Journal of Personality and Social Psychology
2000, Vol. 78, No. 5, 853–870

0022-3514/00/$5.00 DOI: 10.1037//0022-3514.78.5.853

The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals

Philip E. Tetlock, Orie V. Kristel, S. Beth Elson, and Melanie C. Green
Ohio State University

Jennifer S. Lerner
Carnegie Mellon University

Five studies explored cognitive, affective, and behavioral responses to proscribed forms of social cognition. Experiments 1 and 2 revealed that people responded to taboo trade-offs that monetized sacred values with moral outrage and cleansing. Experiments 3 and 4 revealed that racial egalitarians were least likely to use, and angriest at those who did use, race-tainted base rates and that egalitarians who inadvertently used such base rates tried to reaffirm their fair-mindedness. Experiment 5 revealed that Christian fundamentalists were most likely to reject heretical counterfactuals that applied everyday causal schemata to Biblical narratives and to engage in moral cleansing after merely contemplating such possibilities. Although the results fit the sacred-value-protection model (SVPM) better than rival formulations, the SVPM must draw on cross-cultural taxonomies of relational schemata to specify normative boundaries on thought.

Research on social cognition ultimately rests on functionalist assumptions about what people are trying to accomplish when they judge events or make choices. The most influential of these assumptions have been the intuitive scientist and the intuitive economist. The former tradition depicts people whose central objective is to understand underlying patterns of causality, thereby conferring some advantage in anticipating life-enhancing or threatening events (cf. Kelley, 1967). The latter tradition depicts people as decision makers whose overriding goal is to select utility-maximizing options from available choice sets (Becker, 1981; Kahneman & Tversky, 1979). Although theorists often disagree sharply over how well people live up to the high professional ideals of science or economics (Mellers, Schwartz, & Cooke, 1998), theorists agree in placing a normative premium on intellectual flexibility and agility. Good intuitive scientists and economists look for the most useful cues in the environment for generating accurate predictions and making satisfying decisions and quickly abandon hypotheses that do not "pan out." Rigidity is maladaptive within both frameworks.

In this article, we explore the empirical implications of an underexplored starting point for inquiry: the notion that, in many contexts, people are striving to achieve neither epistemic nor utilitarian goals, but rather, as prominent historical sociologists have argued (Bell, 1976), are struggling to protect sacred values from secular encroachments by increasingly powerful societal trends toward market capitalism (and the attendant pressure to render everything fungible) and scientific naturalism (and the attendant pressure to pursue inquiry wherever it logically leads). A sacred value can be defined as any value that a moral community implicitly or explicitly treats as possessing infinite or transcendental significance that precludes comparisons, trade-offs, or indeed any other mingling with bounded or secular values.[1] When sacred values are under assault, the apposite functionalist metaphor quickly becomes the intuitive moralist–theologian metaphor,[2] which depicts people engaged in a continual struggle to protect their private selves and public identities from moral contamination by impure thoughts and deeds (Belk, Wallendorf, & Sherry, 1989). The most emphatic ways to distance oneself from normative transgressions are by (a) expressing *moral outrage*—a composite psychological state that subsumes cognitive reactions (harsh char-

Philip E. Tetlock, Orie V. Kristel, S. Beth Elson, and Melanie C. Green, Department of Psychology, Ohio State University; Jennifer S. Lerner, Department of Decision Sciences, Carnegie Mellon University.

We acknowledge the support of National Science Foundation Grant BSR 9505680 as well as that of the Institute of Personality and Social Research at the University of California, Berkeley, and the Mershon Center of Ohio State University. We also appreciate the helpful advice of Alan Fiske, Danny Kahneman, and Barbara Mellers and the research and clerical assistance of Sara Bassett, Rachel Szteiter, and Megan Berkowitz.

Correspondence concerning this article should be addressed to Philip E. Tetlock, Department of Psychology, 142 Townshend Hall, 1885 Neil Avenue, Ohio State University, Columbus, Ohio 43210-1222. Electronic mail may be sent to tetlock.1@osu.edu.

[1] It should be stressed that the declaratory policy of a moral community toward a sacred value represents an expressed preference, not a revealed preference. As many economists would point out, the actual choices people make may belie high-sounding proclamations that the sacred value is assigned infinite weight.

[2] Sacred values are often ultimately religious in character, but they need not have divine sanction (hence our hybrid designation of the functionalist metaphor as moralist–theologian). Sacred values can range from fundamentalists' faith in God to the liberal–social democratic dogma of racial equality to the radical libertarian commitment to the autonomy of the individual. Although the theoretical framework proposed here does not differentiate sacred values with or without divine mandate, many writers, from Samuel Johnson to Fyodor Dostoyevsky to T. S. Eliot, have drawn sharp distinctions here and have even suggested that only sacred values anchored in faith in God can sustain genuine moral outrage and cleansing. To paraphrase Dostoyevsky, if there were no God, no act, not even cannibalism, would be forbidden.

acter attributions to those who endorse the proscribed thoughts and even to those who do not endorse, but do tolerate, this way of thinking in others), affective reactions (anger and contempt for those who endorse the proscribed thoughts), and behavioral reactions (support for ostracizing and punishing deviant thinkers); and (b) engaging in *moral cleansing* that reaffirms core values and loyalties by acting in ways that shore up those aspects of the moral order that have been undercut by the transgression. Within this framework, rigidity, accompanied by righteous indignation and by blanket refusal even to contemplate certain thoughts, can be commendable—indeed, it is essential for resolutely reasserting the identification of self with the collective moral order (cf. Durkheim, 1925/1976). What looks irrationally obdurate within the intuitive scientist and economist research programs can often be plausibly construed as the principled defense of sacred values within the moralist–theologian research program (Tetlock, 1999).

In this article, we identify three types of normative proscriptions—taboo trade-offs, forbidden base rates, and heretical counterfactuals—that people consciously or unconsciously impose on cognitive processes that are fundamental to rationality in the intuitive scientist and economist traditions. Here we consider each proscription in turn.

Taboo Trade-Offs

Trade-off reasoning is widely viewed as a minimal prerequisite for economic rationality (Becker, 1981). Utility maximization presupposes that people routinely factor reality constraints into their deliberations and explicitly weigh conflicting values. Indeed, economic survival in competitive markets requires that people make at least implicit trade-offs between objectives such as work versus leisure, saving versus consumption, and consumption of alternative products. The moralist–theologian metaphor warns of sharp resistance to efforts to translate all values into a common utility metric. Fiske and Tetlock (1997) documented that, in most cultures, people are chronic "compartmentalizers" who deem some trade-offs legitimate (goods and services routinely subject to market-pricing rules) but vehemently reject others—in particular, those that treat "sacred values" like honor, love, justice, and life as fungible.

