large population initiates a terrorist action with probability a, and the agency apprehends an individual. If a terrorist action

is not initiated, the apprehended person is necessarily innocent. If a terrorist action is initiated, there is a probability b that the apprehended person is guilty. Thus, $\alpha := ab$ is the probability that the apprehended person has guilty knowledge, and $1 - \alpha$ is the probability that the person has no knowledge. Because we are interested in the logic of the ticking bomb scenario, we set $b = 1$, and therefore $\alpha = a$. We are assuming, in other words, that we are in the case in which the benefits from effective torture would be the largest.

At time 2, not knowing whether or not a terrorist action was initiated, the agency chooses effort $x \geq 0$ to stop a terrorist action by means other than torture. The effort might involve various forms of intelligence gathering and security checks. The cost of x is $c(x)$, where $c' > 0$ and $c'' > 0$. At time 3, if a terrorist action was initiated, the agency's effort stops it with probability $\varphi(x)$, where $\varphi' > 0$ and $\varphi'' \leq 0$. Thus, with probability $\alpha\varphi(x)$, the agency stops a terrorist action, and with probability $1 - \alpha\varphi(x)$, the agency does not stop a terrorist action and infers that either no terrorist action was initiated or one was initiated and the agency's effort did not stop it.

If the agency does not stop a terrorist action, at time 4, it receives evidence ε about whether the apprehended individual initiated a terrorist action. The evidence can be high, ε_H, or low, ε_L. The probability of high evidence is $q_1(x)$ if a terrorist action was initiated and $q_2(x)$ if one was not initiated, where $q_1(x) \geq q_2(x)$ for any $x > 0$, $q_1(0) = q_2(0) = y \in (0, 1)$, $q_1' > 0$, $q_1'' \leq 0$, $q_2' < 0$, and $q_2'' \geq 0$. If the agency increases x, it obtains a more accurate signal, but it still gets a signal even if it chooses $x = 0$.[9]

At time 5, if torture is legal, the agency chooses whether or not to torture the apprehended individual given the available evidence (T or $\neg T$). We initially assume that the agency never

[9] We assume that the same effort x is determining the prevention probability, $\varphi(x)$, and the quality of the evidence about suspects, as represented by $q_1(x)$ and $q_2(x)$. If instead we had $x = (x_1, x_2)$, where x_1 is about prevention while x_2 is about increasing the quality of evidence on suspects, then allowing torture could only reduce prevention effort in the model (see footnote 12 below).

uses torture illegally. If torture is not used and a terrorist action was initiated, the terrorist action succeeds, causing social damages D. If a terrorist action was initiated and torture is used, then with probability θ, torture is effective and the agency extracts the information from the guilty individual and stops the terrorist action, and with probability $1-\theta$, torture is ineffective and the terrorist action succeeds.[10] If torture is used and no terrorist action was initiated, society incurs damages t from torturing an innocent individual.

We assume that the agency internalizes a fraction, $\delta \in [0,1]$, of social damages, D, and does not internalize the costs of torturing innocent people, t, which is part of the agency problem. This assumption motivates possible constraints on the agency's behavior. If the agency is not as concerned as society about protecting safety or individual rights, society may want to prevent the agency from using torture rather than leave the decision to use torture at the agency's discretion.[11] If a terrorist action is initiated but the agency stops it (either through torture or other means) or if a terrorist action is not initiated, then the agency's payoff is $-c(x)$. If a terrorist action is initiated and the agency does not stop the terrorist action, then the agency's payoff is $-\delta D - c(x)$.

Assuming that the agency obeys directives, we compare outcomes when (1) torture is totally banned, regime B; (2) torture is legal only when evidence is high, regime H; and (3) torture is legal whether the evidence is high or low, regime LH. The three regimes have effects on agency behavior, measured by non-torture efforts, and on welfare, measured by public safety, the probability of torturing the innocent, and effort costs.

We perform this comparison both in generality, when the functions $q_1(x)$, $q_2(x)$, $\varphi(x)$, and

[10] As specified, x does not affect the probability θ that torturing the guilty individual yields the information necessary to stop an attack. Making θ an increasing function of x would not affect the main qualitative results of the paper but would add an additional source of complementarity between the agency's use of torture and its other efforts (see footnote 12 below).

[11] We think of this as an extreme version of a reduced form of the difference between agency incentives and society's preferences. For a general analysis of optimal agency discretion when agency objectives are different from those of society, see Shavell (2007).

$c(x)$ satisfy only the derivative/inequality conditions listed above, and for the following plausible parameterization: $x \geq 0$, $q_1(x) = 1 - (1/2)e^{-\gamma_1 x}$, $q_2(x) = (1/2)e^{-\gamma_2 x}$, $\varphi(x) = 1 - e^{-\lambda x}$, and $c(x) = c(e^x - 1)$, where $\gamma_1, \gamma_2, \lambda, c > 0$. The more general analysis yields results phrased in terms of derived quantities and a tradeoff between a decommitment and a complementarity effect. The parametric analysis is complementary, giving additional understanding of when one might expect one effect or the other to be larger.

2.2 Agency Behavior

Let $\mathbb{D}(x) = (1 - \varphi(x))[-Z]$, where $Z = \alpha\delta D$, denote expected damages under regime B. Under regime B, the agency does not use torture whether it has low or high evidence, strategy $(\neg T, \neg T)$. Thus, its optimal action, x_B^*, solves

$$\max_{x \geq 0} EU^B_{Agency}(x|\neg T, \neg T) = \mathbb{D}(x) - c(x). \tag{1}$$

Under regime H, the agency uses tortures only when it has high evidence, strategy $(\neg T, T)$. Thus, its optimal action, x_H^*, solves

$$\max_{x \geq 0} EU^H_{Agency}(x|\neg T, T) = \psi(x)\mathbb{D}(x) - c(x), \tag{2}$$

where $\psi(x) = (1 - q_1(x)\theta)$ is the probability that strong evidence and torture fail. Under regime LH, the agency uses torture whether it has low or high evidence, strategy (T, T). Thus, its optimal action, x_{LH}^*, solves

$$\max_{x \geq 0} EU^{LH}_{Agency}(x|T, T) = (1 - \theta)\mathbb{D}(x) - c(x), \tag{3}$$

where $(1 - \theta)$ is the probability that torture of a guilty person fails. We assume that x_B^*, x_H^*, and x_{LH}^* are strictly positive. We now show that having no ban on torture unambiguously reduces other agency effort and characterize conditions under which a total ban elicits higher effort than a partial ban. (Proofs are in the Appendix.)

Proposition 1 *(a)* $\min(x_B^*, x_H^*) > x_{LH}^*$*, and (b)* $x_B^* > x_H^*$ *iff*

$$DCT(x_B^*) > CP(x_B^*) \text{ or equivalently } DCT(x_H^*) > CP(x_H^*), \tag{4}$$

where $DCT(x^*) = q_1(x^*)\varphi'(x^*)$ *and* $CP(x^*) = q_1'(x^*)(1-\varphi(x^*))$ *for* $x^* \in \{x_B^*, x_H^*\}$.

Proposition 1(b) shows that allowing torture when evidence of terrorist action is high has two effects on the agency's incentives to counter terrorism by means other than torture. First, it lowers these incentives because the agency might come to rely on torture to avert attacks, a decommitment effect. Second, it increases these incentives because other efforts may increase the probability of having strong enough evidence to warrant the use of torture if other efforts fail to stop an attack, a complementarity effect. Formally, the term $DCT(x^*)$ measures the decommitment effect. To the extent that other agency efforts increase the probability of stopping an attack without using torture ($\varphi' > 0$) when they generate enough evidence to use torture (q_1), allowing torture reduces other agency efforts. The term $CP(x^*)$ measures the complementarity effect. To the extent that other agency efforts increase the probability of having enough evidence to use torture ($q_1' > 0$) when they fail to stop an attack without using torture ($1-\varphi$), allowing torture increases other agency efforts.[12]

According to Proposition 1(a), allowing torture even when evidence is low unambiguously reduces the agency's other efforts. It does not increase the chances of having enough evidence to warrant the use of torture if other efforts fail, since the agency always has enough evidence to warrant torture if torture is allowed even when evidence is low. Thus, allowing torture even when evidence is low has no complementarity effect and only a decommitment effect on other efforts and therefore unambiguously reduces other efforts.

