

People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility [☆]

Philip E. Tetlock ^{a,*}, Penny S. Visser ^b, Ramadhar Singh ^c, Mark Polifroni ^d, Amanda Scott ^d, Sara Beth Elson ^d, Philip Mazzocco ^d, Phillip Rescober ^a

^a Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900, USA

^b University of Chicago, Chicago, IL, USA

^c National University of Singapore, Singapore, Singapore

^d The Ohio State University, Columbus, OH, USA

Received 4 May 2005; revised 3 February 2006

Available online 17 April 2006

Abstract

Three experiments explored determinants of punitive character attributions to norm violators. Experiment 1 showed that ideological conservatism and manipulated threat to society increased anger and attributional punitiveness when there was ambiguity about culpability. Experiment 2 showed that informing observers that norm violations were widespread and rarely punished increased attributional punitiveness by activating anger-charged retributive goals. Experiment 3 showed that liberals and conservatives alike felt justified in assigning greater blame to high-status perpetrators who commit acts of negligence with more severe consequences but that only conservatives felt justified in doing so for low-status perpetrators. Overall, the results reinforce the hypothesis that societal threat activates a prosecutorial mindset identifiable by a correlated cluster of attributions, emotions, punishment goals and punitiveness.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Attributions; Responsibility; Intuitive prosecutor; Social control; Ideology; Self-correction; Severity effect; Extenuation; Exacerbation; Accountability

Introduction

When confronted by norm violations, people have many coping options. Attribution theories expect people to consider a wide array of causal accounts, from the internal-controllable (e.g., evil intent) to the external-uncontrollable (e.g., exploitative system), that can trigger a corresponding array of reactions, from punitiveness to forgiveness (Skitka

& Mullen, 2002; Weiner, 1995). Culpable control theory expects people to favor explanations for norm violations that justify punishing perpetrators, ideally those with tainted reputations who fit the template for good suspects (Alicke, 2000). Just-world and system-justification theories expect people to neutralize threats to core beliefs by derogating victims (Lerner & Lerner, 1978) and affirming the legitimacy of the existing status hierarchy (Jost, Banaji, & Nosek, 2004).

Each framework captures key empirical regularities. The current article advances a new framework that draws on this past work but highlights neglected determinants of attributions of responsibility. The theoretical focus is on the fair-but-biased-yet-correctible (FBC) model grounded in the intuitive-prosecutor branch of functionalism (Tetlock, 2002). The FBC model builds on three uncontroversial sets of assumptions: (1) the fairness postulate—most people see themselves as fair-minded, anchor this self-image in their

[☆] Authors thank James Myers, Rachel Szteiter, Sara Hohenbrink, Jeanette Porubin, Meaghan Quinn, Brooke Curtiss, David Madan, Megan Berkowitz, Yuen Foong Khong, Kris Preacher, Jason Mills, Sarah Lowe, Galo Falchettore, and Shi-jie Chang for assistance in data collection and analysis. We also appreciate the support of the Burt Chair in psychology at The Ohio State University, the Mitchell Chair at the Haas School of Business at the University of California, Berkeley, and Grant # R-107-000-030-112 from the National University of Singapore to Ramadhar Singh.

* Corresponding author.

E-mail address: tetlock@haas.berkeley.edu (P.E. Tetlock).

adherence to shared norms of fair play, and are roused to retributive wrath when others display contempt for these norms (Lerner & Lerner, 1978; Miller & Vidmar, 1981); (2) the bias postulate—as creatures of bounded rationality with imperfect cognitive self-control, people often fall prey to judgmental biases that cause them either to over-weight irrelevant criteria or under-weight relevant criteria (Gilovich, Griffin, & Kahneman, 2002); (3) the self-correction postulate—when people catch themselves straying from their own private standards of good judgment, they try to correct themselves and self-correction is easiest in repeated-measures settings that facilitate monitoring the cues they are using (Petty & Wegener, 1998).

Fig. 1 lays out the logical structure of the model. Here we divide key FBC predictions into two categories: those bearing on situational and dispositional triggers of the prosecutorial mindset and those bearing on factors that influence ability and motivation to engage in cognitive self-correction.

Situational and dispositional triggers of the prosecutorial mindset

Theorists have long puzzled over how human beings manage to sustain intricate patterns of interdependence, and do so under conditions that game theorists see as profoundly unfavorable: strangers interacting in large groups

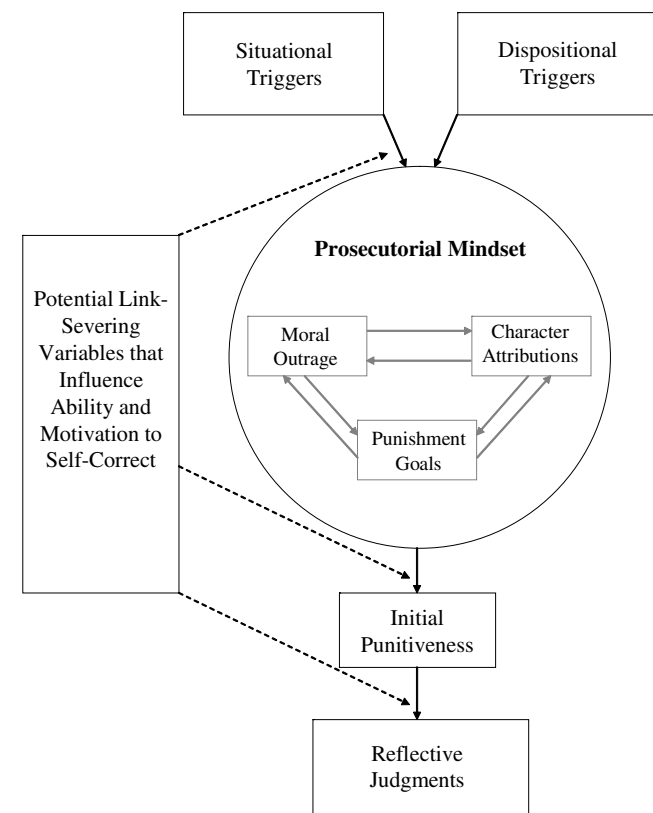


Fig. 1. The key logical components of the FBC model. The circled region designates the hypothetical construct, prosecutorial mindset, which subsumes the directly observable variables noted inside the circle.

in which monitoring by authorities is spotty and there is often little expectation of future interaction (Axelrod & Hamilton, 1981). Recently, evolutionary psychologists and behavioral economists have tried to solve this puzzle by positing a moralistic streak in human nature that predisposes people to value, as an end in itself, the punishment of norm violators. They argue that, for countless millennia before the emergence of centralized authority (the Hobbesian sovereign of political theory), social norms were upheld by censorious third-party observers, intuitive prosecutors, who willingly made sacrifices to punish cheaters, even those who have not cheated them (Fehr & Fischbacher, 2004).

Building on this work, the FBC posits a view of human nature that stresses both the readiness of people to internalize the normative order and to adopt a stance of prosecutorial vigilance toward norm violators. However, the FBC adds psychological complexity to the stylized social-watchdog view of human nature. The model posits a hypothetical construct, the prosecutorial mindset, which subsumes a cluster of causally interrelated indicators: a heightened likelihood of causal attributions holding norm violators culpable, of moral outrage, of endorsing retribution and general deterrence as goals of punishment, and of punishing both violators and those who fail to punish violators. Observers shift into this mindset to the degree they have been induced to believe, or were predisposed to believe that: (a) norm violations are widespread; (b) violations are intentional; (c) violators are flaunting their contempt for shared values; (d) violators are routinely escaping punishment; (e) the social order is legitimate; (f) the norm violations offend shared moral values. Observers disengage from the mindset only when reassured that norm violations have been: (a) punished by legitimate representatives of the collectivity; (b) punished sufficiently to satisfy the goals of retribution (inflict pain in just-deserts proportion to that inflicted) and general deterrence (warn those tempted to violate norms not to do so, thereby bringing crime down).

Thus defined, the prosecutorial mindset bears striking resemblance to Durkheim's (1925/1976) descriptions of collective reactions to normative transgressions, especially in its emphasis on how outraged people are by disrespect for their way of life and how satisfied they are when those who put themselves above the law are brought low. Linking these motivational postulates to recent work on social cognition, especially Kunda (1999), the FBC treats the mindset as a special case of motivated reasoning: (a) that is driven by shifting mixes of moral outrage and two punishment goals, the more impulsive, anger-charged, value-expressive goal of retribution and the more calculating instrumental goal of general deterrence; (b) that is capable—whenever there is sufficient causal ambiguity—of biasing inference in favor of internal-controllable attributions that depict violations as intentional, minimize wiggle room for escaping blame via excuses, and portray perpetrators as meriting punishment (Tetlock, 2000; Weiner, 1995).