This sharp resistance is rooted, in part, in the familiar incommensurability problem. Decision theorists have long stressed that people find interdimensional comparisons cognitively difficult and resort to noncompensatory choice heuristics such as elimination-by-aspects to avoid them (Payne, Bettman, & Johnson, 1992). The moralist–theologian framework, however, treats this explanation as incomplete. Apple–orange comparisons are difficult, but people often make them when they go to the supermarket. Moreover, people do not find it shameful to make trade-offs between money and consumption goods. The moralist–theologian framework traces opposition to reducing all values to a single utility metric to a deeper, more intractable form of incommensurability: constitutive incommensurability, a pivotal concept in modern moral philosophy (Raz, 1986) as well as in classic sociological theory (Durkheim, 1925/1976). As Tetlock, Peterson, and Lerner (1996) argued, the guiding idea is that our commitments to other people require us to deny that we can compare certain things—in particular, things of finite value with things that we are normatively obligated to treat as infinitely important. To transgress this boundary, to attach a monetary value to one's friendships, children, or loyalty to one's country, is to disqualify oneself from the accompanying social roles. Constitutive incommensurability can thus be said to exist whenever comparing values subverts one of the values (the putatively infinitely significant value) in the trade-off calculus. Taboo trade-offs are, in this sense, morally corrosive: The longer one contemplates indecent proposals, the more irreparably one compromises one's moral identity. To compare is to destroy.

Forbidden Base Rates

We find just as solid a normative consensus that good intuitive scientists and/or statisticians should use base rates as that good intuitive economists should confront trade-offs. Decision theorists routinely invoke Bayes' theorem as the appropriate principle for aggregating base-rate and case-specific information (cf. Fischhoff and Beyth-Marom, 1983). We also find considerable consensus that people often deviate from Bayesian prescriptions and ignore base rates. For many years, the base-rate fallacy, with its compellingly counterintuitive demonstrations such as the lawyer–engineer problem, has been regularly trotted out in influential textbooks as a lead exhibit in the case for human irrationality (e.g., Myers, 1993). The standard explanation has been that people make subjective-likelihood judgements by relying on simple error-prone heuristics such as representativeness, in which judgments about the probability of category membership hinge entirely on the perceived similarities of the target to the defining features of the category (Kahneman & Tversky, 1972).

The base-rate literature is both enormous and enormously controversial (Koehler, 1996). Our goal is not, however, just to add to the already formidable list of moderators of whether, and to what degree, people use base rates. Rather, it is to demonstrate that relying on error-prone heuristics is not the only pathway to base-rate neglect. In many contexts, accuracy is neither the only nor even the primary standard for evaluating quality of judgment. A classic example is the U.S. legal system in which procedural justice trumps judgmental accuracy whenever, as often occurs, diagnostic evidence is excluded from trial. Indeed, in exactly this vein, prominent legal theorists have proposed that base-rate evidence is fundamentally inconsistent with the legal ideal of individual justice and should be categorically excluded (Tribe, 1971).

Forbidden base rates refer to any statistical generalization that devoted Bayesians would not hesitate to enter into their probability calculations but that deeply offends a religious or political community. The primary obstacle to using the putatively relevant base rate is not cognitive, but moral. In a society committed to racial, ethnic, and gender egalitarianism, forbidden base rates include observations bearing on the disproportionately high crime rates and low educational test scores of certain categories of human beings. Putting the accuracy and interpretation of such generalizations to the side, people who use these base rates in judging individuals are less likely to be applauded for their skills as good intuitive statisticians than they are to be condemned for their racial and gender insensitivity.

Heretical Counterfactuals

These propositions take the form of assertions about historical causality (framed as subjunctive conditionals with false anteced-

ents) that pass conventional cognitive tests of plausibility but that many people greet with indignation because the assertions subvert a core tenet of their religious belief systems. In Kahneman and Miller's (1986) norm theory and, more generally, in the extensive philosophical literature on what could or might have been in history (Tetlock & Belkin, 1996), there is wide agreement that compelling counterfactuals should pass such tests as "imaginability of the antecedent" and "soundness of antecedent-consequent linkages." Claims such as "if Hitler had perished as a foot soldier in World War I, there would have been no Nazi regime" rise or fall in credibility as a function of whether listeners can easily imagine the antecedent occurring in the actual world and of whether listeners possess causal schemata that specify alternative pathways to Nazism.

The moralist–theologian framework posits that cognitive theories of counterfactual reasoning need to acknowledge the emotionally charged normative boundaries that religious and political movements erect against what–if speculation. Particularly irksome are counterfactuals that apply normal laws of human nature and of physical causality to heroic founders of the movement. Consider the reaction of the Ayatollah Khomeini to Salmon Rushdie's heretical counterfactual in *Satanic Verses* that invited readers to imagine that the Prophet Mohammed kept the company of prostitutes. For this transgression, the theocratic regime in Iran sentenced Rushdie to death (the ultimate expression of moral outrage).

Within the Christian faith in the modern era, such theological ferocity is rare, but it is not difficult to identify counterfactuals that strike the faithful as bizarre or repugnant. Classic examples include counterfactual conjectures that undermine the faith in the "unique historicity" of Jesus Christ (Buckley, 1997)—the view that Jesus was God made man, divine yet also human, that he was born to a virgin Mary, that he died to atone for humanity's sins, and that the events of his life as revealed in the New Testament gospels were the product of a divine plan and hence shielded from the random contingencies that distort the lives of ordinary mortals. From a fundamentalist perspective, the life of Christ had to unfold as it did and devout believers should react indignantly to counterfactuals such as the following that imply otherwise: "If Joseph had left Mary because he did not believe she had conceived a child with the Holy Ghost, Jesus would have grown up in a one-parent household and formed a different personality." From a secular point of view, though, such counterfactuals are eminently reasonable. They introduce schematic chains of causal propositions—in Abelson's (1981) terms, "scripts"—that virtually all of us apply reflexively in everyday life to a text that many of us deem divinely inspired.

Sacred-Value-Protection Model (SVPM)

This article has two guiding objectives, one conceptual and one empirical. The conceptual objective is to move beyond abstract metaphorical posturing and to articulate a testable middle-range theory of how people function as intuitive moralists–theologians. In principle, many middle-range theories could serve this role. Just as we now have a host of middle-range theories of people as intuitive scientists and economists that vary (among other things) on a rationality continuum, so it is easy to imagine that we could have a host of theories of people as intuitive moralists–theologians that vary on a ferocity–forgiveness continuum—a continuum that could be personified at one end by Torquemada of the Spanish Inquisition and at the other end by open-minded and compassionate 20th century Judaeo-Christian thinkers such as Archbishop Tutu. But it is necessary to start somewhere, and our point of departure is the SVPM (Tetlock, 1999). The SVPM initially made no "content" assumptions about what people deem to be sacred, but it did make strong motivational and process assumptions about how people cope with threats to sacred values. Key hypotheses focus on two coping strategies, moral outrage and moral cleansing.