[12]If we were to also make the effectiveness of torture, θ, an increasing function of non-torture effort, x, then condition (4) would become $\theta(x^*)q_1(x^*)\varphi'(x^*) > \theta(x^*)q_1'(x^*)(1-\varphi(x^*)) - \theta'(x^*)q_1(x^*)(1-\varphi(x^*))$. The complementarity effect would then have the additional term $-\theta'(x^*)q_1(x^*)(1-\varphi(x^*))$.

If we were to have $x = (x_1, x_2)$, where x_1 is prevention effort and x_2 is effort to improve the quality of evidence on suspects if x_1 fails, so that $\varphi(x_1)$, $q_1(x_2)$, and $c(x_1, x_2) = f(x_1) + g(x_2)$, then there would be no complementarity effect at all and legalizing torture could only reduce prevention effort.

The effects of an increase in the effectiveness of torture (θ) on non-torture efforts under regimes where torture is allowed are analogous to the effects on non-torture efforts of shifting from a regime in which torture is banned to one in which it is allowed.

Proposition 2 *(a)* $\partial x_B^*/\partial\theta = 0$*, (b)* $\partial x_{LH}^*/\partial\theta < 0$*, and (c)* $\partial(x_B^* - x_H^*)/\partial\theta = -\partial x_H^*/\partial\theta > 0$ *iff* $DCT(x_B^*) > CP(x_B^*)$*.*

According to Proposition 2(b), under regime LH, an increase in the effectiveness of torture always reduces non-torture efforts, because it only has a decommitment effect on non-torture efforts under this regime. As part (c) shows, an increase in the effectiveness of torture under regime H has the same decommitment and complementarity effects as those produced by a shift from regime B to regime H, which are identified in Proposition 1(b), and it too reduces non-torture effort if and only if the decommitment effect dominates. This coincidence of conditions is intuitive since, in terms of effects on non-torture effort, using a more effective means of torture is equivalent to using torture in more cases.

What factors make it more likely that the decommitment effect dominates the complementarity effect? Let us explore the role of $Z = \alpha\delta D$ (the attack threat).

Proposition 3 *(a)* $\partial x_B^*/\partial Z > 0$*,* $\partial x_{LH}^*/\partial Z > 0$*, and* $\partial x_H^*/\partial Z > 0$*, (b)* $\partial CP(x_B^*)/\partial Z < 0$ *and* $\partial DCT(x_B^*)/\partial Z \gtrless 0$*, and (c)* $\partial[DCT(x_B^*) - CP(x_B^*)]/\partial Z > 0$ *if* $\varphi'(x)/(1-\varphi(x))$ *is non-decreasing in* x*.*

Proposition 3(a) shows that the agency's non-torture efforts are increasing in Z under all three regimes. According to part (b), the complementarity effect on the agency's non-torture efforts of shifting from regime B to regime H is decreasing in Z. Allowing torture in high-evidence cases increases the agency's incentives for non-torture efforts insofar as these efforts improve the chances of gaining enough evidence to warrant torture if they fail to stop an attack, the complementarity effect. However, if Z is higher, then non-torture

efforts are higher, so these efforts are less likely to fail to stop an attack, thereby lowering the complementarity effect. Moreover, as part (c) shows, if the hazard $\varphi'/(1-\varphi)$ is non-decreasing (i.e., conditional on failing to stop an attack with a given non-torture effort level, the probability of succeeding with a small additional effort grows or remains constant), then the decommitment effect gets higher relative to the complementarity effect as Z increases.

We now employ our parameterization to gain additional understanding of when the decommitment effect is likely to dominate.

Corollary 1 *Suppose* $q_1(x) = 1 - (1/2)e^{-\gamma_1 x}$, $q_2(x) = (1/2)e^{-\gamma_2 x}$, $\varphi(x) = 1 - e^{-\lambda x}$, *and* $c(x) = c(e^x - 1)$, *where* $\gamma_1, \gamma_2, \lambda, c > 0$. *Assume* $(1-\theta)Z\lambda > c$, *which guarantees that* x_B^*, x_H^*, *and* x_{LH}^* *are strictly positive. Then (a)* $x_B^* > x_{LH}^*$ *, and (b)* $x_B^* > x_H^*$ *iff*

$$[\frac{1}{2}(1+\frac{\gamma_1}{\lambda})]\rho^{\gamma_1} < 1, \text{ where } \rho = \left(\frac{c}{Z\lambda}\right)^{\frac{1}{(1+\lambda)}} < 1.$$

The condition for $x_B^* > x_H^*$ in Corollary 1(b) is more likely to be satisfied if λ (the effectiveness of non-torture effort in stopping attacks) is higher, γ_1 (the effectiveness of non-torture effort in turning up strong evidence when the suspect is guilty) is lower, c (the cost of non-torture effort) is lower, or $Z = \alpha\delta D$ is higher. A decrease in the effectiveness of non-torture efforts in turning up strong evidence when the suspect is guilty reduces the complementarity effect of legalizing torture in high-evidence cases, while an increase in the effectiveness of non-torture efforts in stopping attacks increases the decommitment effect, making it more likely that $x_B^* > x_H^*$.

2.3 Welfare Effects

We first analyze two central components of welfare, the probability of being safe from a terrorist action and the probability of torturing the innocent. If the agency does not torture, then the probability of safety and the probability of torturing the innocent are, respectively, $S_B^* = 1 - \alpha(1 - \varphi(x_B^*))$ and $Q_B^* = 0$. If the agency tortures only when it has high evidence,

then the probabilities are $S_H^* = 1 - \alpha(1 - \varphi(x_H^*))\psi(x_H^*)$ and $Q_H^* = (1 - \alpha)q_2(x_H^*)$. If the agency tortures on the basis of any evidence at all, then the probabilities are $S_{LH}^* = 1 - \alpha(1 - \varphi(x_{LH}^*))(1 - \theta)$ and $Q_{LH}^* = (1 - \alpha)$.

The three safety probabilities are strictly increasing in agency effort, and the three probabilities of torturing the innocent are either flat or strictly decreasing in agency effort. Legalizing torture unambiguously reduces the likelihood of torturing the innocent.

Corollary 2 $Q_{LH}^* > Q_H^* > Q_B^* = 0$.

Legalizing torture has a direct effect on security–it stops terrorist attacks when they have been initiated and a torture-susceptible guilty person is available. By altering agency behavior, it also has an indirect effect on security. The overall effect on safety depends on which of these effects dominate.

Corollary 3 *(a)* $S_B^* > S_{LH}^*$ *iff* $[\varphi(x_B^*) - \varphi(x_{LH}^*)] > \theta[1 - \varphi(x_{LH}^*)]$, *(b)* $S_H^* > S_{LH}^*$ *iff* $[\varphi(x_H^*) - \varphi(x_{LH}^*)] > \theta[1 - \varphi(x_{LH}^*) - q_1(x_H^*)(1 - \varphi(x_H^*)]$, *(c)* $S_B^* > S_H^*$ *iff* $[\varphi(x_B^*) - \varphi(x_H^*)] > \theta q_1(x_H^*)[1 - \varphi(x_H^*)]$.

In Corollary 3(c), the left-hand side is positive if the decommitment effect dominates the complementarity effect as identified in Proposition 1(b), so that $x_B^* > x_H^*$. In this case, $S_B^* > S_H^*$ is possible. However, if $x_B^* < x_H^*$, the left hand side of the inequality is always negative, implying that if the complementarity effect dominates, public safety is always higher when torture is legalized than when torture is banned.

Corollary 4 *(a)* $\partial(S_B^* - S_H^*)/\partial\theta = -\partial S_H^*/\partial\theta < 0$ *if* $DCT(x_H^*) < CP(x_H^*)$ *and (b)* $\partial(S_B^* - S_H^*)/\partial Z > 0$ *if* $DCT(x_H^*) > CP(x_H^*)$.