It is worth clarifying though that the FBC is a social-functional theory that makes no strong process assumptions

about intrapsychic combinatorics. The model is agnostic on the degree to which: (a) the effects of chronic individual differences in the accessibility of cognitions and transitory manipulations of accessibility are additive or interactive (available evidence favors additivity, Higgins, 1996); (b) attributional appraisals drive outrage or outrage drives appraisals (available evidence suggests both cognitive and emotional mediation, Goldberg, Lerner, & Tetlock, 1999; Sunstein, Schkade, & Kahneman, 2000); (c) appraisals and outrage drive punitiveness or play a post-decisional role in justifying punitiveness. Indeed, the FBC views attributions, anger, punishment goals, and punitiveness as all indicators of one underlying construct, the prosecutorial mind, and thus expects these variables to fit a single-factor model.

Cognitive self-correction

The FBC model recognizes that people have limited access to their own cognitive processes and sometimes either over- or undershoot: making more or less punitive judgments than they can justify given the meta-norms (norms for enforcing norms) prevailing within valued reference groups. The FBC also recognizes that, although there is often consensus on the meta-norms for modulating accountability for norm violations, this consensus occasionally fissures along sub-cultural and ideological lines.

For instance, the FBC posits substantial consensus in Western societies on the factors—foreseeability, intentionality, duress—that determine the culpability of norm transgressors (the legal doctrine of *mens rea*, Hart, 1961). These factors define an exculpation continuum anchored at one end by free-will accounts that sway virtually everyone toward punitiveness (e.g., callous premeditation), at the other end by deterministic accounts that sway virtually everyone toward leniency (e.g., neurological disorder), and populated in the middle by accounts that invoke murky mixes of extenuating and exacerbating considerations (Schlenker, 1997). Linking the exculpation continuum with research on motivated reasoning, the FBC predicts that transitory manipulations of the prosecutorial mindset and chronic ideological variation in the mindset will exert most impact in the normatively hazy middle zone of the exculpation continuum, where the evidence is most open to interpretation and self-correction most difficult. Here, observers should be especially likely to fill in missing information with worst-case assumptions and to interpret existing information in ways that permit rejecting moderately mitigating accounts (e.g., peer pressure) that open the door to abuse excuses but permit accepting moderately exacerbating accounts (e.g., lack of remorse), that justify “cracking down” (Tetlock, 2000).

Integral though volition is to Western conceptions of responsibility, the FBC acknowledges that factors beyond individual control—such as crime waves and the severity of accidents—can influence punitiveness via familiar mechanisms such as affective spillover (Goldberg et al., 1999) and hindsight-bias distortions (Hawkins & Hastie, 1990). Such

effects are widely regarded as biases within psychology (Tetlock, 2002) but the FBC warns that this view is not universally shared: those who most value defending a given aspect of the normative order are most likely to affirm the justifiability of taking into account factors beyond the perpetrator’s control. Here the FBC yields unexplored predictions about how liberals and conservatives will vary in their views on the justifiability of allowing society-wide threat or accidental consequences to influence punitiveness. On the one hand, liberals and conservatives may be equally roused to righteous wrath by societal threat, via rapid-fire, often unconscious, cognitive-emotional mechanisms that, as human beings, they share (Goldberg et al., 1999). On the other hand, one group may subscribe to an intuitive theory of justice that, in this context, bars factors beyond individual control from swaying their judgment whereas another group may subscribe to an opposing theory. These two groups should be equally susceptible to influence by societal threat when they make their initial judgments in a between-subjects context but diverge in a repeated-measures context in which they can use covariation cues to infer their cue-utilization policy and correct for inappropriate influences.

Linking work on cognitive self correction (Petty & Wegener, 1998) with survey work on ideological belief systems (McClosky & Brill, 1983; Sniderman & Tetlock, 1986), the FBC posits that conservatives, for whom law and order is a trump values, generally feel more justified in allowing society-wide threat or severity of accidents to sway their judgments of culpability whereas liberals, for whom due-process rights trump, feel less justified in being so swayed. The FBC must, however, qualify this generalization because it traces punitiveness—in Durkheimian fashion—to the violation of shared moral values. It follows that, when liberal values have been more deeply violated (e.g., corporate wrong-doing), liberals will feel more justified in letting factors beyond individual control influence judgments of culpability.

Finally, drawing on Schlenker (1997) and Hamilton and Sanders (1992), the FBC acknowledges the power of organizational roles to influence judgments of culpability. When higher-status perpetrators betray the trust placed in them, they merit more, not less, outrage (Marques, Robalo, & Rocha, 1992). This analysis suggests that people will feel most justified in making punitive attributions to high-status perpetrators whose negligence causes severe consequences. In legal terms, people will feel freer to impose strict-liability standards for high-status perpetrators.

Overview

This article reports three studies that test hypotheses derived from the FBC and related models. Experiment 1 explores the impact of threat to social order, ambiguity about culpability, and observer ideology on judgments of norm violators. Experiment 2 explores the impact of adequacy of punishment of past violators and the mediating

role of punishment goals. Experiment 3 examines the sensitivity of managers to the severity of the unintended consequences of negligence by higher or lower status employees.

Experiment 1

This experiment was a 3 (activated-mindset, control, deactivated-mindset) \times 3 (liberal, moderate or conservative observers) \times 5 (repeated measure: exculpation continuum) mixed-factorial design that tested five hypotheses:

Hypotheses (1) and (2). Observers will direct more punitive attributions and anger at norm violators to the degree either that observers have been induced to believe that the social order is threatened or that observers were already predisposed to that belief;

Hypothesis (3). Inductions of societal threat and chronic individual differences in threat perceptions will however increase punitiveness only when there is enough ambiguity about the volitional nature of the violation to leave room for alternative interpretations from which motivated thinkers can “choose;”

Hypothesis (4). Conservatives, for whom the mindset is usually more accessible, will be more open to modulating blaming of individuals in response to hypothetical shifts in societal threat;

Hypothesis (5). Although plausible cases can be made for the primacy of either cognition (attributions) or affect (anger) as mediating variables, the dependent measures of attributions, anger, and punitiveness are most simply modeled as inter-correlated indicators of the same underlying construct, the prosecutorial mindset.

Method

Participants

A total of 294 undergraduate students (156 females and 126 males) were recruited from courses with research-participation requirements. Participants completed the study in groups of 11 to 16, and were randomly assigned to conditions within each session. Twelve participants with missing data were deleted from the analysis.

Procedure

The experiment was introduced as part of a “citizen-perspectives-on-crime” exercise designed to “give courts a better sense of how ordinary citizens size up the characters of criminal defendants and reach conclusions about appropriate punishment.” Participants were told that there are no objectively right answers: “even judges disagree over how appropriate it is to base sentences on the causes of conduct as opposed to the conduct itself.”

Ideology. Participants classified themselves on a three-category scale as liberal (agree that “in trying to strike the right balance between protecting society from crime and protecting the rights of citizens accused of crimes, our society too often fails to protect the rights of the accused;” $n=93$),

moderate (“society has struck the right balance ...;” $n=99$) or conservative (“society too often goes overboard protecting the rights ...;” $n=90$). This index captures the most theoretically relevant facet of the complex concept of ideology (Tetlock, 2000).

Manipulating mindsets. Participants were then randomly assigned to one of three conditions:

- (a) Participants in the *activation-of-prosecutorial-mindset* condition learned of a city that, 20 years ago, was a remarkably safe place to live. The vast majority of citizens respected the rule of law and each other’s rights. Back then, only a tiny fraction of citizens got into trouble with the law. And conviction rates for those few crimes averaged 96%. But, several years ago, conviction rates started falling and crime soared. So did contempt for authority. Last year was really bad: violent crime doubled and a record number of people, 5287 assault victims were injured badly enough to need hospital care. The police were overwhelmed. Conviction rates fell to an historic low, 4%: only 4 of every 100 serious crimes led to conviction. There was a record surge in complaints from citizens fearing for their safety. To vivify the threat, participants learned of a recent brutal attack on a grandmother escorting children to school. The grandmother was repeatedly stabbed in her stomach, back and neck by a purse snatcher who became enraged when she resisted. She died two days later. The perpetrator was never caught;
- (b) Participants in the *deactivation* condition learned of a remarkably law-abiding city identical to the one described in the activated-mindset condition before crime soared. This city, however, never changed. Citizens feel perfectly safe in walking the streets at night;
- (c) Participants in the control condition received no information on crime and conviction rates.