Moral Outrage

Building on Durkheim's (1925/1976) classic observations of how people respond to affronts to the collective conscience that disturb the normative equilibrium of society, the SVPM predicts that when observers believe that decision makers have entertained proscribed thoughts, they will respond with moral outrage, which has cognitive, affective, and behavioral components: lower thresholds for making harsh dispositional attributions to norm violators; anger, contempt, and even disgust toward violators; and enthusiastic support for both norm enforcement (punishing violators) and metanorm enforcement (punishing those who shirk the burdensome chore of punishing deviants; cf. Coleman, 1991). Pursuing the logic of constitutive incommensurability (to compare is to destroy), the model also postulates that the longer observers believe that decision makers contemplated compromising sacred values, even if they ultimately do the right thing and support sacred values, the more intense the outrage they direct at those decision makers.

Moral Cleansing

Revealing its kinship with self-affirmation variants of dissonance theory (Steele, 1988) and social identity theory (Schlenker, 1982), the SVPM predicts that decision makers themselves will feel at some level of consciousness tainted by merely contemplating taboo trade-offs, forbidden base rates, and heretical counterfactuals and will engage in symbolic acts of moral cleansing designed to reaffirm their solidarity with their moral community. The SVPM deviates from virtually all variants of dissonance theory, however, in four key ways. First, the SVPM predicts a "mere contemplation effect": It is not necessary to commit a counternormative act; it is sufficient for counternormative thoughts to flicker briefly through consciousness prior to rejecting them. That brief prerejection interval, during which our natural first reaction to propositions is apparently to consent (Gilbert, 1991), can produce a subjective sense—however unjustified—that one has been cognitively contaminated and has fallen from moral grace in the community. Second, the logic of constitutive incommensurability dictates that the longer one contemplates taboo-breaching proposals, the greater the subjective contamination and estrangement from the collective. Unlike dissonance theory, which focuses solely on the intrapsychic function of maintaining mental equilibrium (original Festingerian emphasis) or of protecting the self-image (the emphasis in revisionist self-oriented variants of dissonance; see, e.g., Greenwald & Ronis, 1978), the SVPM assigns a double-barreled functional role to outrage and cleansing: an intrapsychic-expressive function in which the goal is to convince oneself of one's moral worthiness and an interpersonal-instrumental function in which the goal is to shore up the external

moral order. Third, and closely related, the SVPM stresses the close symbolic connections between the breach in the moral order and the norm-defending outrage and the norm-exemplifying cleansing responses. When the defensive perimeter of the moral order begins to crumble, priority should go to sealing the breach, not to strengthening those parts of the perimeter that remain strong (cf. Stone, Wiegand, Cooper, & Aronson, 1997). By contrast, Steele's (1988) self-affirmation variant of dissonance theory maintains that the connection between identity-damaging acts and identity-restoration tactics is much looser and that a wide range of self-enhancing affirmations can mitigate the dissonance created by counterattitudinal acts. Fourth, although dissonance and self-esteem researchers frequently find substitutability among coping responses to ego threat (Simon, Greenberg, & Brehm, 1995; Stone et al., 1997; Tesser & Cornell, 1991; but also see Aronson, Blanton, & Cooper, 1995), the SVPM allows for both compensatory and overkill relationships between outrage and cleansing responses to threats to sacred values. A subset of the experiments deploy question-ordering manipulations to explore these two possibilities: (a) the compensatory hypothesis that, once people have had an opportunity to distance themselves from proscribed cognitions by means of either moral outrage or cleansing, they need to do nothing else; and (b) the overkill hypothesis that people often rely on multiple, seemingly redundant, strategies of distancing themselves from proscribed cognitions.

Experiments 1 and 2: Taboo Trade-Offs

In Experiment 1, we explored the reactions of a broad spectrum of political activists to routine or secular–secular trade-offs (money for goods and services legally exchanged in the market economy of late 20th century America) and taboo or secular–sacred trade-offs (money for goods and services that cannot legally be bought or sold in late twentieth century America). Tetlock et al. (1996) hypothesized that what counts as a taboo trade-off should vary dramatically across ideological subcultures and historical periods. Free-market libertarians should be most inclined to allow individuals to enter into whatever contractual understandings they wish—be it buying or selling lettuce or votes, newspapers or body organs, or future options for commodities or adoption rights for children. Their wrath will be reserved for those meddlesome souls who invent moral externalities (adverse effects on third parties) designed to justify constraining consenting adults from making trade-offs and agreements that each contracting party agrees leaves him or her better off. By contrast, Marxists will be most offended. They will object not only to proposals to render sacred values fungible, but even to the exploitative character of many routine market transactions in American society. Finally, in the broad middle of American political spectrum, there should be considerable consensus on what is a taboo trade-off. Conservative Republicans and liberal Democrats should agree on most items in Experiment 1: Human body organs and babies, and basic rights and responsibilities of democratic citizenship all fall outside boundaries of the fungible, whereas cars, houses, and the services of gardeners all fall within the domain of the fungible. Still, disagreements should erupt. Liberals may object that market pricing of medical and legal services effectively assigns dollar values to life and justice, whereas conservatives may view such transactions with casual equanimity.

Experiment 2 differed from Experiment 1 in several key respects. No special effort was made to sample political extremists. And the focal comparison shifted from one between routine and taboo trade-offs to one between taboo trade-offs (pitting secular against sacred values as in money vs. lives) and tragic trade-offs (pitting sacred against sacred values such as one life vs. another). The central hypothesis derived from the constitutive-incommensurability postulate of the SVPM is the longer observers believe a decision maker considered a taboo trade-off, the more punitively they will judge that decision maker, even if, in the end, the decision maker does what most people consider to be the "right thing" and affirms the sacred value. By contrast, the longer observers believe that a decision maker considered a tragic trade-off, the wiser and more judicious observers will deem the decision maker, regardless of the outcome of the decision. Lengthy deliberation on tragic trade-offs reaffirms the solemnity of the occasion and the transcendent significance of the competing sacred values; lengthy deliberation on taboo trade-offs exacerbates the transgression of weighing a sacred value on a secular scale.