Part (a) of Corollary 4 shows that an increase in the effectiveness of torture makes safety under regime H higher relative to safety under regime B if the complementarity effect

dominates. Part (b) shows that an increase in the attack threat ($Z = \alpha\delta D$) makes it more likely that legalizing torture in high-evidence cases reduces safety if the decommitment effect dominates. From Proposition 3(c), if $\varphi'(x)/(1-\varphi(x))$ is non-decreasing, the decommitment effect is more likely to dominate if the attack threat is higher. Thus, in this case, legalizing torture in high-evidence cases is both more likely to reduce non-torture efforts and more likely to reduce safety if it reduces non-torture efforts as the attack threat increases.

The full welfare levels under each of the regimes are $W_B^* = -(1-S_B^*)D - c(x_B^*)$, $W_H^* = -(1-S_H^*)D - Q_H^* t - c(x_H^*)$, and $W_{LH}^* = -(1-S_{LH}^*)D - Q_{LH}^* t - c(x_{LH}^*)$. The agency's payoffs under each of the regimes are the same as society's payoffs except with D replaced by δD and with $t = 0$. The next corollary compares welfare across regimes.

Corollary 5 *(a) $W_B^* > W_{LH}^*$ iff $[S_B^* - S_{LH}^*]D + Q_{LH}^* t > [c(x_B^*) - c(x_{LH}^*)]$, (b) $W_H^* > W_{LH}^*$ iff $[S_H^* - S_{LH}^*]D + [Q_{LH}^* - Q_H^*]t > [c(x_H^*) - c(x_{LH}^*)]$, and (c) $W_B^* > W_H^*$ iff $[S_B^* - S_H^*]D + Q_H^* t > [c(x_B^*) - c(x_H^*)]$.*

The availability of torture in high-evidence cases may reduce safety if it reduces the agency's non-torture efforts, but the reduction in non-torture efforts still reduces the agency's costs. Nonetheless, as Corollary 5 shows, the availability of torture reduces welfare if the social costs of torturing the innocent, t, are sufficiently high (since t only enters welfare directly, by assumption). However, since the agency does not internalize t by assumption (and internalizes only a fraction δ of damages D), the availability of torture can increase the agency's payoff although it reduces welfare. In such a case, the agency has incentives to disobey directives on torture, which leads us to the enforcement problem.

3. The Enforcement Problem

Our analysis of the agency problem has focused on the choice of non-torture efforts assuming that the agency obeys directives on torture. We now suppose that the agency faces a penalty,

p, if it tortures when not allowed to.[13] We examine and compare what happens when (1) torture is banned by penalties, regime $\mathcal{B}(p,p)$ with penalties $(p_L, p_H) = (p,p)$, and (2) torture is regulated by a warrant system, regime $\mathcal{W}(p,0)$, in which torture does not carry a penalty if evidence is high, i.e., the penalties are $(p_L, p_H) = (p,0)$.

Regime B of the previous section achieves a complete torture ban because the agency obeys directives, while regime $\mathcal{B}(p,p)$ penalizes any torture, and only achieves the ban if the penalty is sufficiently high. The comparison between the regimes H and $\mathcal{W}(p,0)$ is similar. Though we analyze regime $\mathcal{W}(p,0)$ as a system of torture warrants, it is formally identical to nonprosecution of torture in high-evidence cases, e.g., if the "necessity" defense successfully avoids prosecution of torturers in cases where evidence of the suspect's guilt is compelling.

We first gather results that allow us to study agency behavior under the two regimes. With these in place, we then analyze, within the context of the model, the Dershowitz argument for reductions in the frequency of torture arising from a torture warrant system. We then turn to the existence of slippery slopes.

3.1 Agency Behavior

Let $P(\varepsilon_L|x)$ and $P(\varepsilon_H|x)$ be the likelihoods of a low and a high evidence suspect, and let $\beta_L(x)$ and $\beta_H(x)$ be the likelihoods of a low and a high evidence suspect being guilty, respectively. From Bayes' rule, we have

$$\beta_L(x) = \frac{\alpha(1-\varphi(x))(1-q_1(x))}{P(\varepsilon_L|x)} \text{ and } \beta_H(x) = \frac{\alpha(1-\varphi(x))q_1(x)}{P(\varepsilon_H|x)}, \text{ where} \tag{5}$$

$$P(\varepsilon_L|x) = \alpha(1-\varphi(x))(1-q_1(x)) + (1-\alpha)(1-q_2(x)) \text{ and} \tag{6}$$

$$P(\varepsilon_H|x) = \alpha(1-\varphi(x))q_1(x) + (1-\alpha)q_2(x). \tag{7}$$

[13] The penalty p might correspond to the utility cost of the possibility of being prosecuted within the domestic judiciary system or being declared a war criminal by international courts after choosing to use torture. *Ex post* enforcement creates a "liability rule" against torture. For an economic analysis of liability rules in the protection of individual rights, see Kontorovich (2004) and Kaplow and Shavell (1996).

Lemma 1 *For all x, (a) $\beta_L(x) < \beta_H(x)$ and $\beta_L'(x) < 0$, (b) the agency's optimal choice is T at ε_L if $\beta_L(x) > \frac{p_L}{\theta\delta D}$ and T at ε_H if $\beta_H(x) > \frac{p_H}{\theta\delta D}$, and (c) $[P(\varepsilon_L|x) + P(\varepsilon_H|x)]' < 0$, which implies that at least one of the derivatives, $[P(\varepsilon_L|x)]'$ and $[P(\varepsilon_H|x)]'$, is negative.*

$\beta_L(x)$ is decreasing in x because increases in x reduce the probability of a terrorist action eluding the agency's non-torture efforts; reduce the probability of false exculpatory evidence; and increase the probability of valid exculpatory evidence.

The agency's behavior under the different regimes can be understood in two parts: (A) for a given torture policy, finding the agency's optimal effort as a function of the penalty p and calculating the associated payoffs, i.e. the value function; and (B) comparing the value functions to pick the optimal torture policy as a function of p. From the first two parts of Lemma 1, we know that under regime $\mathcal{B}(p,p)$, the agency's optimal torture policy is either (T,T), $(\neg T,T)$, or $(\neg T,\neg T)$, and that under regime $\mathcal{W}(p,0)$, it is either (T,T) or $(\neg T,T)$.

For part (A) of the analysis of regime $\mathcal{B}(p,p)$, let $f_{\mathcal{B}}^{T,T}(p)$, $f_{\mathcal{B}}^{\neg T,T}(p)$, and $f_{\mathcal{B}}^{\neg T,\neg T}(p)$ be the agency's value functions at the optimal effort levels for a given p when the agency's policy is (T,T), $(\neg T,T)$, and $(\neg T,\neg T)$, respectively. Specifically, these are

$$f_{\mathcal{B}}^{T,T}(p) = \max_{x\geq 0}[(1-\theta)\mathbb{D}(x) - c(x)] - p\left\{P(\varepsilon_L|x) + P(\varepsilon_H|x)\right\}, \tag{8}$$

$$f_{\mathcal{B}}^{\neg T,T}(p) = \max_{x\geq 0}[\psi(x)\mathbb{D}(x) - c(x)] - p\left\{0 + P(\varepsilon_H|x)\right\}, \text{ and} \tag{9}$$

$$f_{\mathcal{B}}^{\neg T,\neg T}(p) = \max_{x\geq 0}[\mathbb{D}(x) - c(x)] - p\left\{0 + 0\right\}. \tag{10}$$

For part (B) of the analysis of $\mathcal{B}(p,p)$, agency behavior is the policy part of the solution to

$$V_{\mathcal{B}}(p) = \max\{f_{\mathcal{B}}^{T,T}(p), f_{\mathcal{B}}^{\neg T,T}(p), f_{\mathcal{B}}^{\neg T,\neg T}(p)\}. \tag{11}$$