Crime scenarios. Participants were provided with “a factual description of a crime” and asked to judge the perpetrator’s responsibility and set punishment. In the scenario, a young man followed a well-dressed woman who was walking home at 10 p.m. Waiting until the street seemed empty, he grabbed her purse and, when she resisted, punched her several times. The woman needed to be hospitalized for two days for a broken jaw and for stitches to cuts on her face.” Participants then made their initial punitiveness, attributional and anger judgments. Participants were next shown, in randomized order, four possible explanations that pre-testing showed ranged from extremely extenuating to extremely exacerbating and that participants were told to assume captured the “true causes of the crime.” After reading each explanation, participants could adjust their judgments of the defendant if they so wished. The most extenuating scenario attributed the crime to a neurological problem, “a large tumor growing in a region of the

defendant’s brain that medical research has shown to be critical for making moral judgments and maintaining self-control in the face of temptation;” the moderately extenuating scenario cited a pre-sentencing report showing that the defendant could tell right from wrong but also noting that the defendant had recently lost his job, had money problems, and was afraid of being evicted from his apartment the moderately exacerbating scenario noted some financial need but cited a pre-sentencing report showing that the crime “was motivated by a conscious decision by the defendant that it is easier to make money by stealing than by working;” the most exacerbating scenario deleted all reference to financial need and added that “the defendant planned to use the money from the robbery to buy cocaine and the services of a prostitute.”

Dependent measures

Punishment. The punishment scale had a midpoint (5) anchored as “average punishment for first-time offender found guilty of robbery and assault and battery” and the endpoints as much less severe (1) and much more severe than average (9);

Locus of causation. Participants assessed responsibility on two 9-point scales: (a) one anchored by “due entirely to forces outside defendant’s control” (1) and “due entirely to forces in the defendant’s control” (9), with “5” as unsure; (b) the other anchored by how much the incident reveals about the defendant’s moral character (“reveals very little” or “a great deal”). These items were correlated ($r = .72$) and averaged to create a personal-responsibility index;

Anger. Participants rated how angry the norm violation made them feel (9 point scale from “not at all” to “extremely,” with the midpoint anchored at “somewhat”).

Repeated-measures: Would-you-change-your-mind-if...?

Participants were told there is debate among judges over whether the overall crime rate in society should sway judgments

of individual perpetrators. Participants in the activated-mindset condition were asked to suppose that the crime had occurred in a safe city identical to the one in the deactivated-mindset condition. Participants in the deactivated-mindset condition were asked to suppose that the crime had occurred in a dangerous city identical to the one in the activated-mindset condition. Control participants were asked to suppose that the high-crime rates in the activated-mindset condition held true. All participants then rated how much these new circumstances would change their minds about punishment (same 9-point severity scale) and personal responsibility. For this latter variable, we asked: “There is always uncertainty when we draw conclusions about how truly sorry law breakers are for their crimes. Would the new circumstances move you toward a “better-safe-than-sorry” policy in estimating how dangerous the defendant is (9) or toward a “give-him-the-benefit-of-the-doubt” policy (1)? The midpoint, 5, represented “no change.”

Post-experimental questions. These items probed the believability of the activate- and deactivate-mindset manipulations (1 = very hard to believe; 9 = very easy; 5 = unsure).

Results

The punitiveness and attribution-of-responsibility variables were merged to create a composite measure of “attributional punitiveness.” The justification was twofold: the hefty correlation between the two indicators (across repeated measures, $r = .79$) and the almost identical profiles of main and interactive effects. Anger correlated .56 with the composite indicator but yielded a somewhat different pattern of effects.

Table 1 presents the mean attributional punitiveness and anger judgments. The 3 (mindset) × 3 (ideology) × 5 (extenuation/exacerbation) repeated measures analyses of variance of attributional punitiveness and of anger tested the five sets of hypotheses:

Table 1

Average attributional punitiveness (AP) and anger (AN) toward the defendant as a function of prosecutorial mindset, ideology of the observer and the degree to which exacerbating or extenuating circumstances were present

Prosecutorial mindset	Ideology of the observer	Exculpation continuum									
		Extr. Exac.		Mod. Exac.		Neutral		Mod. Exten.		Extr. Exten.	
		AP	AN	AP	AN	AP	AN	AP	AN	AP	AN
Activated	Liberal ($n = 28$)	6.9 (1.1)	6.7 (1.8)	5.6 (1.0)	6.5 (1.7)	4.9 (0.9)	6.0 (1.4)	4.4 (0.9)	5.0 (1.6)	2.2 (1.3)	2.5 (1.2)
	Moderate ($n = 30$)	7.1 (1.1)	7.4 (1.5)	6.2 (1.0)	6.3 (1.4)	5.4 (0.9)	5.5 (1.6)	5.2 (0.7)	4.5 (1.3)	2.1 (0.8)	2.9 (1.3)
	Conservative ($n = 36$)	7.1 (1.2)	7.7 (1.0)	6.8 (1.0)	6.8 (1.0)	5.9 (0.9)	6.1 (1.3)	5.6 (0.8)	5.3 (1.3)	2.0 (0.6)	2.5 (0.9)
Deactivated	Liberal ($n = 35$)	6.8 (1.1)	6.9 (1.6)	5.3 (0.9)	5.8 (1.4)	4.8 (0.9)	5.4 (1.1)	4.1 (0.9)	4.4 (0.9)	2.0 (0.8)	2.7 (1.4)
	Moderate ($n = 35$)	7.0 (1.1)	7.2 (1.7)	5.8 (1.1)	5.9 (1.2)	4.8 (0.8)	5.6 (1.5)	4.7 (1.0)	4.7 (1.6)	1.9 (0.6)	2.8 (1.2)
	Conservative ($n = 29$)	7.0 (1.3)	7.2 (1.4)	6.0 (1.0)	6.4 (1.2)	5.2 (0.8)	5.9 (1.3)	4.8 (0.8)	5.1 (0.9)	1.8 (0.7)	2.7 (1.4)
Neutral	Liberal ($n = 28$)	6.6 (1.2)	6.9 (1.3)	5.2 (1.2)	6.3 (1.0)	4.5 (0.9)	5.8 (0.9)	3.8 (0.7)	5.2 (1.4)	2.2 (1.2)	2.7 (1.6)
	Moderate ($n = 30$)	6.8 (1.3)	7.0 (1.4)	5.6 (1.1)	6.0 (1.3)	4.8 (0.9)	5.2 (1.2)	4.9 (0.6)	4.4 (1.1)	2.1 (0.9)	2.4 (1.2)
	Conservative ($n = 36$)	6.9 (1.2)	7.3 (1.3)	5.7 (1.0)	6.0 (1.5)	5.2 (0.9)	5.8 (1.4)	4.9 (0.7)	5.0 (1.1)	2.2 (1.0)	2.5 (1.3)

“Presented second” means deactivated mindset for the activated-prosecutorial mindset condition and means activated mindset for the deactivated and control conditions.

Hypothesis (1). Planned contrasts found that, relative to the control and deactivated conditions, the prosecutorial-mindset manipulation increased attributional punitiveness, M 's = 5.16 versus 4.79 and 4.76, $F(1,271) = 11.95$, $p < .001$, $\omega^2 = .09$, and anger, M 's = 5.40 versus 5.25 and 5.18, $F(1,271) = 3.32$, $p < .05$, $\omega^2 = .03$. Contrary to expectation, the deactivated-mindset manipulation had no impact on either attributional punitiveness or anger ($F < 1$).

Hypothesis (2). Planned contrasts showed that conservatives expressed greater attributional punitiveness than did moderates and liberals, M 's = 5.13 versus 4.96 and 4.61, $F(1,271) = 10.32$, $p < .001$, $\omega^2 = .10$, and greater anger, M 's = 5.45 versus 5.18 and 5.25, $F(1,271) = 4.05$, $p < .05$. Moderates expressed greater attributional punitiveness than did liberals, $F(1,271) = 5.26$, $p < .05$, M 's = 4.96 versus 4.61, but not greater anger ($p > .15$).

In line with past work, the effects of the transitory manipulation of societal threat and of chronic ideological differences in perceived threat were additive. The interactions for both attributional punitiveness and anger were not significant (p 's $> .15$).

Hypothesis (3). Fig. 2 highlights the powerful linear trend toward less attributional-punitiveness as judges moved from exacerbating to extenuating scenarios, $F(1,271) = 120.13$, $p < .001$, $\omega^2 = .31$. (Given the semantic content of the manipulation, this effect is best viewed as a manipulation check.) Fig. 2 also reveals support for the predicted mindset-by-exculpation and ideology-by-exculpation interaction effects on attributional punitiveness, $F(2,271) = 10.67$, $p < .01$, $\omega^2 = .06$, and $F(2,271) = 10.67$, $p < .01$, $\omega^2 = .07$. Consistent with the motivated-reasoning hypothesis, the mindset induction increased attributional punitiveness (relative to the control and deactivated

conditions) only for moderately exacerbating, neutral, and moderately extenuating scenarios along the continuum, $F(1,271) = 18.45$, $p < .01$, and not for the extremely exacerbating and extenuating scenarios (F 's < 1). Planned contrasts also showed that conservatives were more attributionally punitive than moderates and liberals only for the moderately exacerbating, neutral, and moderately extenuating scenarios, $F(1,271) = 22.54$, $p < .01$, and, again, not for the extremely exacerbating and extenuating scenarios (F 's < 1). The three-way interaction of mindset, ideology, and exculpation was not significant ($F < 1$).