Method: Experiment 1

Participants. Between 1991 and 1994, a sample of 127 undergraduates was recruited from campus political organizations that spanned the political spectrum from the Libertarian Party (and an affiliated Rand–Hayek Study Group), the Republican Party, the Democratic Party, and the Socialist Workers Party (and an affiliated Marxist group, the Spartacist Youth League). From this initial sample, it was possible to identify ideologically coherent and consistent advocates of the four political factions designated earlier: libertarian (n = 12), mainstream liberal (n = 34), mainstream conservative (n = 30), and Marxist socialist (n = 14). Group membership was a necessary but not sufficient condition for ideological classification, which was convergently validated against responses to questions designed to differentiate the groups. To qualify as Marxist socialists, respondents also had to endorse public control of the economy as well as a radical leveling of incomes; to qualify as liberals, respondents had to disagree with the socialist items but to endorse a moderate leveling of incomes by means of progressive tax rates and to support guaranteed access to medical care; to qualify as conservatives, respondents had to disagree with the liberal sentiments but to agree that government regulations on business are excessive and to endorse some restrictions on abortion; to qualify as libertarians, respondents had to agree that regulations on business are excessive but to reject any state role in redistributing income and to reject state interference not only in abortion but also in personal decisions to use marijuana or to engage in any form of consensual sex.

Assessing reactions to value trade-offs. Participants were told that the goal of the study was to explore the attitudes that Americans have about what people should be allowed to buy and sell in competitive market transactions:

> Imagine that you had the power to judge the permissibility and morality of each transaction listed below. Would you allow people to enter into certain types of deals? Do you morally approve or disapprove of those deals? And what emotional reactions, if any, do these proposals trigger in you?

Respondents then judged two types of trade-offs: routine (secular–secular) and taboo (secular–sacred). The five secular–secular trade-offs included "paying someone to clean my house," "buying a house," "buying food," "paying a doctor to provide medical care to me or my family," and "paying a lawyer to defend me against criminal charges in court." The nine secular–sacred trade-offs included buying and selling of human body parts for medical transplant operations, surrogate motherhood contracts (paying

someone to have a baby whom the buyer subsequently raises), adoption rights for orphans, votes in elections for political offices, the right to become a U.S. citizen, the right to a jury trial, sexual favors (prostitution), someone else to serve jail time to which the buyer had been sentenced by a court of law, and paying someone to perform military service that the buyer had a draft obligation to perform.

For each activity, respondents made the following judgments on 7-point scales, anchored at 1 and 7: *should be banned–should be permitted* (midpoint: *permitted with major restrictions*), *highly moral–highly immoral* (midpoint: *unsure*), *highly upsetting–not at all upsetting* (midpoint: *moderately upsetting*), *not at all sad–extremely sad* (midpoint: *moderately sad*), *not at all tragic–tragic* (midpoint: *moderately tragic*), *not at all offensive–highly offensive* (midpoint: *moderately offensive*), *no anger–great deal of anger* (with the midpoint: *angers me somewhat*). Respondents also rated what they thought of someone willing to permit this type of trade-off: *very irrational–very rational* (midpoint: *neutral*), *very compassionate–very cruel* (midpoint: *neutral*), and *completely crazy–completely sane* (midpoint: *neutral*), and how they would react if: (a) They were asked in ordinary conversation about their views on the subject (*I'd be deeply insulted–it would not bother me at all to be asked* and *I would want to end the conversation quickly–I would want to continue the conversation*); (b) an elected member of the student government refused to oppose funding for a campus group that had invited a speaker who favors a ballot proposition that "would treat children without parents like commodities that could be sold to the highest responsible bidder" (*very negative–very positive;* midpoint: *neutral*).

All respondents were given a moral-cleansing opportunity to express behavioral intentions that affirmed their commitment to insulating a sacred value from monetary encroachments: They were asked on a 7-point scale (*not at all interested–extremely enthusiastic;* midpoint: *unsure*) how willing they were to volunteer to help a political-action group fighting to prevent passage of a (fictitious) ballot proposition that would legalize the buying and selling of adoption rights for children in need of parents. Half the respondents in each ideological group answered this question prior to examining and evaluating the list of hypothesized taboo trade-offs, and the other half answered this question after doing so. Insofar as merely contemplating taboo trade-offs is morally contaminating, participants in the "after" condition should express stronger intentions to engage in moral cleansing.

Results: Experiment 1

Constructing the moral-outrage index. The hypothesized cognitive components of moral outrage (attributions of cruelty, irrationality, and insanity) were positively correlated with each other (average $r = .58$) just as the affective components were with each other (angry, upset, insulted by any implication that one might endorse a taboo trade-off; average $r = .41$). The aggregated cognitive and affective components were also correlated with each other ($r = .52$) as well as with desire to ban market exchanges that embody taboo trade-offs ($rs = .65$ and $.59$), with punitive behavioral reactions to people who endorse taboo trade-offs (desire to sever contact, $r = .35$ and $.39$), and with willingness to punish those who fail to punish violations of taboo trade-offs (metanorm enforcement, $r = .29$ and $.36$). To simplify analysis, we created a composite moral-outrage index by subjecting these correlations to maximum-likelihood factor analysis (oblimin rotation) and deriving scores for each respondent on the first factor which, judging from the rotated factor loadings, captured each component of moral outrage: Negative Affect (e.g., anger), Dispositional Attributions (e.g., irrational, cruel), and Sanctioning (e.g., desire to sever contact). Participants' scores on the Outrage factor were computed by summing scores on all high-loading (greater than .3) factors and averaging.

Analyses of variance. A 4 (ideological faction) × 2 (timing-of-cleansing measure) × 2 (repeated measure: routine trade-off vs. taboo trade-off) analysis of variance (ANOVA) assessed effects on moral outrage. As Figure 1 indicates, a main effect emerged: taboo trade-offs elicited far greater outrage ($M = 4.48$) than did routine trade-offs ($M = 2.68$), $F(1, 73) = 26.32$, $p < .001$. The hypothesized interaction between ideology and trade-off status also emerged. Taboo trade-offs triggered outrage from liberal Democrats, conservative Republicans, and radical socialists but scarcely a flicker of annoyance from libertarians, $F(1, 73) = 23.74$, $p < .001$. The differences among ideological groups fell to nonsignificance, however, for routine trade-offs that evoked minimal outrage, with two notable exceptions: (a) Socialists were more offended by routine trade-offs than all other groups, $F(1, 73) = 4.61$, $p < .05$; (b) liberals were more offended by two of the five secular–secular trade-offs (buying and selling medical and legal services) than were conservatives, $F(1, 73) = 5.05$, $p < .05$. Timing of the moral-cleansing measure had no main or interactive effects on moral outrage directed at taboo trade-offs.