Similarly, for part (A) of the analysis of regime $\mathcal{W}(p,0)$, we have

$$f_{\mathcal{W}}^{T,T}(p) = \max_{x\geq 0}[(1-\theta)\mathbb{D}(x) - c(x)] - p\left\{P(\varepsilon_L|x) + 0\right\}, \text{ and} \tag{12}$$

$$f_{\mathcal{W}}^{\neg T,T}(p) = \max_{x\geq 0}[\psi(x)\mathbb{D}(x) - c(x)] - p\left\{0 + 0\right\}. \tag{13}$$

and part (B) asks for the policy part of the solution to

$$V_{\mathcal{W}}(p) = \max\{f_{\mathcal{W}}^{T,T}(p), f_{\mathcal{W}}^{\neg T,T}(p)\}. \tag{14}$$

Under both regimes $\mathcal{B}(p,p)$ and $\mathcal{W}(p,0)$, as the agency tortures in fewer circumstances, i.e., as we move down through the three cases in (8), (9), and (10) under regime $\mathcal{B}(p,p)$, and down through the two cases in (12) and (13) under regime $\mathcal{W}(p,0)$, for all fixed values of x, the terms that multiply the penalties p decrease. This suggests that the value functions in the two regimes have a "single-crossing from above in p" property. This turns out to be true but is somewhat more subtle because we are evaluating the penalty terms at different values of x. The following lemma addresses this issue.

Lemma 2 *The value functions for part (A) of the analysis of regimes $\mathcal{B}(p,p)$ and $\mathcal{W}(p,0)$ have the following properties: the value function for any policy that tortures is strictly decreasing in p; convex in p; and crosses the value function for a policy that tortures in fewer circumstances exactly once, from above, as p increases.*

Define $\underline{p_{\mathcal{B}}}$, $\overline{p_{\mathcal{B}}}$, and $p_{\mathcal{W}}$ to be the penalty levels at which we have the crossings,

$$f_{\mathcal{B}}^{T,T}(\underline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,T}(\underline{p_{\mathcal{B}}}),\ f_{\mathcal{B}}^{\neg T,T}(\overline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,\neg T}(\overline{p_{\mathcal{B}}}), \text{ and} \tag{15}$$

$$f_{\mathcal{W}}^{T,T}(p_{\mathcal{W}}) = f_{\mathcal{W}}^{\neg T,T}(p_{\mathcal{W}}). \tag{16}$$

3.2 The Dershowitz Argument

As noted above, the core of the Dershowitz (2002) argument is that torture happens although it is illegal and that an enforced system of judicial warrants could bring this under control, resulting in less torture. In our model, Dershowitz's case corresponds to regime $\mathcal{B}(p^\circ, p^\circ)$ for any p° at which the agency's torture policy is $(\neg T, T)$. The appropriate comparison is the change of regime to regime $\mathcal{W}(p^\circ, 0)$. From Lemma 1(b), the agency's choice of torture

policy in regime $\mathcal{W}(p^\circ, 0)$ is either $(\neg T, T)$, the case we analyze now, or (T, T), which we analyze in Proposition 6 as part of our discussion of slippery slopes.

The patterns of the crossings of the value functions determine when $(\neg T, T)$ is simultaneously optimal under regime $\mathcal{B}(p, p)$ and regime $\mathcal{W}(p, 0)$.

Proposition 4 *The agency's optimal torture policy is $(\neg T, T)$ under both regimes $\mathcal{B}(p^\circ, p^\circ)$ and $\mathcal{W}(p^\circ, 0)$ if $\underline{p_{\mathcal{B}}} < \overline{p_{\mathcal{B}}}$, $p_{\mathcal{W}} < \overline{p_{\mathcal{B}}}$, and $p^\circ \in (\max\{p_{\mathcal{W}}, \underline{p_{\mathcal{B}}}\}, \overline{p_{\mathcal{B}}})$.*

We now compare agency effort, safety, and frequency of torture of the innocent under regimes $\mathcal{B}(p, p)$ and $\mathcal{W}(p, 0)$ when the optimal policy is $(\neg T, T)$ under both regimes.

Proposition 5 *Suppose the agency's optimal torture policy is $(\neg T, T)$ under both regimes $\mathcal{B}(p^\circ, p^\circ)$ and $\mathcal{W}(p^\circ, 0)$. (a) If $[P(\varepsilon_H|x)]' < 0$ for all x, regime $\mathcal{W}(p^\circ, 0)$ has lower effort and thus lower safety and more frequent torture of the innocent. (b) If $[P(\varepsilon_H|x)]' > 0$ for all x, regime $\mathcal{W}(p^\circ, 0)$ has higher effort and safety and less frequent torture of the innocent.*

The effects of moving from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ depend on the sign of $[P(\varepsilon_H|x)]'$:

$$[P(\varepsilon_H|x)]' = \alpha[q_1'(x)(1 - \varphi(x)) - \varphi'(x)q_1(x)] + (1 - \alpha)q_2'(x). \tag{17}$$

Thus, the sign of $[P(\varepsilon_H|x)]'$ depends on the relative sizes of the three terms, $DCT(x) = -\alpha\varphi'(x)q_1(x) < 0$, $CP(x) = \alpha q_1'(x)(1 - \varphi(x)) > 0$, and $DCP(x) = (1 - \alpha)q_2'(x) < 0$, which correspond to the following three effects. Decommitment effect (DCT): moving from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ tends to reduce agency effort x because it eliminates the agency's cost of using torture if the individual is guilty and the agency's evidence turns out to be high, which reduces the agency's commitment not to use torture. Complementarity effect (CP): moving from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ tends to increase x because increasing x increases the probability that the agency has high evidence if the individual is guilty, in which case the agency can use torture without punishment. Decomplementarity effect (DCP): moving

from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ tends to reduce x because reducing x increases the probability that the agency has high evidence if the individual is innocent, in which case the agency can escape punishment for using torture on an innocent individual.

If the complementarity effect dominates both the decommitment and decomplementarity effects, then moving from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ increases agency effort x, and thereby increases safety and reduces the probability of torturing the innocent. However, if the decommitment and decomplementarity effects together dominate the complementarity effect, then moving from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ reduces agency effort x, and thereby reduces safety and increases the probability of torturing the innocent.

Corollary 6 *Suppose* $q_1(x) = 1 - (1/2)e^{-\gamma_1 x}$, $q_2(x) = (1/2)e^{-\gamma_2 x}$, $\varphi(x) = 1 - e^{-\lambda x}$, *and* $c(x) = c(e^x - 1)$, *where* $\gamma_1, \gamma_2, \lambda, c > 0$. *Then* $[P(\varepsilon_H|x)]' < 0$ *if (a)* $\lambda > \gamma_2$ *and (b)* $\alpha\gamma_1 < \alpha\lambda + (1-\alpha)\gamma_2$.

Conditions (a) and (b) are both satisfied if λ is sufficiently high. Thus, for our parameterization, moving from $\mathcal{B}$ to $\mathcal{W}$ is likely to reduce agency non-torture efforts and security if non-torture efforts are relatively effective at stopping attacks.

3.3 Slippery Slopes

We have a slippery slope if legalizing torture in dire circumstances increases the set of circumstances in which torture occurs. This happens if either $(\neg T, T)$ or $(\neg T, \neg T)$ are optimal for the agency under $\mathcal{B}(p, p)$ while (T, T) is optimal under the torture warrant system $\mathcal{W}(p, 0)$. In either case, switching to a torture warrant system lowers agency effort and increases the frequency of torture of the innocent.