Turning to emotions, planned contrasts revealed: (a) a linear decline in anger along the exculpation continuum, $F(1,271) = 119.52$, $p < .001$ (unsurprisingly, it was harder to get mad at violators suffering brain disorders); (b) a quadratic interaction in which anger ebbed more slowly in the middle of the exculpation continuum among conservatives than among moderates and liberals, $F(1,271) = 5.57$, $p < .05$. No other interactions were significant (all p 's $> .10$).

Hypothesis (4). To map ideological variation in willingness to change one's mind, difference scores were computed between revised punitiveness and character judgments and original judgments. As hypothesized, conservatives in the control and deactivated conditions were more willing than liberals and moderates to increase their punitiveness and attribution-of-responsibility judgments in response to questions about how they would shift their views if crime were a more severe problem than specified earlier (the average attributional punitiveness difference scores for conservatives, 1.24 and 1.31, $F(1,271) = 8.45$, $p < .01$, for moderates .71 and .43, $F(1,271) = 3.95$, $p < .05$, and for liberals .21 and .26, $F(1,271) = 1.47$, ns). Contrary to hypothesis, no group decreased its punitiveness and responsibility judgments

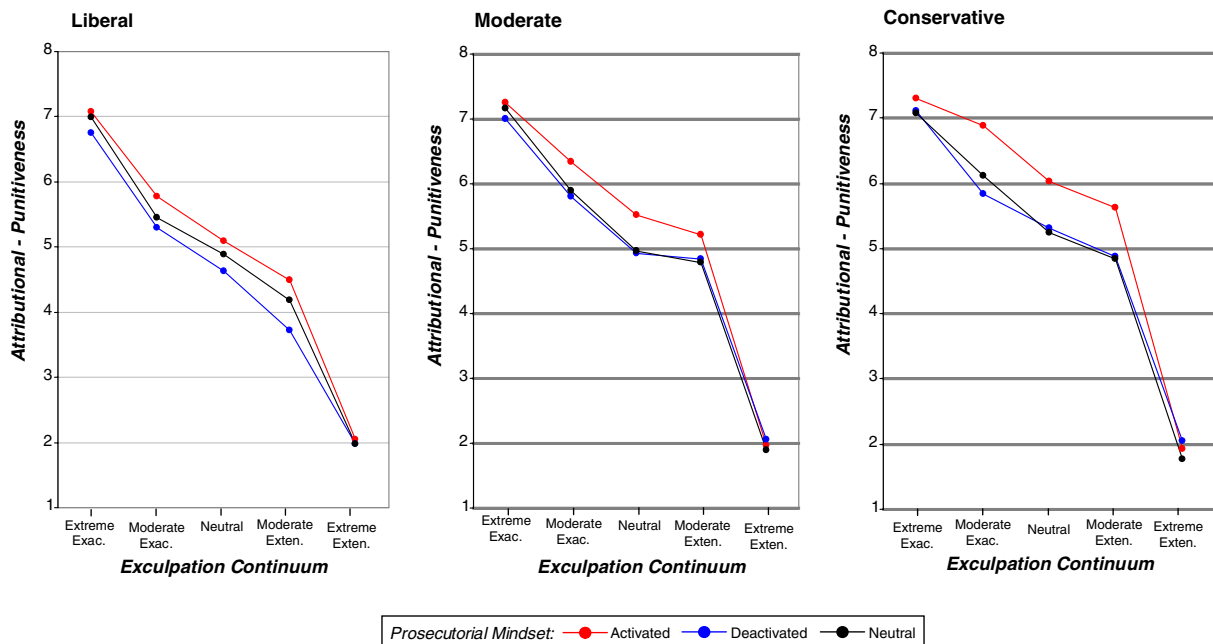


Fig. 2. Plot showing average attributional punitiveness as a function of mindset manipulation, ideology of observer, and location of crime scenarios along exculpation continuum.

when asked to suppose that crime was less of a problem than stated earlier ($F < 1$). The heightened willingness of conservatives to become more punitive in response to rising crime rates also only held up when participants judged scenarios along the exculpation continuum in which the norm violation was, to some degree, volitional, disappearing when the norm violation was the product of clear biological causation, M 's for conservative difference scores for four volitional scenarios, 1.43 and 1.48, and M 's for brain tumor scenario, .08 and .11, $F(1, 271) = 22.75$, $p < .01$.

Hypothesis (5). Sobel tests explored the role of anger in mediating the effects of mindset, ideology and exculpation on attributional punitiveness. These tests showed that anger fully mediated the exculpation effect (Sobel coefficient = .26, $SE = .06$, $z = 4.14$, $p < .001$), rendering it non-significant ($b = .07$, $SE = .05$, $t = 1.43$, ns), and partly mediated the effects of both mindset and of ideology (Sobel coefficients $b = .24$ and $.29$, $SE = .11$ and $.13$, $z = 2.21$ and 2.29 , $p < .05$), leaving both direct effects significant at .05. Sobel tests were also, however, consistent with the hypothesis that causality flowed in the opposite cognition-to-emotion direction. Attributional punitiveness largely mediated the effects of mindset, ideology, and exculpation on anger. Indeed, only the exculpation-anger relationship withstood controlling for attributional punitiveness, $b = .10$, $SE = .05$, $t = 1.98$, $p < .05$. Principal-axis factor analysis reinforced the futility of statistical efforts to sort out causality: a single factor model accounted for 74% of the variance, with loadings for the separate indicators of attributions (.83), anger (.63) and punitiveness (.79). No other factor had an eigenvalue greater than 1.0.

Discussion

Participants punish norm violators more harshly, hold them more responsible, and are angrier at them to the degree that either participants have been induced to believe that the social order is under siege or participants were already predisposed to this view. These findings mesh well not just with FBC predictions but with a variety of past work, including: Durkheim's (1925/1976) sociological studies of collective reactions to transgressions; contemporary survey studies of the links between conservatism and concern for social order (McClosky & Brill, 1983) and the power of "sociotropic" threat to inflame public opinion (Davis & Silver, 2004); and neuroscience studies of the pleasure people derive from punishing cheaters (Fehr & Fischbacher, 2004).

The FBC should, however, get its biggest credibility boost from the support for its more subtle interaction predictions. The transitory manipulation of prosecutorial mindset and chronic variation in ideology had their greatest effects on attributional punitiveness when there was greatest potential for motivated reasoning to bias causal interpretations: for scenarios in the normatively hazy middle zone of the exculpation continuum. Turning to the change-your-mind judgments, conservatives were more

open than liberals to strengthening their attributions of culpability in response to deteriorating societal conditions in all but the tumor scenario. These latter results show how groups can be equally swayed by an independent variable in between-subject designs in which participants have no clue what is being manipulated but respond differently in repeated-measures designs in which manipulations are transparent.

Experiment 1 also connected the FBC to two other lines of social psychological work. It replicated: (a) the well-established positive correlations among internal-controllable attributions, punitiveness and anger (Weiner, 1995). Indeed, the size of these correlations supported the contention that these dependent variables are most parsimoniously viewed as indicators of a single common construct (the prosecutorial mindset), underscoring the futility of teasing apart statistically the causal priority of cognitive appraisals versus emotions in the mediational sequence; (b) the additive effects of chronic individual differences in accessibility and transitory experimental inductions (Higgins, 1996). Liberals are as easy to rouse to punitiveness as conservatives via manipulations of societal threat—even though they are more reluctant than conservatives to use such information openly in a repeated measures context.

The most unexpected result was how much easier it was to elevate than to depress punitiveness. This asymmetry may, in part, have methodological causes. Post-experimental probes revealed that, although participants mostly believed the high-threat induction (credibility rating of assertion that "there is a city in the United States in which only 4% of violent crimes are punished," $M = 5.5$), they doubted the low-threat one (credibility rating of "there is a city in which 96% of violent crimes are punished," $M = 3.4$). The asymmetry may however have deeper psychological causes. Change-your-mind questioning revealed that: (a) when low-societal-threat participants (especially conservatives) imagined shifting into the high-threat condition, many felt justified in cracking down; (b) by contrast, when high-threat participants imagined shifting into the low-threat condition, few felt justified in easing up—and treating perpetrators more leniently because almost no one is getting away with breaking the law. Post-experimental questioning revealed a mix of retributivist and deterrence motives: when norm violations were rising, the modal participant saw it as a sign that penalties had been set too low but when norm violations were falling, few thought it acceptable to "lighten up." Rather, they concluded that penalties were working and it would be wrong not to make criminals "pay for their crimes" just because there was less crime. This inflexibility may explain why, in time-series analyses of public opinion, the desire to get tough rises more in response to growing crime than it declines in response to falling crime, creating punitiveness traps (Cullen, Fisher, & Applegate, 1998). We measure the mediational role of punishment goals, retributive and deterrence motives, in the next two studies.