A 4 × 2 ANOVA assessed effects on moral cleansing (volunteering for a campaign to block baby auctions). Moral cleansing was more pronounced among conservatives ($M = 4.97$), liberals ($M = 5.02$), and socialists ($M = 5.39$) than among libertarians ($M = 2.10$). This ideology effect held up, moreover, regardless of whether moral cleansing was assessed before or after exposure to the taboo trade-offs—an unsurprising result in view of the sharp opposition across the nonlibertarian groups to auctioning babies, $F(1, 73) = 18.06$, $p < .01$. The timing manipulation (whether respondents received the request to join the campaign against the baby-auctioning ballot initiative before or after judging taboo trade-offs) had no main effect on moral cleansing. The predicted Ideology × Timing interaction did, however, emerge. The two mainstream groups (liberal Democrats and conservative Republicans) expressed stronger desires to stop baby auctions when they were first exposed to the taboo trade-offs and then asked to help ($Ms = 5.68$ and 5.46) as opposed to first being asked to help and then contemplating the taboo trade-offs ($Ms = 4.35$ and 4.48). By contrast, the order manipulation had no effect on the two relatively extreme groups, libertarians (who thought baby auctions to be a good idea, $Ms = 2.08$ and 2.12) and Marxists (who found even many routine trade-offs distasteful; $Ms = 5.31$ and 5.47). As a result, the timing effect was significantly greater among the mainstream groups than among the relatively extreme groups, planned contrast, $F(1, 73) = 6.09$, $p < .05$. Finally, although the correlation between moral cleansing and outrage was nonsignificant when cleansing was assessed prior to contemplating the taboo trade-offs, $r(43) = .06$, the same correlation became significant, $r(44) = .44$, when cleansing was assessed after people had contemplated and been outraged by taboo trade-offs.

Method: Experiment 2

A total of 228 participants were presented with a health-care decision-making questionnaire that contained one of eight versions of the following scenario, generated by a 2 (taboo–tragic trade-off) × 2 (length of deliberations) × 2 (saving or not saving "Johnny") factorial. Robert, the key decision maker, was described as the Director of Health Care Management

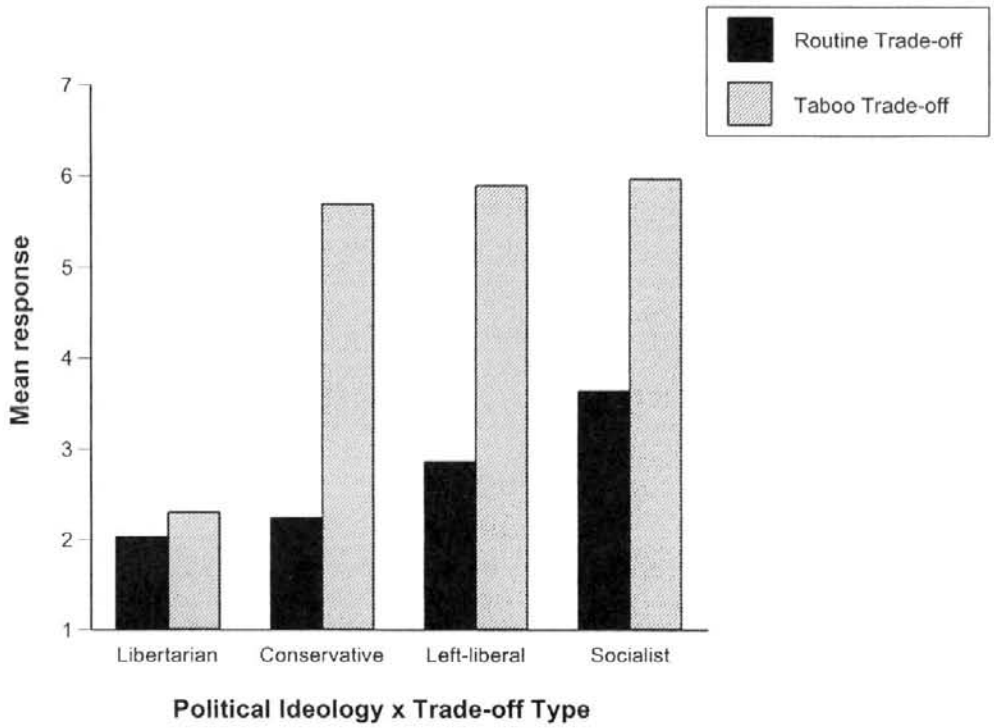

Figure 1. Average moral outrage as a function of political ideology and routine and taboo trade-offs (Experiment 1).

at a major hospital who confronted a "resource allocation decision." At this point, the experimental manipulation of taboo versus tragic trade-offs was introduced. The tragic trade-off was

> Robert can either save the life of Johnny, a five year old boy who needs a liver transplant, or he can save the life of an equally sick six year old boy who needs a liver transplant. Both boys are desperately ill and have been on the waiting list for a transplant but because of the shortage of local organ donors, only one liver is available. Robert will only be able to save one child.

The taboo-tradeoff was

> Robert can save the life of Johnny, a five year old who needs a liver transplant, but the transplant procedure will cost the hospital $1,000,000 that could be spent in other ways, such as purchasing better equipment and enhancing salaries to recruit talented doctors to the hospital. Johnny is very ill and has been on the waiting list for a transplant but because of the shortage of local organ donors, obtaining a liver will be expensive. Robert could save Johnny's life, or he could use the $1,000,000 for other hospital needs.

The second independent variable, the speed and ease with which Robert made the decision, was always inserted immediately after the characterization of the problem: "Robert sees his decision as an easy one, and is able to decide quickly," or "Robert finds this decision very difficult, and is only able to make it after much time, thought, and contemplation." The third independent variable, whether Robert decided to save Johnny's life, was always introduced immediately after the information on how quickly Robert made the decision (in the tragic–trade-off condition, either Johnny or the other child was saved; in the taboo–trade-off condition, either Johnny was saved or the money was directed to other hospital functions).

Dependent variables involved rating Robert's decision (7-point scales on *bad–good, wise–foolish, positive–negative,* and *moral–immoral*) and feelings about the decision (*fair–unfair, not at all disgusted–disgusted, excited–upset,* and *sad–happy*). Participants also rated on 7-point scales whether they agreed that "Robert should be removed from his job" and that "Robert should not be punished for his decision." Finally, participants were asked "If Robert was a friend of mine, and I knew the decision he made, I would end the friendship over this issue" (7-point scales on, *agree–disagree*) and whether they would be willing "to volunteer some of their time to aid a city campaign to increase organ donations" (7-point scale to assess moral cleansing).