Proposition 6 *If* p° *is such that either (a) the agency's optimal torture policy is* $(\neg T, \neg T)$ *under regime* $\mathcal{B}(p^\circ, p^\circ)$ *and is* (T, T) *under regime* $\mathcal{W}(p^\circ, 0)$, *or (b) the agency's optimal torture policy is* $(\neg T, T)$ *under regime* $\mathcal{B}(p^\circ, p^\circ)$ *and is* (T, T) *under regime* $\mathcal{W}(p^\circ, 0)$, *then switching from regime* $\mathcal{B}(p^\circ, p^\circ)$ *to regime* $\mathcal{W}(p^\circ, 0)$ *reduces agency effort and increases the frequency of torture of the innocent.*

If the torture policy changes in either of the ways indicated, we have direct evidence of lower agency effort—from Lemma 1, the agency only tortures at ε_L if $\beta_L(x)$ is sufficiently high, and higher values of β_L can only arise from lower values of x. More intuitively, the slippery slopes in Proposition 6(a) and (b) both arise from lower effort reducing the quality of exculpatory evidence. Moving from regime $\mathcal{B}$ to regime $\mathcal{W}$ eliminates the agency's penalty for using torture in high-evidence cases. This reduces the agency's commitment to non-torture efforts, whether or not the agency was already using torture in high-evidence cases under $\mathcal{B}$. A reduction in non-torture efforts reduces the quality of the agency's evidence, which increases the agency's incentives to adopt a strategy that uses torture even when evidence is low. Adoption of such a strategy further reduces agency efforts, further reinforcing the agency's incentives to use torture even when evidence is low.

Whether or not moving down either type of slippery slope translates to lower safety depends, as in Corollary 3, on the effectiveness of torture. If torture is sufficiently ineffective relative to other efforts, then lowering other efforts and torturing more reduces safety. Whether or not each type of slippery slope arises depends on the relative locations of the crossing points defined in (15) and (16).

Proposition 7 *Suppose $\underline{p_{\mathcal{B}}} < \overline{p_{\mathcal{B}}}$. (a) The agency's optimal torture policy is uniquely $(\neg T, \neg T)$ under regime $\mathcal{B}(p^\circ, p^\circ)$ and is uniquely (T, T) under regime $\mathcal{W}(p^\circ, 0)$ if and only if $\overline{p_{\mathcal{B}}} < p^\circ < p_{\mathcal{W}}$. (b) The agency's optimal torture policy is uniquely $(\neg T, T)$ under regime $\mathcal{B}(p^\circ, p^\circ)$ and is uniquely (T, T) under regime $\mathcal{W}(p^\circ, 0)$ if and only if $\underline{p_{\mathcal{B}}} < p^\circ \min\{\overline{p_{\mathcal{B}}}, p_{\mathcal{W}}\}$).*

The crossing point conditions in Proposition 7(a) and (b) are more easily satisfied if the value functions for policies involving torture in regime $\mathcal{B}$ are steeper functions of p while the value function for the policy involving torture in regime $\mathcal{W}$ is shallower. The value functions for $\mathcal{B}$ are relatively steeper if the level of $P(\varepsilon_H|x)$ is higher, because this term is the difference between the penalties across $\mathcal{B}$ and $\mathcal{W}$. If the probability of having high evidence suspects is

higher, the agency is more tempted to rely on torture as a counterterrorism tool, and moving to $\mathcal{W}$ is more likely to lower agency effort and increase torture of the innocent. Moreover, if $P(\varepsilon_H|x)$ does not vary much with x, then the value functions in $\mathcal{B}$ are less convex, meaning that they stay steeper for a larger range of penalties. If the probability of having a high evidence suspect is less sensitive to effort, then there is less loss to lowering effort if torture is used more, which also makes a slippery slope more likely to arise.

4. Summary and Future Work

We developed a model of counterterrorism to analyze the effects of allowing the government to use torture when evidence of terrorist involvement is strong. We first examined the case in which the agency tasked with counterterrorism places a different weight on torture than society does, but in which it follows any directives. In this case, we showed that allowing the agency to use torture in strong-evidence cases may reduce its efforts to stop terrorism by means other than torture. This effect blunts any gain to safety that may arise through torture, and the net effect may be a reduction in security.

We then extended our analysis to encompass the possibility that there is an enforcement problem and the agency is willing to disobey torture directives at the risk of legal sanction. This extension allowed us to examine conditions under which the Dershowitz argument that a system of torture warrants could reduce torture holds or fails, showing that its validity depends on the agency's ineffectiveness at stopping attacks. It also brought to light a slippery slope that works through the endogeneity of the quality of information. Allowing torture in strong-evidence cases may reduce the agency's non-torture efforts. The resulting agency deskilling may then reduce the quality of exculpatory evidence, which may lead to torture even in weak-evidence cases. The main arguments we developed have a simple

outline: loosening constraints on torture may induce changes in agency behavior that may compromise security and reduce the quality of the agency's evidence to such an extent that it motivates the use of torture even in the face of potentially exculpatory evidence.

In future work, it might be interesting to endogenize terrorism in our model and examine the effects of legalizing torture on the probability of attack. If legalizing torture has a sufficiently large decommitment effect, it might indirectly increase the probability of attack by reducing the agency's preventive effort and thereby increasing the probability that an attack would succeed. On the other hand, torture could directly reduce the attack probability if it is effective as a punishment or means of intimidation. However, if illegitimate regimes are more prone to use torture (e.g., for intimidation) than legitimate ones, legalizing torture might also signal that a regime is illegitimate, which may increase individuals' benefits of attacking it. Dreher, Gassebner, and Siemers (2010) find that terrorist attacks are positively associated with human rights violations (including torture) across regions. By engaging in torture, a government risks pooling with illegitimate regimes and inciting a terrorist backlash. In addition, the use of torture may affect the structure of terrorist organizations. Torture may be used to elicit information about other terrorist suspects or members of the terrorist network, which may induce the terrorist organization to become more decentralized.

A Mathematical Appendix

Proof of Proposition 1. As we assume interior solutions and smooth and concave objective functions, analyses of the first order conditions (FOCs) are sufficient. For (b), the FOC characterizing x_B^* is (1′) $\varphi'(x_B^*)Z = c'(x_B^*)$ and the FOC characterizing x_H^* is (2′) $Z[\varphi'(x_H^*)(1-q_1(x_H^*)\theta)+q_1'(x_H^*)\theta(1-\varphi(x_H^*))] = c'(x_H^*)$. Since $c''(x) > 0$, we have $x_B^* > x_H^*$ iff $Z[\varphi'(x_B^*)(1-q_1(x_B^*)\theta)+q_1'(x_B^*)\theta(1-\varphi(x_B^*))] < c'(x_B^*) = \varphi'(x_B^*)Z$. Similarly, we have $x_B^* >$

x_H^* iff $Z[\varphi'(x_H^*)(1-q_1(x_H^*)\theta)+q_1'(x_H^*)\theta(1-\varphi(x_H^*))]=c'(x_H^*)<\varphi'(x_H^*)Z$. Thus, $x_B^*>x_H^*$ iff $q_1(x_B^*)\varphi'(x_B^*)>q_1'(x_B^*)(1-\varphi(x_B^*))$ or equivalently $q_1(x_H^*)\varphi'(x_H^*)>q_1'(x_H^*)(1-\varphi(x_H^*))$. For (a), the FOC characterizing x_{LH}^* is (3′) $(1-\theta)[\varphi'(x_{LH}^*)Z]=c'(x_{LH}^*)$. From $(1-\theta)<1$, it follows that $x_B^*>x_{LH}^*$. Moreover, since $c''(x)>0$, (2′) and (3′) deliver $x_H^*>x_{LH}^*$ iff $q_1'(x_H^*)(\varphi(x_H^*)-1)<(1-q_1(x_H^*))\varphi'(x_H^*)$, which always holds. ∎

Proof of Proposition 2. Let SOC_B, SOC_H, and SOC_{LH} be the second order conditions for x_B^*, x_H^*, and x_{LH}^*, respectively. Then, $SOC_B<0$, $SOC_H<0$, and $SOC_{LH}<0$. For (a), the FOC for x_B^* is $\varphi'(x_B^*)Z-c'(x_B^*)=0$. Thus, $\partial x_B^*/\partial\theta=0$. For (b), the FOC for x_{LH}^* is $Z(1-\theta)\varphi'(x_{LH}^*)-c'(x_{LH}^*)=0$. Thus, $\partial x_{LH}^*/\partial\theta=(1/SOC_{LH})[\varphi'(x_{LH}^*)Z]<0$. For (c), the FOC for x_H^* is $Z[\varphi'(x_H^*)(1-q_1(x_H^*)\theta)+q_1'(x_H^*)\theta(1-\varphi(x_H^*))]-c'(x_H^*)=0$. Thus, $\partial x_H^*/\partial\theta=-(1/SOC_H)[-q_1(x_H^*)\varphi'(x_H^*)+q_1'(x_H^*)(1-\varphi(x_H^*))]<0$ iff $DCT(x_B^*)>CP(x_B^*)$. ∎