Experiment 2

Experiment 2 is a 2 (activation-versus-deactivation-of-prosecutorial-mindset) \times 3 (no prior case or either inadequate-or-adequate punishment in prior case) factorial that tests three sets of hypotheses: (1) the combination of rising crime and inadequate punishment in a salient precedent will be a potent trigger of attributional punitiveness, anger and the punishment goals of retribution and deterrence toward norm violators; (2) the attributional punitiveness effect will be mediated jointly by the inter-correlated indicators of anger and the activation of retribution and general deterrence goals; (3) the more participants endorse deterrence and retribution, the more justified they will feel in increasing punitiveness in response to societal threat.

Method

Participants

A total of 99 undergraduates (54 females; 45 males) were drawn from courses with research requirements. Participants worked in groups of six to ten, and were randomly assigned to conditions within sessions. Four participants with missing data were dropped from analyses.

Procedure

The same cover story was used as in Experiment I. Half of the participants were assigned to the activation-of-prosecutorial-mindset condition in which they learned that perjury is reaching epidemic proportions in a large American city. People do not take oaths to tell the truth as seriously as they once did and, as a result, there have been horrible miscarriages of justice. A recent survey found that most citizens now had “deep doubts” about the ability of courts to “get to the truth.” The other half were assigned to the deactivation-of-mindset condition in which they learned that perjury is rare and, moreover, on the decline. A recent survey found great confidence among citizens in the ability of courts to get to the truth.

All participants (except those in the no-prior-case, control condition) then learned of a case in which two executives framed an innocent colleague by inserting false dates into documents and by lying about it to Securities-and-Exchange-Commission investigators and the court. The colleague was convicted, sentenced to five years of imprisonment, lost his family, and was so depressed that he attempted suicide. Participants also learned that, three years later, the deception was uncovered through careful analysis of computer records of who was where in the building at critical times. The two executives, a man and a woman, were convicted of perjury.

Here the second independent variable was introduced. Participants in the adequate-punishment condition learned that the judge was unmoved by the defendants’ pleas for leniency and sentenced them to the term recommended by the official sentencing guidelines: five years of imprisonment (the median recommendation in pilot testing). The

judge stressed how important it was that the defendants spend at least as much time in prison as the defendants had inflicted on an innocent person. Depending on whether participants had been randomly assigned to the activated or deactivated mindset condition, the judge explained his decision in one of two ways: “perjury is getting out of control and cracking down feels like the right course of action in this case” or “even though perjury is not at all that common, cracking down feels like the right course of action in this case.” Participants in the inadequate-punishment condition learned that the judge felt so much sympathy for the defendants’ pleas for leniency that he changed his mind and that, instead of sentencing them to five years as specified by the law, he imposed no jail time. Depending on whether participants had been assigned to the activated or deactivated mindset condition, the judge explained his decision in one of two ways: “even though perjury is getting out of control, leniency feels like the right course of action in this case” or “perjury is not all that common and leniency feels like the right course of action in this case.”

Participants rated the appropriateness of the judge’s decision on a 9-point scale (a manipulation check for the adequacy-of-punishment manipulation). Participants were then given an opportunity to offer character assessments and to make punishment recommendations in a new case that involved perjury by two business executives who had covered up the failure of the owners of the factory to implement worker-safety measures. This failure resulted in serious injury (loss of an arm) to a worker. The executives were MBAs in their mid-30s who were eager to maximize their bonuses by economizing on safety; the worker was a divorced mother, in her mid-40s, who was trying to support her family.

Dependent measures

Participants responded to the same punishment, locus of causality and anger scales used in Experiment I. They also judged on 9-point scales: (a) the justifiability of allowing punitiveness judgments of individual perpetrators to be swayed by the goals of general deterrence (sending a message to others) and retribution (inflicting pain commensurate with that inflicted on others); (b) the justifiability of becoming more or less lenient if: (i) perjury were less common (for those in the activated-mindset condition) or more common than they had supposed (for those in the deactivated condition); (ii) the defendants in the previous case had gotten a tougher sentence (for those in the “inadequate-punishment” condition) or a more lenient sentence (for those in the adequate-punishment and control conditions).

Results

Punitiveness and attributions of responsibility were again sufficiently correlated ($r = .73$), and similarly related to independent variables, to justify a composite measure of attributional-punitiveness. Table 2 presents the means and

Table 2

Mean attributional punitiveness, anger, retribution and general deterrence scores (9-point scales) in response to the prosecutorial-mindset and adequacy-of-punishment manipulations in Experiment II

Prosecutorial mindset	Adequacy of punishment in previous case	Attributional punitiveness	Anger	Retribution	General deterrence
Weak-activation-of-mindset	Inadequate ($n = 16$)	6.53 (.95)	6.13 (.72)	6.00 (1.21)	7.51 (0.98)
	No prior case ($n = 15$)	5.95 (.90)	5.87 (1.3)	5.4 (1.35)	7.19 (1.05)
	Adequate ($n = 17$)	5.97 (1.01)	5.75 (1.0)	5.25 (1.06)	7.15 (1.09)
Strong-activation-of-mindset	Inadequate ($n = 18$)	7.35 (1.02)	6.88 (1.15)	6.78 (1.30)	7.71 (0.85)
	No prior case ($n = 14$)	6.43 (.77)	6.44 (.73)	6.25 (1.29)	7.49 (0.80)
	Adequate ($n = 16$)	6.12 (.87)	6.19 (.75)	5.63 (1.204)	7.52 (0.77)

standard deviations for attributional punitiveness and three other hypothesized indicators of the prosecutorial mindset, anger, retribution and general deterrence.

Hypothesis (1). Relative to the no-mindset control, the prosecutorial-mindset stimulated attributional punitiveness, M 's = 6.63 versus 6.02, $F(1, 89) = 6.25$, $p < .05$, $\omega^2 = .07$; anger, M 's = 6.5 versus 5.91, $F(1, 89) = 4.88$, $p < .05$, $\omega^2 = .05$, and retribution goals, 6.56 versus 5.53, $F(1, 89) = 5.39$, $p < .05$, $\omega^2 = .08$, but did not stimulate general deterrence goals, M 's = 7.58 versus 7.28, $F(1, 89) = 1.54$, ns. Relative to the no-prior-case control and the adequate-punishment-in-prior-case conditions, which did not differ from each other on any of the four dependent variables ($F < 1$), the inadequate-punishment manipulation also stimulated attributional punitiveness, M 's = 6.94 versus 6.50, $F(1, 89) = 7.98$, $p < .01$, $\omega^2 = .09$; anger, M 's = 6.50 versus 6.06, $F(1, 89) = 3.15$, $p < .05$, $\omega^2 = .05$, and retribution, M 's = 6.39 versus 5.64, $F(1, 89) = 6.22$, $p < .05$, $\omega^2 = .09$, but did not stimulate general deterrence goals ($p > .15$). Planned contrasts supported the FBC's interaction prediction on three of the four key dependent variables. As Table 2 shows, attributional punitiveness, anger, and retribution goals peaked in the prosecutorial-mindset-and-inadequate-punishment condition (M 's = 7.35, 6.88, and 6.78), values that were greater than those in the other five conditions taken as a whole, $F(1, 89) = 8.32$, $p < .01$, $\omega^2 = .09$, $F(1, 89) = 6.01$, $p < .05$, $\omega^2 = .06$, $F(1, 89) = 5.88$, $p < .05$, $\omega^2 = .06$, or taken separately. The general deterrence variable did not register the interaction ($p > .15$).

Hypothesis (2). Multiple regressions assessed whether the main and interactive effects of mindset and adequate punishment were mediated by the arousal of anger and retribution and general deterrence goals. These tests entered six predictors (three manipulated independent variables and three possible mediators) and showed that anger and retribution were both robust predictors of attributional punitiveness, b 's = .29 and .25, SE 's = .06 and .05, t 's = 4.71 and 4.9, p 's < .01, that jointly reduced the mindset effect to non-significance, $b = .08$, $SE = .13$, $t = .64$, ns and that jointly reduced but did not eliminate the adequate-punishment main effect and mindset-by-adequate-punishment interaction, b 's = $-.21$ and $-.17$, SE 's = .08 and .07, t 's = -2.62 and -2.44 , p 's < .05. Contrary to hypothesis, general deterrence did not emerge as an independent predictor, $b = .15$, $SE = .10$, $t = 1.51$, ns. Principal axis factor analysis again

reinforced the futility of statistical efforts to sort out causality: a single factor model accounted for 52% of the variance and yielded loadings for attributional punitiveness (.76), anger (.61), retribution (.59) and punitiveness (.79) and deterrence (.29). Dropping deterrence allows the first factor to explain 64% of the variance.