Results: Experiment 2

Constructing the moral-outrage and punitive-interpersonal stance indexes. Maximum-likelihood factor analysis (with oblimin rotation) was used to combine scales into a composite index. Although examination of a scree plot and fit measures indicated a three-factor solution (root-mean-square error of the approximation [RMSEA] = .06), only two clearly interpretable factors emerged: an Outrage factor (loadings greater than .3 included: bad, foolish, negative, immoral, unfair, and disgust) and a Punitive Stance factor (loadings greater than .3 included: dismiss from job, should be punished, end friendship). Both the Moral-Outrage and the Interpersonal-Punitiveness Scales showed good reliability (Cronbach's α = .96 and .73 respectively) and were positively correlated (r = .63).

Moral-outrage effects. Table 1 presents average responses across conditions. A three-way ANOVA revealed a main effect of outcome. Observers directed less outrage at the decision maker who saved Johnny rather than directing money to other hospital functions (Ms = 3.09 vs. 4.58), $F(1, 220) = 84.44$, $p < .01$. A two-way interaction revealed that a decision maker contemplating

Table 1
Mean Levels of Outrage, Sanctioning, and Moral Cleansing as a Function of Trade-Off Type, Ease and Speed of Decision, and Outcome (Experiment 2)

Experimental condition	Dependent variables		
	Outrage	Sanctioning	Moral cleansing
Taboo tradeoff			
Difficult decision			
1. Hospital	$5.71_{2,4-8}$	4.13_{2-8}	$5.36_{4-6,8}$
2. Johnny	$3.43_{1,3,4,7,8}$	$3.13_{1,4-6}$	$5.31_{4-6,8}$
Easy decision			
3. Hospital	$5.14_{2,4-8}$	$3.35_{1,4-6}$	4.78
4. Johnny	$1.51_{1-3,5-8}$	$1.66_{1-3,7,8}$	$4.17_{1,2}$
Tragic tradeoff			
Difficult decision			
5. Other child	$3.33_{1,3,4,7,8}$	$2.24_{1-3,7,8}$	$3.97_{1,2}$
6. Johnny	$3.05_{1,3,4,7,8}$	$1.92_{1-3,7,8}$	$3.77_{1,2}$
Easy decision			
7. Other child	4.31_{1-6}	$3.37_{1,4-6}$	4.45
8. Johnny	4.43_{1-6}	$3.29_{1,4-6}$	$4.10_{1,2}$

Note. Range = 1 to 7, with higher levels indicating greater outrage, sanctioning, and cleansing. Subscripts for each mean indicate the row numbers of those means that are significantly different from that mean (LSD test, $p < .05$).

a taboo trade-off evoked much outrage if he failed to save Johnny ($M = 5.41$) and little outrage if he did save Johnny ($M = 2.45$), whereas the decision maker contemplating a tragic trade-off evoked low to moderate outrage, regardless of whether he saved Johnny ($M = 3.72$) or the other child ($M = 3.81$), $F(1, 220) = 75.58$, $p < .001$. An additional two-way interaction emerged between trade-off and ease of decision, $F(1, 220) = 53.81$, $p < .001$. For taboo trade-offs, decision makers were met with greater outrage if the decision had been difficult ($M = 4.48$) rather than easy ($M = 3.23$), whereas the reverse pattern held for tragic trade-offs (M easy = 4.37, M difficult = 3.18). These effects were qualified by a three-way interaction, $F(1, 220) = 7.11$, $p < .01$. The administrator in the taboo condition who chose Johnny quickly was judged least negatively ($M = 1.51$), whereas the administrator in the taboo condition who chose slowly and chose the hospital ($M = 5.71$) was judged most negatively. The tragic–trade-off decision maker who decided slowly was judged more positively than when he made up his mind quickly, regardless of selection (M tragic, difficult and slow = 3.18; M tragic, easy and quick = 4.37).

Interpersonal-punitiveness. As Table 1 indicates, similar patterns emerged for the sanctioning index, including a main effect for outcome, $F(1, 220) = 17.89$, $p < .01$, and two-way interactions between trade-off and speed or ease of decision process, $F(1, 220) = 42.12$, $p < .01$, and trade-off and decision outcome, $F(1, 220) = 9.87$, $p < .01$. Additionally, a main effect of trade-off emerged, $F(1, 220) = 3.87$, $p = .05$. Decision makers facing taboo trade-offs ($M = 3.01$) were punished more than those facing tragic trade-offs ($M = 2.67$). A planned contrast revealed the greatest sanctioning when the decision maker in the taboo condition required a long time to make up his mind and wound up affirming the secular value (hospital salaries and/or infrastructure) over the sacred value (Johnny's life), a mean different from all seven other means, $t(220) = 4.88$, $p < .001$. As Table 1 shows, sanctioning reached its nadir when the decision maker resolved the taboo trade-off quickly in favor of the sacred value, a condition mean significantly different from all other means, but not significantly different from the two conditions in which the tragic–trade-off decision maker thought long and hard about the choice.

Moral cleansing. This variable was, as in Study 1, positively correlated with both moral outrage, $r(228) = .21$, $p < .01$, and sanctioning, $r(228) = .32$, $p < .01$. As Table 1 indicates, a main effect of trade-off type emerged, $F(1, 220) = 8.60$, $p < .05$. Participants who read about a taboo trade-off were more likely to volunteer for the organ-donation campaign than were those who read about a tragic trade-off (M taboo = 4.88, M tragic = 4.06). The two-way interaction between trade-off and ease was also significant, $F(1, 220) = 5.00$, $p < .05$; decision makers faced with a taboo trade-off inspired more cleansing if the decision was difficult ($M = 5.33$) rather than easy ($M = 4.46$), whereas the reverse was true for decision makers faced with a tragic choice (M difficult = 3.87, M easy = 4.27). Planned contrasts revealed a surge in moral cleansing in the two conditions in which decision makers thought long and hard about a taboo trade-off and either affirmed the sacred value or allowed the secular value to trump the sacred value, $t(220) = 4.61$. Post hoc (least significant difference) tests revealed that these two conditions were different from all other conditions but two: when the decision maker contemplated the taboo trade-off only briefly and chose hospital salaries and when the decision maker contemplated the tragic trade-off briefly and chose the other child.