Proof of Proposition 3. For (a), we have $\partial x_B^*/\partial Z=(1/SOC_B)[\varphi'(x_B^*)]>0$, $\partial x_{LH}^*/\partial Z=(1/SOC_{LH})[(1-\theta)\varphi'(x_{LH}^*)]>0$, and $\partial x_H^*/\partial Z=-(1/SOC_H)[c'(x_H^*)/Z^2]>0$. For (b), we have $\partial CP(x_B^*)/\partial Z=q_1''(x_B^*)(1-\varphi(x_B^*))-q_1'(x_B^*)\varphi'(x_B^*)<0$ and $\partial DCT(x_B^*)/\partial Z=q_1'(x_B^*)\varphi'(x_B^*)+q_1(x_B^*)\varphi''(x_B^*)\gtrless 0$. For (c), $DCT(x_B^*)-CP(x_B^*)>0$ iff $\varphi'(x_B^*)/(1-\varphi(x_B^*))>q_1'(x_B^*)/q_1(x_B^*)$. Since $\partial[q_1'(x)/q_1(x)]/\partial x<0$ and $\partial x_B^*/\partial Z>0$, $\{\partial[DCT(x_B^*)-CP(x_B^*)]/\partial x_B^*\}(\partial x_B^*/\partial Z)>0$ if the hazard rate $\varphi'(x)/(1-\varphi(x)$ is non-decreasing in x. ∎

Proof of Corollary 1. Since x_{LH}^* is minimal, strictly positive optima are guaranteed by $(1-\theta)\ \varphi'(0)Z>c'(0)$, i.e. $(1-\theta)Z\lambda>c$. Solving $\varphi'(x_B^*)Z=c'(x_B)$, $(1-\theta)\varphi'(x_{LH}^*)Z=c'(x_{LH})$, and $Z[\varphi'(x_H^*)(1-q_1(x_H^*)\theta)+q_1'(x_H^*)\theta(1-\varphi(x_H^*))]=c'(x_H^*)$, we find $x_B^*=(1/(1+\lambda))\ln(Z\lambda/c)$, $x_{LH}^*=(1/(1+\lambda))\ln((1-\theta)(Z\lambda/c))$, and x_H^* solves $(Z\lambda/c)\left[1-\theta+(\theta/2)[1+(\gamma_1/\lambda)]e^{-\gamma_1 x}\right]=e^{(1+\lambda)x}$. Clearly, $x_B^*>x_{LH}^*$. Moreover, from Proposition 1(b), $x_B^*>x_H^*$ iff $q_1(x_B^*)\varphi'(x_B^*)>q_1'(x_B^*)(1-\varphi(x_B^*))$. Plugging in $x_B^*=(1/(1+\lambda))\ln(Z\lambda/c)$, we get that $x_B^*>x_H^*$ iff $[(1+\gamma_1/\lambda)/2]\rho^{\gamma_1}<1$, where $\rho=(c/Z\lambda)^{1/(1+\lambda)}<1$. ∎

Proof of Corollary 4. For (a) $\partial(S_B^* - S_H^*)/\partial\theta = -\partial S_H^*/\partial\theta = -(\partial x_H^*/\partial\theta)[(1 - q_1(x_H^*)\theta)\varphi'(x_H^*) + q_1'(x_H^*)(1 - \varphi(x_H^*))] < 0$ if $\partial x_H^*/\partial\theta > 0$. From Proposition 2, we have $\partial x_H^*/\partial\theta > 0$ if $DCT(x_H^*) < CP(x_H^*)$. Thus, $\partial(S_B^* - S_H^*)/\partial\theta$ if $DCT(x_H^*) < CP(x_H^*)$. For (b), $\partial(S_B^* - S_H^*)/\partial Z = \{\partial[\varphi(x_B^*) - \varphi(x_H^*)]/\partial(x_B^* - x_H^*)\}(\partial(x_B^* - x_H^*)/\partial Z) + \theta[q_1(x_H^*)\varphi'(x_H^*) - q_1'(x_H^*)(1 - \varphi(x_H^*))](\partial x_H^*/\partial Z)$. By construction, $\partial[\varphi(x_B^*) - \varphi(x_H^*)]/\partial(x_B^* - x_H^*) > 0$. From Proposition 3, $\partial x_H^*/\partial Z > 0$ and $\partial(x_B^* - x_H^*)/\partial Z > 0$. Thus, if $q_1(x_H^*)\varphi'(x_H^*) - q_1'(x_H^*)(1 - \varphi(x_H^*)) > 0$ or $DCT(x_H^*) > CP(x_H^*)$, then $\partial(S_B^* - S_H^*)/\partial Z > 0$. ∎

Proof of Lemma 1. For the first part of (a), note that $\beta_L(x) < \beta_H(x)$ because $q_1(x) > q_2(x)$. For the second part, note that

$$\beta_L(x) = \frac{\alpha(1 - \varphi(x))(1 - q_1(x))}{\alpha(1 - \varphi(x))(1 - q_1(x)) + (1 - \alpha)(1 - q_2(x))} = \frac{a(x)}{a(x) + b(x)}, \tag{18}$$

where $a(\cdot)$ is decreasing in x and $b(x)$ is increasing. (b) is a direct implication of utility maximization. For (c), note that $P(\varepsilon_L|x) + P(\varepsilon_H|x) = \alpha(1 - \varphi(x)) + (1 - \alpha)$, so $[P(\varepsilon_L|x) + P(\varepsilon_H|x)]' < 0$ since $\varphi'(x) > 0$. ∎

Proof of Lemma 2. By the envelope theorem, the value functions in (8), (9), and (12) are strictly decreasing in the penalty p. For convexity, note that all three problems are of the form $f(p) = \max_{x \geq 0} n(x) - pm(x)$ and that, $x^*(p)$, the optimal effort x for a given value of p has the property that $\frac{d}{dp}x^*(p)$ has the opposite sign of $m'(x^*(p))$. Now, $f(p) = n(x^*(p)) - pm(x^*(p))$ where $x^*(p)$ satisfies $[n'(x^*(p)) - pm'(x^*(p))] \equiv 0$. Therefore, $f'(p) = [n'(x^*) - pm'(x^*)]\frac{d}{dp}x^*(p) - m(x^*(p))$ and $f''(p) = -m'(x^*)\frac{d}{dp}x^*(p)$. Since the two terms in this product have the opposite sign, we have $f''(p) \geq 0$. We now show that the value functions in the two regimes have the "single-crossing from above in p" property. We treat the simpler case, regime $\mathcal{W}$, first. Since (T, T) is the agency's optimal policy at $p = 0$, $f_{\mathcal{W}}^{T,T}(0) > f_{\mathcal{W}}^{\neg T,T}(0)$. By the envelope theorem, under regime $\mathcal{W}$ with penalties $(p, 0)$, $\frac{d}{dp}f_{\mathcal{W}}^{T,T}(p) = -P(\varepsilon_L|x) < 0$, while $\frac{d}{dp}f_{\mathcal{W}}^{\neg T,T}(p) = 0$. For the same reasons, under regime