Hypothesis (3): Participants were again divided over the justifiability of changing their minds in response to shifts in societal threat. They felt more justified in increasing punitiveness when the social order was deteriorating than in decreasing punitiveness when the social order was improving, $F(1, 89) = 6.34$, $p < .01$. This willingness to adjust judgments in response to societal threat was especially true for respondents who valued retribution and general deterrence as legitimate goals in setting punishment for transgressors ($r = .37$ and $.25$).

Discussion

Consistent with the work of Goldberg et al. (1999), high threat to society and inadequate punishment in a salient precedent were potent triggers, individually and jointly, of punitiveness. Consistent with the work of Carlsmith, Darley, and Robinson (2002), the activation of emotionally charged retribution goals played a mediating role that completely eclipsed general deterrence goals. These findings suggest that the FBC model should give greater weight to retribution and less to general deterrence as a driver of punitiveness—at least for intentional norm violations. Post-experimental questioning buttresses this conclusion. A common reaction to the inadequate-punishment scenario was controlled outrage. Several participants criticized the judge for “falling for another con job.” A common reaction to the adequate-punishment scenario was that the defendants “got what they deserved.”

Experiment 2 showed that again, even in scenario experiments, participants' self-reported cue-utilization strategy (the claimed irrelevance of prior case information) can be fallible. These self-reports clash with the evidence of between-subject effects for the prior-case manipulation, raising the possibilities that participants either lack access to the causes of their judgments or are unwilling to acknowledge those causes.

Post-experimental questioning also probed whether participants thought it appropriate to reduce punishment if:

(a) an “infallible lie detector could raise the detection rate for perjury to 100% (the answer was 75% negative); (b) the defendants were injured in an accident prior to sentencing (the answer was 80% negative). These results underscore the view that punitiveness was largely driven by the backward-looking goal of retributive justice: making perpetrators suffer in just-desert proportion to their offense and ensuring that this suffering is no fluke of nature but rather has been methodically administered in the name of authorities that observers endow with legitimacy (Duff, 1994; Feinberg, 1984; Miller & Vidmar, 1981).

Experiment 3

Experiment 3 extends the retribution-and-deterrence logic of the FBC to the realm of torts—in particular, to the severity effect in the attribution-of-responsibility literature. Numerous experiments have shown that, holding all else constant (including foreseeability), people view those who cause accidents as more culpable as the consequences grow in severity (Burger, 1981). Most researchers have treated the effect as a bias and explained it by invoking cognitive and emotional mechanisms to which people, regardless of ideology, are unlikely to have introspective access (such as hindsight bias).

The FBC predicted: (1) a replication of the severity effect—as the severity of the consequences of an accident grows, observers will make more punitive attributions to both the individual causal agent and the individual’s employer and endorse both retribution and general deterrence more strongly as goals of punishment; (2) a replication of organizational-role effects (Hamilton & Sanders, 1992)—as the status of the causal agent rises, observers will make more punitive attributions to both the individual and the employer and endorse both retribution and general deterrence as punishment goals; (3) the severity and status effects will be mediated by the activation of the goals of general deterrence and retribution; (4) the severity and status effects will be moderated by ideology, especially after the manipulations have been explicitly cued by “would-you-change-your-mind” questioning. When judging the low-status perpetrator, conservatives will be more willing than liberals to defend the severity effect—and this effect will be driven by the stronger activation of deterrence and retribution goals among conservatives. However, both tendencies should be reversed when the perpetrator belongs to a class (highly paid executives) that stimulates more resentment among egalitarian liberals than among hierarchy-defending conservatives. The prediction should thus be an interaction. Whereas conservatism should be correlated with stronger severity effects in the low-status conditions, liberalism should be correlated with stronger effects in the high-status conditions. Both tendencies should be mediated by the activation of deterrence and retributive goals.

To test these hypotheses, the third experiment took the form of a 2 (low-high severity) \times 2 (low-high status of perpetrator) factorial that explored reactions to negligence by

low or high status individuals that produced minor or major consequences.

Method

Participants

Respondents were business undergraduate and graduate (MBA) students who took part in a class exercise on ethics and who included 44 men and 28 women.

Procedure

The scenario described a multinational firm that produces a wide range of canned food products. Production specialists had reported the malfunctioning of one of the machines that mixes the broth for soups and warned the Vice President for Production Operations of the potential danger of bacterial contamination. The Vice President drafted a detailed ALERT memo to all plant supervisors that told them to suspend all production until the problem is fixed.

Status and severity manipulations. In the high-status-perpetrator condition, participants learn that, before the executive could give the memo to his clerical assistant in charge of distributing such information, he received a phone call in which he learned that he had received a job offer, with a big increase in salary in a more attractive city to which he and his wife had long wanted to move. The time was 5:30 p.m. on a Friday afternoon and the executive forgot to forward the memo for distribution. Indeed, he did not remember to do this until he returned to work Monday morning. By that time, nearly 200,000 cans of soup had been shipped out to stores and it turned out that 2 (low severity) or 20,000 (high severity) of these cans were contaminated by bacteria that can cause illness. As a result, two people suffered mild stomach upset (low severity) or 2477 people suffered severe stomach pain and several older patients nearly died (high severity).

In the low-status perpetrator condition, the executive gave the ALERT memo to his clerical assistant in charge of distributing such information. But before the assistant could initiate the distribution of the information, the assistant received the job-offer phone call and events then follow down the same low or high severity paths.

Dependent variables

Respondents made the following judgments on 9-point scales: (1) punitiveness-toward-individual scales (the degree to which the causal agent should be subject to the maximum disciplinary penalties that: (a) the company can impose (ranging from no penalty (1) to demotion and cuts in salary and benefits (5) to dismissal and loss of all retirement and stock-option benefits (9)); (b) society can impose (ranging from no criminal liability (1) to misdemeanor charges of criminal negligence causing bodily harm (5) to felony charges of criminal negligence causing bodily harm (9)); (2) attributions about character of causal agent (from

1 (no reason to draw conclusions about how moral this person is) to 9 (very strong reasons for drawing conclusions about how immoral this person is), with 5 set at “moderately good reasons”); (3) punitiveness-toward-company scales (the degree to which the company should be subject to fines and damage awards (ranging from no fines or awards (1) to maximum possible fines and awards (9), with five representing the midpoints)); (4) organizational-culture attributions to company (9 point scale from 1 (no reason to draw conclusions about how socially responsible this company is) to 9 (very strong reasons for drawing conclusions about how socially responsible this company is), with 5 as “moderately good reasons”). Participants were also asked to: (a) imagine that the circumstances were not as first described to them but rather as described in each of the other three scenarios in the 2 (severity) \times 2 (status) factorial (e.g., lower status if originally assigned to the high-status condition and less severity if originally in the high severity condition) and then repeat the penalty and character ratings; (b) rate themselves on a 9-point liberalism-conservatism scale as well as rate the justifiability of being guided by “general deterrence” (setting an example that encourages others to be more careful) and retribution (inflict suffering on wrong-doers in just proportion to what they have inflicted on others).

Results

The punitiveness-toward-individual and character-attribution scales were sufficiently correlated ($r = .72$) to warrant merger into a composite index. So too were the scales for punitiveness toward company and organizational-culture attributions ($r = .69$). Table 3 presents the average attributional punitiveness toward individual and company as well as the average endorsement of general deterrence and retribution as goals of punishment.

Hypothesis (1). The severity effect emerged. People displayed greater attributional punitiveness toward both the individual and the company when hundreds of people became very ill than when a handful of people became slightly ill, $M_s = 6.57$ versus 5.52, $F(1,64) = 25.68$, $p < .001$, $\omega^2 = .14$, $M_s = 5.75$ versus 4.95, $F(1,67) = 15.34$, $p < .01$, $\omega^2 = .11$. People also endorsed general deterrence goals more strongly, $M_s = 6.7$ versus 4.45, $F(1,67) = 24.89$, $p < .001$, $\omega^2 = .15$, and to a significant but lesser extent, retri-

bution goals, $M_s = 3.6$ versus 3.0, $F(1,64) = 5.65$, $p < .05$, $\omega^2 = .06$.

Hypothesis (2). The status effect emerged. People directed greater attributional punitiveness at both individual perpetrator and the company when the perpetrator was a senior executive: individual, $M_s = 6.61$ versus 5.9, $F(1,67) = 21.48$, $p < .01$, $\omega^2 = .08$; company, $M_s = 6.75$ versus 3.95, $F(1,67) = 19.38$, $p < .01$, $\omega^2 = .18$. People also endorsed general deterrence goals more strongly, $M_s = 5.9$ versus 5.25, $F(1,67) = 18.89$, $p < .001$, $\omega^2 = .15$, and to a significant but lesser extent, retribution goals, $M_s = 3.45$ versus 3.15, $F(1,67) = 4.98$, $p < .05$, $\omega^2 = .05$. No severity-by-role interaction emerged for attributional punitiveness toward the individual or toward the company when negligence by a high status perpetrator caused severe consequences, $F < 1$, in both cases. And no such interaction emerged for the punishment goals of retribution and general deterrence, $p_s > .15$.