Discussion: Experiments 1 and 2

Why are some trade-offs regarded as so routine that people are baffled that anyone should even bother to ask about them whereas other trade-offs are so controversial that people react with scorn to the mere posing of the question? It explains little just to invoke "culture and socialization." We gain more explanatory leverage, however, by joining Alan Fiske's (1991) theory of relational schemata to the SVPM. Within relational theory, people treat a trade-off as taboo to the degree it inappropriately extends a market-pricing schema into domains that are normatively regulated by one of three alternative schemata: communal sharing, authority ranking, or equality matching. Caring for children is regarded a communal-sharing responsibility of families; obligations to perform military service derive from the legitimate authority-ranking prerogatives of the legal system; the principle of one-person, one vote is a cornerstone equality-matching norm of modern democracies. People who treat these rights and responsibilities as open to the monetary trade-offs of market-pricing relationships show at best ignorance and at worst contempt for the spheres of justice that society insulates from the universal solvent of money (cf. Walzer, 1983). The response to the threat is—not surprisingly from the perspective of any appraisal theory of emotion—moral outrage. Outrage dissipates only within the rarefied ideological subculture of the libertarian movement whose members share a commitment to free choice within competitive markets. It is worth stressing, though, that libertarians are capable of outrage. Free-response data suggested that their wrath was largely

reserved, however, for "moral busy bodies" who are forever inventing injuries to third parties that justify new regulatory restraints.

Support also arose in both experiments for the moral-cleansing hypotheses of the SVPM. Merely contemplating taboo trade-offs spurred declarations of intent to volunteer to halt a ballot proposition to legalize the buying and selling of adoption rights and to assist a campaign to increase organ donation. The obvious parallel is to the transgression–compliance effect in the altruism literature (Carlsmith & Gross, 1969). People induced to believe that they have harmed others seized opportunities to repair their social identities by engaging in prosocial acts. But the parallel is imperfect inasmuch as our respondents neither harmed anyone nor stood accused of any transgression. Our results are, however, open to two other distinct but not mutually exclusive interpretations: (a) Merely contemplating taboo trade-offs may be sufficient to create a sense of moral contamination (feeling dirty, befouled) that people try to eliminate by strenuously reaffirming their commitment to defending the moral order against market intrusions; (b) calling attention to taboo trade-offs may have had the effect in Study 1 of increasing the perceived potency of political forces that sought to legitimize such modes of thinking and in Study 2 of increasing the perceived need to expand medical resources for helping desperately ill people. The former interpretation invokes an automatic, visceral response to contamination of the sort that Rozin and Nemeroff (1995) investigated; the latter invokes a conscious, purposive response to an emergent threat. Although the SVPM posits both expressive and instrumental processes to be at work, they could be disentangled experimentally—a point to which we return later.

Whereas Experiment 1 highlighted the deep differences between routine and taboo trade-offs, Experiment 2 highlighted the equally deep distinctions between taboo and tragic trade-offs. Even when the hospital administrator ultimately affirmed life over money, his social identity was tarnished to the degree that observers believed that he lingered over that decision. It was as though participants reasoned "anyone who thinks that long about the dollar value of a child's life is morally suspect." Although the taboo-breaching decision maker who affirmed life after long deliberation was not rated as negatively as the taboo-breaching decision maker who chose money after long deliberation, he was still rated negatively relative to the decision maker who disposed of the taboo trade-off quickly by affirming the sacred value. The almost mirror-image functional relationship between length of deliberation and evaluations of the decision maker in the tragic trade-off condition underscores not only the acceptability of trading sacred values against each other but the profound distinctions people draw between taboo and tragic trade-offs. Participants in the tragic trade-off conditions apparently reasoned: "The longer the deliberation, the greater respect shown for the solemnity of the decision."

Overall, moral outrage and cleansing rose and fell in tandem across the eight conditions of Experiment 2. They diverged most noticeably when the administrator considered the taboo trade-off a long time but ultimately affirmed the child's life. Here outrage was present but muted in comparison with the conditions in which the taboo decision makers deliberated either a short or long time and made the "wrong" choice. By contrast, moral cleansing was statistically indistinguishable from, and close to, its maximum when the administrator lingered over the taboo trade-off but affirmed life. A post hoc interpretation is that respondents were hard-pressed to justify a strong outrage response to the administrator in this condition (he did finally do the "right thing"), but they were left with the queasy feeling that the decision was a close call, that a precedent had been set for making these types of trade-offs, and that, next time, the decision may go the other way. People thus tried to shore up the normative order, and contribute to the solution of a life-and-death problem, by engaging in moral cleansing with the practical goal of alleviating future organ shortages.

Experiments 3 and 4: Forbidden Base Rates

In Experiment 3, we examined observers' reactions to decision makers who used base rates that either did or did not turn out to be correlated with the racial composition of neighborhoods. The hypotheses included: (a) the symbolic antiracism hypothesis, that people would regard actuarial risk as a legitimate rationale for price discrimination in setting insurance premiums only when the correlation between actuarial risk and racial mix of neighborhoods is not mentioned. When the correlation is highlighted, people—especially liberals—will vehemently reject race-tainted base rates and invoke multiple grounds for rejecting them (a variant of the defensive-overkill hypothesis); (b) the covert-racism hypothesis, that conservatives would deviate from this trend and seize on the base rates as justification for charging steep premiums to a long-standing target of prejudice in American society: Blacks.

In Experiment 4, we examined how decision makers react when they discover that a base rate that they used in setting insurance premiums is correlated with the racial composition of neighborhoods. The hypotheses were that: (a) Decision makers who discover that they inadvertently used race-tainted base rates in setting prices will try to revise their estimates as well as engage in moral cleansing; (b) these effects will be most pronounced among liberals (the symbolic antiracism hypothesis predicts that discovering one has adopted a race-tainted pricing policy will be painfully dissonant for those who conceive of themselves as defenders of the disadvantaged) and may even be reversed among racial conservatives (the not-so-covert racism hypothesis predicts that some people will raise premiums after learning which neighborhoods are predominantly Black).

Methods: Experiments 3 and 4

Procedure for Experiment 3. A sample of 199 undergraduates was randomly assigned to conditions in a 2 (equal vs. unequal pricing) × 2 (racial composition of neighborhoods) factorial design. They learned that the research goal was to explore how people make judgments, that they would be judging an actual business decision-making episode, and that there was a strong chance that the experimenter would call on them to explain why they made their judgments.

Insurance scenario. All participants learned that insurance is required for all bank loans to purchase houses. This insurance can be expensive, which can prevent people with limited means from buying homes for their families. Participants then received one of three versions of the scenario:

> Dave Johnson is an insurance executive who must make a decision about whether his company will start writing home insurance policies in six different towns in his state. He classifies three of the towns as high risk: 10% of the houses suffer damage from fire or break-ins each year. [It turns out that 85% of the population of these towns is Black/no reference to race]. He classifies the other three towns as

> relatively low risk: less that 1% of the houses suffer fire or break-in damage each year. [It turns out that 85% of the population of these towns is White/no reference to race].

To assess the potential discrimination in favor or against largely White towns, another condition was later added in which the high-risk towns were 85% White.