$\mathcal{B}(p,p)$, the crossings of $f_{\mathcal{B}}^{\neg T,\neg T}(p)$ happen from above. All that is left to consider is the crossing of $f_{\mathcal{B}}^{T,T}(p)$ and $f_{\mathcal{B}}^{\neg T,T}(p)$. In order to show that $f_{\mathcal{B}}^{T,T}(p)$ crosses $f_{\mathcal{B}}^{\neg T,T}(p)$ at most once from above as p increases, it is sufficient to show that for at any $p^{\ddagger}$ where $f_{\mathcal{B}}^{T,T}(p^{\ddagger}) = f_{\mathcal{B}}^{\neg T,T}(p^{\ddagger})$, $f_{\mathcal{B}}^{T,T}(\cdot)$ is steeper than $f_{\mathcal{B}}^{\neg T,T}(\cdot)$. By the envelope theorem again, we need to show that $[P(\varepsilon_L|x_{T,T}^*(p^{\ddagger})) + P(\varepsilon_H|x_{T,T}^*(p^{\ddagger}))] > P(\varepsilon_H|x_{\neg T,T}^*(p^{\ddagger}))$, where $x_{T,T}^* = x_{T,T}^*(p^{\ddagger})$ and $x_{\neg T,T}^* = x_{\neg T,T}^*(p^{\ddagger})$ are the agency's corresponding optimal effort levels under regime $\mathcal{B}(p^{\ddagger},p^{\ddagger})$ when following the torture policies (T,T) and $(\neg T,T)$. Now, at $p^{\ddagger}$, both torture policies $(\neg T,T)$ and (T,T) are optimal, which implies that $\beta_L(x_{\neg T,T}^*) \leq p^{\ddagger}/\theta\delta D \leq \beta_L(x_{T,T}^*)$. Since $\beta_L'(x) < 0$, this implies that $x_{\neg T,T}^* \geq x_{T,T}^*$. Then,

$$
\begin{aligned}
&P(\varepsilon_L|x_{TT}^*) + P(\varepsilon_H|x_{TT}^*) - P(\varepsilon_H|x_{\neg TT}^*) \qquad (19)\\
&= \alpha(1-\varphi(x_{TT}^*)) + (1-\alpha) - \alpha(1-\varphi(x_{\neg TT}^*))q_1(x_{\neg TT}^*) - (1-\alpha)q_2(x_{\neg TT}^*)\\
&= (>0) + \alpha\underbrace{\{1-\varphi(x_{\neg TT}^*)\}\{1-q_1(x_{\neg TT}^*)\}}_{>0} + \underbrace{\alpha\{\varphi(x_{\neg TT}^*) - \varphi(x_{TT}^*)\}}_{\geq 0} > 0,
\end{aligned}
$$

where the weak inequality comes from $x_{\neg TT}^* \geq x_{TT}^*$ and $\varphi'(x) > 0$. ∎

Proof of Proposition 4. From Lemma **2**, $f_{\mathcal{B}}^{T,T}(p)$ crosses $f_{\mathcal{B}}^{\neg T,T}(p)$ and $f_{\mathcal{B}}^{\neg T,T}(p)$ crosses $f_{\mathcal{B}}^{\neg T,\neg T}(p)$ once from above as p increases. Let $\overline{p_{\mathcal{B}}}$ and $\underline{p_{\mathcal{B}}}$ be the points at which $f_{\mathcal{B}}^{T,T}(\overline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,T}(\overline{p_{\mathcal{B}}})$ and $f_{\mathcal{B}}^{\neg T,T}(\underline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,\neg T}(\underline{p_{\mathcal{B}}})$, respectively. If $\underline{p_{\mathcal{B}}} < \overline{p_{\mathcal{B}}}$, under $\mathcal{B}(p^\circ,p^\circ)$, for $p^\circ \in [0,\underline{p_{\mathcal{B}}})$, $V_{\mathcal{B}}(p^\circ) = f_{\mathcal{B}}^{T,T}(p^\circ)$, for $p^\circ \in (\underline{p_{\mathcal{B}}},\overline{p_{\mathcal{B}}})$, $V_{\mathcal{B}}(p^\circ) = f_{\mathcal{B}}^{\neg T,T}(p^\circ)$, and for $p^\circ > \overline{p_{\mathcal{B}}}$, $V_{\mathcal{B}}(p^\circ) = f_{\mathcal{B}}^{\neg T,\neg T}(p^\circ)$. The agency chooses $(\neg T,T)$ under $\mathcal{B}$ if $p^\circ \in (\underline{p_{\mathcal{B}}},\overline{p_{\mathcal{B}}})$. Similarly, from Lemma **2**, $f_{\mathcal{W}}^{T,T}(p)$ crosses $f_{\mathcal{W}}^{\neg T,T}(p)$ once from above as p increases. Let $p_{\mathcal{W}}$ be the point at which $f_{\mathcal{W}}^{T,T}(p_{\mathcal{W}}) = f_{\mathcal{W}}^{\neg T,T}(p_{\mathcal{W}})$. Under $\mathcal{W}(p^\circ,0)$, for $p^\circ \in [0,p_{\mathcal{W}})$, $V_{\mathcal{W}}(p^\circ) = f_{\mathcal{W}}^{T,T}(p^\circ)$, and for $p^\circ > p_{\mathcal{W}}$, $V_{\mathcal{W}}(p^\circ) = f_{\mathcal{W}}^{\neg T,T}(p^\circ)$. The agency chooses $(\neg T,T)$ under $\mathcal{W}$ if $p^\circ > p_{\mathcal{W}}$. Thus, if $p_{\mathcal{W}} < \overline{p_{\mathcal{B}}}$, then for $p^\circ \in (\max\{p_{\mathcal{W}},\underline{p_{\mathcal{B}}}\},\overline{p_{\mathcal{B}}})$, the agency chooses $(\neg T,T)$ under both $\mathcal{B}$ and $\mathcal{W}$. ∎

Proof of Proposition 5. For $t \in \{0,1\}$ and $x \geq 0$, define $h(x,t) = [\psi(x)\mathbb{D}(x) - c(x)] -$

$p^\circ \{0 + t \cdot P(\varepsilon_H|x)\}$, so that $\max_{x\geq 0} h(x,1)$ is the agency's optimal effort problem in $\mathcal{B}(p^\circ, p^\circ)$ if following the policy $(\neg T, T)$, and $\max_{x\geq 0} h(x,0)$ is the agency's optimal effort problem in $\mathcal{W}(p^\circ, 0)$ if following the policy $(\neg T, T)$. Simple increasing differences comparative statics (e.g., Corbae, Stinchcombe, and Zeeman, 2009, §2.8.b) show that $x^*(1) \geq x^*(0)$ if for all $x' > x$, $h(x',1) - h(x,1) > h(x',0) - h(x,0)$, and $x^*(1) \leq x^*(0)$ if for all $x' > x$, $h(x',1) - h(x,1) < h(x',0) - h(x,0)$. Applied to $h(x,t)$ defined above, $h(x',1) - h(x,1) > h(x',0) - h(x,0)$ iff $-p^\circ \cdot \{P(\varepsilon_H|x') - P(\varepsilon_H|x)\} > 0$. Thus, $[P(\varepsilon_H|x)]' < 0$ for all x implies that $\mathcal{B}(p^\circ, p^\circ)$ has higher optimal effort than $\mathcal{W}(p^\circ, 0)$, part (a) of the Proposition, and $[P(\varepsilon_H|x)]' > 0$ for all x implies that $\mathcal{B}(p^\circ, p^\circ)$ has lower optimal effort than $\mathcal{W}(p^\circ, 0)$, part (b) of the Proposition. ∎

Proof of Proposition 6. In both (a) and (b), the switch from $\mathcal{B}(p^\circ, p^\circ)$ to $\mathcal{W}(p^\circ, 0)$ involves changing the policy from not torturing when evidence is low to torturing when evidence is low. By Lemma 1(a), $\beta_L'(x) < 0$. Therefore, by Lemma 1(b), the change in torture policy indicates a decrease in effort. The argument for the increasing frequency of torture of the innocent follows, as above, from $0 < q_2(x) < 1$. ∎