Hypothesis (3). Multiple regressions showed that: (a) the severity effect on both individual and corporate culpability was partly mediated by the activation of general deterrence goals, $b_s = .25$ and $.29$, $SE = .11$ and $.10$, $t_s(65) = 2.36$ and 2.92 , $p < .01$, but the severity effects remained significant after controlling for this variable, $b_s = 1.49$ and 1.67 , $SE = .37$ and $.41$, $t_s(65) = 4.03$ and 4.14 , $p < .01$. Retribution did not qualify even as a partial mediator for either individual or corporate culpability, $t_s < 1.5$, ns ; (b) the status effect on individual and corporate culpability was wholly mediated by general deterrence, $b_s = .58$ and $.82$, $SE = .09$ and $.12$, $t(65) = 5.99$ and 6.86 , $p_s < .01$, leaving no direct effects for status, $t_s < 1.5$, ns . Retribution again failed to qualify even as partial mediator, $t_s < 1.5$, ns .

Hypothesis (4). Although ideology was a weak predictor of the susceptibility of first-reaction judgments to the severity effect (multiple regressions revealed no ideology-by-severity or ideology-by-severity-by-status interactions), ideology did predict more reflective second-order judgments on the appropriateness of increasing or decreasing blame as consequences grew more or less severe. But context matters. Conservatives and moderates (identified by tertile split) were consistent defenders of openly using severity information when they judged the individual culpability of both low and high status agents, blaming the agent less when asked to imagine less severe consequences, difference scores representing gaps between initial judgments and change-your-mind judgments, $M_s = -1.55$ and -1.61 , test of difference from zero: $t(22) = 2.15$, $p < .05$, and blaming the agent more when imagining more severe consequences, difference score $M_s = 1.66$ and 1.78 , test of difference from zero: $t(22) = 2.44$, $p < .05$. Liberals were less consistent defenders of using severity information. On the one hand, they readily blamed high-status perpetrators more in response to more severe consequences, difference score $M_s = +1.89$ and $+2.32$, $t(12) = 2.35$, $p < .05$, and blamed low-status perpetrators less in response to less severe consequences, difference score $M_s = -1.99$ and -1.65 , $t(12) = 1.97$, $p < .05$. On the other hand, they

Table 3

Average scores (with standard deviations) in the low and high status and severity conditions for the following variables: attributional punitiveness toward the individual, attributional punitiveness toward company, general deterrence goals, and retribution goals

Severity	Status	Attributional punitiveness to individual	Attributional punitiveness to company	General deterrence	Retribution
Low	Low	4.95 (1.15)	3.8 (0.8)	4.2 (1.3)	2.9 (0.9)
Low	High	5.95 (1.1)	4.1 (1.4)	4.7 (1.2)	3.1 (0.8)
High	Low	6.1 (0.9)	6.1 (1.0)	6.3 (1.4)	3.4 (1.1)
High	High	7.2 (0.85)	7.4 (0.9)	7.1 (1.1)	3.8 (1.2)

resisted: (a) reducing blame in the high-status conditions as consequences became less severe, difference score M 's = -0.67 and -0.82 , $t(12) < 1$, *ns*; (b) increasing blaming in the low-status conditions as consequences grew more severe, difference score M 's = -0.32 , $t(12) < 1$, *ns*.

Principal axis factor analysis of all dependent measures yielded the best support yet for a revised two-factor model of the prosecutorial mindset. The first factor accounted for 41% of the variance. Its highest loadings were punitiveness toward individual (.84) and company (.82), dispositional attributions to individual (.85) and company (.80), and deterrence goals (.84). The second factor accounted for 17% of the variance. Its highest loading variables were retribution ($-.53$) and willingness to change one's mind if severity of consequences or perpetrator status were altered (.53 and .59).

Discussion

Experiment 3 went beyond replicating the severity effect by exploring: (a) role-responsibility moderators of the effect; (b) ideological variation in the perceived justifiability of the effect; (c) the activation of general deterrence goals as a mediator of the effect.

It claims too much to say these results decisively favor the FBC over rival theories of reactions to injustice. Most formulations simply do not make predictions here. The data do, however, caution against sweeping generalizations about: (a) attributional favoritism toward higher-status groups. Consistent with Schlenker's (1997) model of responsibility, high status entails role obligations that, if not honored, provide grounds for punitiveness; (b) the greater punitiveness of conservatives. Liberals and conservatives were initially equally susceptible to the severity effect but diverged in their willingness to apply the effect to low and high status perpetrators. Conservatives adhered to the substantially-deterrence-driven severity effect across the board whereas liberals adhered to the effect only when judging the senior executive.

General discussion

Taken together, the three experiments yield a more nuanced portrait of how people function as intuitive prosecutors. The FBC model highlights an array of determinants of how people respond, cognitively and emotionally, to norm violations in two types of settings—between-subjects designs that make it difficult to gauge what is influencing their judgment and repeated-measures designs that make it easy to do so. The FBC also offers a salutary reminder that people subscribe to a variety of ideologically grounded theories of justice that moderate their reactions to norm violations, especially after they have had a chance to reflect on what factors should influence their judgment. Liberals and conservatives within American culture disagree over the appropriateness of allowing factors beyond individual control, such as threat to society or severity of accidents, to increase individual culpability. Experiment I showed that

conservatives were readier to defend tightening standards of accountability when the social order was under siege; Experiment III showed that conservatives were more consistent defenders of cracking down on perpetrators when the accidental consequences of norm violation were severe. Experiment 3 also showed when liberals become more punitive than conservatives: when participants made “change-your-mind” judgments of high-status perpetrators whose negligence had caused severe consequences.

These findings tie into several strands of social psychological work. First, the FBC complements current work on cognitive self-correction: what counts as a bias, something in need of correction when called to one's attention in a repeated-measures design, depends on one's world view. This ideological variation in standards for judging judgment has implications for our understanding: (a) of the status of the over-attribution and severity effects as biases. The pattern of meta-cognitive judgments meshes with the finding that conservative managers are more disposed to defend no-excuses policies as prudent rather than as evidence of a cognitive defects (Tetlock, 2000); (b) of how cognitive self-correction operates in the real world. Accountability-style manipulations designed to make people think harder may deepen, not bridge, ideological divides because people see accountability as a signal to bring their judgment into closer conformity with their own (ideologically charged) conceptions of good judgment (Lerner & Tetlock, 1999). Such fan effect predictions should be tested.

Second, the findings raise locus-of-effect questions that often bedevil social psychological work. It is unclear to what degree ideological differences in perspectives on bias are rooted deeply in clashing views of human trustworthiness (beliefs about how quickly others take advantage of openness to excuses) or in clashing values (the importance of “never being taken for a sucker” versus “never convicting the innocent”) or in the interplay of facts (subjective-probability judgments) and values (utility functions).

Third, the findings shed light on the debate over the emotional-retributive and calculating-deterrence components of punitiveness (Carlsmith et al., 2002). Whereas anger was a key mediator of punitiveness toward intentional norm violation in Experiment 1 and both anger and the correlated desire for retribution played a parallel role in Experiment 2, pragmatic concerns about deterrence drove reactions to tort-like negligence in experiment 3. One possibility, dating back at least to Durkheim (1933), is that retribution and demands for expiation feel like the right response to violations of sacred values (deontic logic) but deterrence and retribution feel like the right response to civil wrongs (utilitarian logic). If so, the FBC model overstates the unity of the prosecutorial mindset. The second and third experiments suggest the usefulness of distinguishing calculating-instrumental forms of the mindset from emotional, value-expressive forms—in effect, a two-factor model of the prosecutorial mindset.

It is also appropriate to comment both on the limits of the methods used here and on the special challenges confronting theorizing in this domain.

Scenario vignettes are well-equipped to answer some theoretical questions, but not others. The independent variables, and surrounding contextual cues, were all linguistic. So too were the response scales. These limitations raise legitimate concerns. One is that the observed power of linguistic stimuli to activate emotion-driven punitive reactions, although significant, under-estimates the power of more vivid real-world events. In this view, the scenario methods constitute unfairly tough tests for the FBC.

Another concern cuts the opposite direction: the linguistic manipulations were too transparent, making it easy for participants to enact whatever roles they suspected the experimenter wanted. In this view, once one sets the semantic stage, one is guaranteed certain results by virtue of the definitions of key terms on the independent or dependent variable sides of the analysis. To invoke the venerable distinction between analytic and synthetic propositions in philosophy of science (a distinction best viewed as a continuum, Quine, 1953), some of our discoveries may be closer to analytical than to synthetic truths, closer to discovering that people define bachelors as unmarried men than to discovering that one ontologically distinctive construct (say, carbon dioxide content of atmosphere) influences another (say, temperature). The best exhibit for this argument is an effect that we have already treated as a manipulation check in Experiment 1: the tendency for attributions of personal responsibility to rise as we delete extenuating and add exacerbating circumstances. One could also argue that the operational definition of ideology in Experiment 1—concern about crime in general—is uncomfortably close to the punitiveness judgments elicited in specific scenarios.

But this objection can be taken too far. It is hard-pressed to explain why between-subjects independent variables, such as societal threat, adequacy of past punishment, and severity of consequences, had their patterns of both main and interactive effects across experiments—and even harder-pressed to explain how these independent variables influenced the culpability judgments of those observers who did not deem it appropriate to be swayed by such factors. If the linguistic-demand argument claims that scenario paradigms can only demonstrate effects that participants were glad to acknowledge openly—as openly as they would acknowledge any other obvious cultural truism—it claims too much.

Shifting to comparisons between the FBC model and other theories, it is easy to get bogged down in questions about exactly how the FBC model differs from the patchwork quilt of partly overlapping and partly distinct, conceptual approaches to attributions of responsibility. Social psychology, arguably, suffers from a surfeit of mini-theories and the FBC model appears, on its face, to add to the problem. Our view is, however, that the FBC fills a key interdisciplinary niche by linking micro and macro theoretical approaches. A comprehensive theory of attribution of responsibility needs to recognize that such processes are shaped not just by the interplay of cognitive-emotional

processes inside the heads of participants but also by the locations of these persons in societies under varying degrees of internal and external strain.

At the macro end of the theory continuum, the FBC has an affinity with sociological, evolutionary-psychology, and behavioral-economic work that portrays individuals as vigilant defenders of the normative order. Unlike these more macro contenders, however, the FBC focuses on specific situational and dispositional triggers of the prosecutorial mindset: societal threat in Experiments 1 and 2; ideology in Experiments 1 and 3; adequacy of punishment in Experiment 2; the exculpation continuum in Experiment 1; and accident severity and perpetrator status in Experiment 3. The FBC also focuses on the complex inter-relations among the observable manifestations of the prosecutorial mindset, such as character attributions, anger, and retribution and deterrence goals.

At the more micro end of the theory continuum, the FBC overlaps with several formulations. Like both Just World and system-justification theory, the FBC assumes that most people think of themselves as fair-minded and of their social order as legitimate. But, unlike system-justification theory, the FBC neither elevates defense of existing hierarchies to the role of master motive nor predicts widespread attributional favoritism toward high-status groups. Rather, the FBC stresses the defense of shared societal values of fair play and predicts a willingness (true of both liberals and conservatives in Experiment 3) to come down especially hard on high-status perpetrators who violate the trust placed in them. Here, the FBC reveals its affinity with Durkheimian theories in sociology, which focus on the solidarity-expressive functions of norm enforcement (when one of us has been victimized by a violation, we have all been victimized). By contrast, system-justification theory reveals its affinity with Marxist theories, which focus on the role of false consciousness in justifying inequality (Alexander, 1987). Put another way, the FBC focuses more on retributive justice whereas system-justification focuses more on distributive justice.

The FBC diverges from Just-World theory in two respects worth testing in future work: (1) the FBC places more emphasis on the efficacy of ritualized, collectively mandated punishment of perpetrators in restoring intrapsychic equilibrium; (2) the FBC views victim derogation as a last-resort coping option, engaged only when affirming social solidarity with victims is either unattractive (because observers subscribe to demeaning stereotypes of victims) or impractical (because perpetrators, for whatever reason, cannot be brought to justice).

At the most micro end of the continuum, the FBC has an affinity with classic legal theories of responsibility (Hart, 1961) and attribution theory (Weiner, 1995): each builds on the premises that the explanations people form for misconduct are potent determinants of their subsequent reactions. But, breaking with the strict intuitive-scientist tradition and in line with Alicke's (2000) culpable control model, the FBC posits a strong proclivity for making

internal-controllable attributions that justify punitive responses to norm violators. The FBC then breaks with the culpable control model by treating this proclivity not as the by-product of flawed cognitive processing (a form of the fundamental attribution error or outcome bias), but as the natural product of activating the prosecutorial goal of defending the social order by inflicting retributive justice on norm violators.

The data thus suggest an intriguing possibility: conservatives better fit the FBC predictions whereas liberals better fit those of the culpable-control model. Judging by responses to change-your-mind questions, conservatives saw nothing wrong with holding norm violators more responsible when the social order was under strain. By contrast, liberals apparently did see something wrong with this practice (although they adopted this practice in between-subjects settings in which it was not obvious what factors they were taking into account). They were unwilling to openly modify their responsibility attributions for norm violators when societal conditions deteriorated in a repeated-measures setting—an unwillingness one would expect if liberals subscribed to the value judgment (embedded in the culpable control model) that it is wrong to become more punitive in response to circumstances extraneous to the violation.

A final comment on functionalism and falsificationism is appropriate. Functionalist explanations are often caricatured as non-falsifiable and the caricature contains kernels of truth. The FBC is merely one of an array of possible middle-range theories of how people function as intuitive prosecutors, some more punitive, some more forgiving, some stressing a cold deterrence calculus, others stressing hot retributive reactions. With that caveat, however, the FBC itself would be falsified by two lines of possible results: (a) repeated failure of manipulations of societal threat, or ideological variation in perceived threat, to induce attributional punitiveness, anger and seeking of retribution; (b) repeated demonstrations that the prosecutorial mindset can be disengaged without inflicting pain on norm violators (programs that stress third-party-provided victim restitution and rendering perpetrators harmless via low-discomfort medical procedures). The exact breaking point for falsification depends, as always, on the patience of the research community.

References

- Alexander, J. (1987). *Twenty lectures: Sociological theory since World War II*. New York: Columbia University Press.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 190–1396.
- Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident. A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin*, *90*, 496–512.
- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish?: Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.
- Cullen, F. T., Fisher, B. S., & Applegate, B. K. (1998). Public opinion about punishment and corrections. In M. Tonry (Ed.), *Crime and justice: An annual review of research* (Vol. 27, pp. 1–79). Chicago: University of Chicago Press.
- Davis, D. W., & Silver, B. D. (2004). Civil liberties vs. security: Public opinion in the context of the terrorist attacks on America. *American Journal of Political Science*, *48*, 28–46.
- Duff, A. (1994). *Punishment: A reader*. New York: Oxford University Press.
- Durkheim, E. (1925/1976). *The elementary forms of the religious life* (2nd ed.). London: Allen and Unwin.
- Durkheim, É. (1933). *The division of labor in society* (George Simpson, Transl.). New York: The Free Press.
- Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Science*, *8*, 187–190.
- Feinberg, J. (1984). *Harm to others: The moral limits of the criminal law*. New York: Oxford University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology*, *29*, 781–795.
- Hamilton, V. L., & Sanders, J. (1992). *Everyday justice: Responsibility and the individual in Japan and the United States*. New Haven: Yale University Press.
- Hart, H. L. A. (1961). *The concept of law*. Oxford: Oxford University Press.
- Hawkins, S., & Hastie, R. (1990). Hindsight: biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.
- Jost, J. T., Banaji, M., & Nosek, B. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–919.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.
- Lerner, M., & Lerner, S. (1978). *The justice motive*. New York: Plenum Press.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255–275.
- Marques, J., Robalo, E., & Rocha, S. (1992). In-group bias and the black sheep effect: Assessing the impact of social identification and perceived variability on group judgments. *European Journal of Social Psychology*, *22*, 331–352.
- McClosky, H., & Brill, A. (1983). *Dimensions of tolerance: What Americans believe about civil liberties*. New York: Russell Sage.
- Miller, D. T., & Vidmar, N. (1981). The social psychology of punishment reactions. In M. J. Lerner & S. C. Lerner (Eds.), *The justice motive in social behavior* (pp. 145–172). New York: Plenum Press.
- Petty, R., & Wegener, D. (1998). Attitude change: multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Vol. Eds.), *The handbook of social psychology* (Vol. 1, *The handbook of social psychology* (Vol. 1, Fourth ed., pp. 323–390). New York: Oxford University Press.
- Quine, W. O. (1953). *From a logical point of view: 9 logico-philosophical essays*. Cambridge, MA: Harvard University Press.
- Schlenker, B. R. (1997). Personal responsibility: Applications of the Triangle Model. In L. L. Cummings & B. Staw (Eds.), *Research in organizational behavior* (Vol. 19, pp. 241–301). Greenwich, Connecticut: JAI Press.
- Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, *28*, 1419–1429.
- Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues*, *42*, 129–150.

- Sunstein, C. R., Schkade, D., & Kahneman, D. (2000). Do people want optimal deterrence? *Journal of Legal Studies*, 29, 237–253.
- Tetlock, P. E. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the political beholder? *Administrative Science Quarterly*, 45, 293–326.
- Tetlock, P. E. (2002). Social-functionalist metaphors for judgment and choice: The Intuitive politician, theologian, and prosecutor. *Psychological Review*, 109, 451–471.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.