Proof of Proposition 7. From Lemma **2**, $f_{\mathcal{B}}^{T,T}(p)$ crosses $f_{\mathcal{B}}^{\neg T,T}(p)$ and $f_{\mathcal{B}}^{\neg T,T}(p)$ crosses $f_{\mathcal{B}}^{\neg T,\neg T}(p)$ once from above as p increases. Let $\overline{p_{\mathcal{B}}}$ and $\underline{p_{\mathcal{B}}}$ be the points at which $f_{\mathcal{B}}^{T,T}(\overline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,T}(\overline{p_{\mathcal{B}}})$ and $f_{\mathcal{B}}^{\neg T,T}(\underline{p_{\mathcal{B}}}) = f_{\mathcal{B}}^{\neg T,\neg T}(\underline{p_{\mathcal{B}}})$, respectively. Under $\mathcal{B}(p^\circ, p^\circ)$, if $\underline{p_{\mathcal{B}}} < \overline{p_{\mathcal{B}}}$, then for $p^\circ \in (\underline{p_{\mathcal{B}}}, \overline{p_{\mathcal{B}}})$, $V_{\mathcal{B}}(p^\circ) = f_{\mathcal{B}}^{\neg T,T}(p^\circ)$, and thus, the agency chooses $(\neg T, T)$, and for $p > \overline{p_{\mathcal{B}}}$, $V_{\mathcal{B}}(p^\circ) = f_{\mathcal{B}}^{\neg T,\neg T}(p^\circ)$, and thus, the agency chooses $(\neg T, \neg T)$. Similarly, from Lemma **2**, $f_{\mathcal{W}}^{T,T}(p)$ crosses $f_{\mathcal{W}}^{\neg T,T}(p)$ once from above as p increases. Let $p_{\mathcal{W}}$ be the point at which $f_{\mathcal{W}}^{T,T}(p_{\mathcal{W}}) = f_{\mathcal{W}}^{\neg T,T}(p_{\mathcal{W}})$. Under $\mathcal{W}(p^\circ, 0)$, for $p^\circ \in [0, p_{\mathcal{W}})$, $V_{\mathcal{W}}(p^\circ) = f_{\mathcal{W}}^{T,T}(p^\circ)$, and thus, the agency chooses (T, T). Case (a): if $\overline{p_{\mathcal{B}}} < p_{\mathcal{W}}$, the agency chooses $(\neg T, \neg T)$ under regime $\mathcal{B}$ and (T, T) under regime $\mathcal{W}$ if $p^\circ \in (\overline{p_{\mathcal{B}}}, p_{\mathcal{W}})$. Case (b): if $\underline{p_{\mathcal{B}}} < p_{\mathcal{W}}$, the agency chooses $(\neg T, T)$ under $\mathcal{B}$ and (T, T) under $\mathcal{W}$ if $p^\circ \in (\underline{p_{\mathcal{B}}}, \min\{\overline{p_{\mathcal{B}}}, p_{\mathcal{W}}\})$. ∎

REFERENCES

Bagaric, Mirko and Clarke, Julie (2006). *Torture: When the Unthinkable is Morally Permissible*, State University of New York Press: New York.

Bagnoli, Mark and Bergstrom, Ted (2005). "Log-concave Probability and Its Applications," *Economic Theory* 26, 445-469.

Baliga, Sandeep and Ely, Jeffrey C. (2010). "Torture," Northwestern University Working Paper.

Berman, Eli and Laitin, David D. (2008). "Religion, Terrorism and Public Goods: Testing the Club Model," *Journal of Public Economics* 92, 1942-1967.

Chen, Kong-Ping, Tsai, Tsung-Sheng, and Leung, Angela (2009). "Judicial Torture as a Screening Device," Academia Sinica Working Paper.

Chen, Kong-Ping, Chou, Chien-Fu, and Tsai, Tsung-Sheng (2009). "Judicial Torture as a War of Attrition," Academia Sinica Working Paper.

Corbae, Dean, Stinchcombe, Maxwell B., and Zeeman, Juraj (2009). *An Introduction to Mathematical Analysis for Economic Theory and Econometrics*, Princeton University Press: New Jersey.

Dershowitz, Alan M. (2002). *Why Terrorism Works*, Yale University Press: New Haven.

Dershowitz, Alan M. (2003). "The Torture Warrant: A Response to Professor Strauss," *New York Law School Law Review* 48, 275-294.

Dreher, Axel, Gassebner, Martin, and Siemers, Lars-H. R. (2010). "Does Terror Threaten Human Rights? Evidence From Panel Data," *Journal of Law and Economics* 53, 65-93.

Eggen, Dan (2007). "Torture Stance Raises Doubts on Mukasey," *Washington Post* October 27.

Enders, Walter and Sandler, Todd (2004). "An Economic Perspective on Transnational Terrorism," *European Journal of Political Economy* 20, 301-316.

Enders, Walter and Sandler, Todd (2005). "Transnational Terrorism 1968-2000: Thresholds, Persistence, and Forecasts," *Southern Economic Journal* 71, 467-483.

Fay, Major General George R. (2004). "Fay Report: Investigation of Intelligence Activities At Abu Ghraib: Executive Summary," Available at Website http://fl1.findlaw.com/news.findlaw.com/hdocs/docs/dod/fay82504rpt.pdf.

Garoupa, Nuno, Klick, Jonathan, and Parisi, Francesco (2006). "A Law and Economics Perspective on Terrorism," *Public Choice* 128, 147-168.

Imseis, Ardi (2001). "'Moderate' Torture On Trial: The Israeli Supreme Court Judgement Concerning the Legality of the General Security Service Interrogation Methods," *International Journal of Human Rights* 5, 71-96.

Kaplow, Louis and Shavell, Steven (1996). "Property Rules versus Liability Rules: An Economic Analysis," *Harvard Law Review* 109, 713-790.

Kontorovich, Eugene (2004). "Liability Rules for Constitutional Rights: The Case of Mass Detentions," *Stanford Law Review* 56, 755-833.

Leshem, Shmuel (2010). "The Benefits of a Right to Silence for the Innocent," *RAND Journal of Economics* 41, 398-416.

Luban, David (2005). "Liberalism, Torture, and the Ticking Bomb," *Virginia Law Review* 91, 1425-1461.

Mialon, Hugo M. (2005). "An Economic Theory of the Fifth Amendment," *RAND Journal of Economics* 36, 834-849.

Mialon, Hugo M. and Rubin, Paul H. (2008). "The Economics of the Bill of Rights," *American Law and Economics Review* 10, 1-60.

Rejali, Darius (2007). *Torture and Democracy*, New Jersey: Princeton University Press.

Rizzo, Mario J. and Whitman, Douglas Glen (2004). "The Camel's Nose is in the Tent: Rules, Theories, and Slippery Slopes," *UCLA Law Review* 51, 539-592.

Seidmann, Daniel J. and Stein, Alex (2000). "The Right to Silence Helps the Innocent: A Game-Theoretic Analysis of the Fifth Amendment Privilege," *Harvard Law Review* 114, 431-510.

Shavell, Steven (2007). "Optimal Discretion in the Application of Rules," *American Law and Economics Review* 9, 175-194.

Siqueira, Kevin and Sandler, Todd (2007). "Terrorist Backlash, Terrorism Mitigation, and Policy Delegation," *Journal of Public Economics* 91, 1800-1815.

Sobel, Joel (2000). "A Model of Declining Standards," *International Economic Review* 41, 295-303.

Strauss, Marcy (2003). "Torture," *New York Law School Law Review* 48, 201-274.

The Economist (2006), "Saying No to Torture," *The Economist*, October 20.

The Pew Research Foundation (2009), "America's Place in the World 2009: An Investigation of Public and Leadership Opinion About International Affairs," December 2009, Available at Website http://people-press.org/reports/pdf/569.pdf.

Volokh, Eugene (2003). "The Mechanisms of the Slippery Slope," *Harvard Law Review* 116, 1026-1134.

Wantchekon, Leonard and Healy, Andrew (1999). "The 'Game' of Torture," *Journal of Conflict Resolution* 43, 596-609.

Wickelgren, Abraham L. (2010). "A Right to Silence for Civil Defendants?" *Journal of Law, Economics, and Organization* 26, 92-114.