Respondents then agreed or disagreed with the following five assertions on 9-point scales: (a) The executive should offer insurance policies for sale in all of the towns and for the same price across all of the towns; (b) The executive should offer insurance policies for sale in all six towns but charge higher premiums for people who live in the high-risk towns; (c) The executive should feel free to offer insurance policies for sale only where he feels he can make a reasonable profit, and if that means only selling policies in the low-risk towns, so be it; (d) If the executive won't write policies for all of the towns, he should write policies for none of the towns; (e) If the executive offers insurance policies for sale only in the low-risk towns, the government should have the right to prosecute him and his company for its discriminatory behavior.

At this juncture, the second independent variable was introduced. The executive decided either to write policies for the same price for all six towns (the egalitarian or ignore-the-base-rates decision) or to write policies for only the low-risk towns (the profit-maximizing or heed-the-base-rate decision):

> [He decided that the fair and compassionate thing to do was to sell policies in both the mostly White low-risk and mostly Black high-risk towns and to charge the same price in all towns/no reference to race.] [He decided that maximizing profits was the right business decision. His decision, therefore, was to sell policies only in the mostly White, low-risk towns and to refuse to service the mostly Black, high-risk towns/no reference to race.]

Respondents then rated their reactions to the decision on 9-point scales: (a) Angry, (b) Saddened, (c) Pleased, (d) Outraged, (e) Would Criticize His Decision If I Met Him. They also rated the decision per se: (a) Fair, (b) Immoral, (c) Foolish, (d) Shows Good Business Sense, (e) Contemptible. Respondents then answered four policy questions that assessed (on 9-point scales) the perceived accuracy of the base-rate information provided, the appropriateness of using such information in setting insurance rates, the appropriateness of focusing solely on profit, and the plausibility of strictly financial rationales for treating people equally.

Procedure for Experiment 4. This study shifted the role that participants played from observers of the process of setting insurance premiums to role-playing participants. A total of 330 participants were randomly assigned to a 2 (race-taint vs. no taint to base rate) × 2 (order-of-questions) design. Subjects were asked to imagine that they were insurance agents responsible for setting premiums for policies to be sold in different zones of the city of Columbus, Ohio. Participants learned that because of the aging state of many houses in Columbus and because of the steep increase in the use of electrical appliances in modern society, the threat of fire to homes is at the greatest level in years. Because of this increased threat, mortgage lenders require all home owners to obtain fire insurance. For an insurance company to make a profit, rates must be set so as to cover the predicted amount of money lost from fires in a specific risk category. Participants were then given specific case information: Houses can be classified into three categories of neighborhood risk for fire damage: a 1 in 1,000; a 1 in 500; and a 1 in 100 chance of fire damage or loss per year. Accountants have compiled a table that insurance agents can use in setting insurance premiums. This table indicated that the company would need to sell policies for an average of $100 in the low-risk neighborhood; $200 in the medium-risk neighborhood; and $1,000 in the high-risk neighborhood. These premiums would permit the company to make "a fair profit" in each zone. Participants were also provided with the price that the company would have to charge if it were to charge the same rate across all neighborhoods and still make a fair profit ($430).

In the exercise, participants played the role of company representatives responsible for setting prices. They imagined that a homeowner from the high-risk zone had inquired about a fire-insurance policy. Insurance agents, participants were told, have some leeway in their decisions. They are allowed to charge an insurance rate based on neighborhood or to charge the same rate across neighborhoods. Participants were told to keep in mind that the numbers provided by the company's actuaries indicate the minimum price for the company to make a fair profit. Participants were then asked: "Based on the information your accountants have given you about the applicant's neighborhood, how much would you charge for this person's insurance policy?"

Participants in the race-tainted base-rate conditions were randomly assigned to two conditions that varied when they got a chance to change their pricing decisions. In the first condition, immediately after making their estimates, participants learned of the close correlation between neighborhood risk and a percentage of Blacks in the neighborhood, with only 10% of the population of the low-risk zone being African American, 30% of the population of the medium-risk zone, and 70% of the population of the high-risk zone. Participants were told,

> In short, the people who wind up paying the highest rates—the people in the high-risk zone—are mostly Black. When such information becomes available, some decision-makers feel that they need to change or update their decision. However, some do not. Based on this additional information about the applicant's neighborhood, would you change your earlier recommended price for homeowner's insurance?

Participants could then respond "yes" or "no" and, if yes, to provide a revised monetary estimate. Next, participants answered five policy questions that explored perceptions of the accuracy of the base-rate information and the appropriateness of using it. Then, participants responded to three moral-cleansing dependent variables on 9-point scales: (1) the emphasis participants planned to put (relative to last year) on attending organized cultural activities such as an African American art show; (2) the interest expressed in participating in a campus-wide rally for racial equality; (3) the interest expressed in participating in an organized publicity drive to locate a student who had mysteriously disappeared.

In the second condition, the other half of the race-tainted base-rate participants received identical instructions but were not given an opportunity to revise their estimates immediately after learning of the adverse impact on Blacks. Instead, they first responded to the five policy and three moral-cleansing questions and, only after doing so, were given an opportunity to revise their judgments.

Participants in the no-racial-taint base-rate conditions received the same general instructions as did those in the race-tainted conditions but received no indication that zonal risk might covary with racial mix of populations. As in the race-tainted conditions, however, the order of the premium-estimation and moral-cleansing questions was counterbalanced.

Race Relations Questionnaire. Prior to completing the tasks described in Experiments 3 and 4, all participants responded on 5-point scales to the following items drawn from past survey research (items that Sniderman and Piazza [1993], among others, argue provide a valid measure of "racial liberalism–conservatism"). Illustrative items included: "Government officials usually pay less attention to a request or complaint from a Black person than from a White person"; "Over the past few years, Blacks have gotten less than they deserve"; "Most Blacks who receive money from welfare programs could get along without it if they tried"; "Irish, Italian, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors"; "It's really a matter of some people not trying hard enough; if Blacks would only try harder they could be just as well off as Whites."

Results: Experiment 3

Racial Liberalism Measure. This scale, derived mostly from items in National Election Studies, had impressive reliability

Table 2
Support for Egalitarian Versus Profit-Maximizing Policies as a Function of Racial Ideology and Type of Base Rate (Experiment 3)

Experimental condition	Dependent measures: Sell policies in all towns and for the same price	Dependent measures: Sell policies everywhere but use differential pricing	Dependent measures: Combined index (of all five policy measures)
"Black-tainted" base rate			
1. Racial liberals	$6.6_{2-7,9}$	6.2_{2-9}	$31.4_{2-6,8,9}$
2. Racial moderates	$4.9_{1,3}$	4.4_{1}	$23.4_{1,3}$
3. Racial conservatives	$3.6_{1,2,4}$	$3.4_{1,4,5}$	$19.0_{1,2,4,5,7,8}$
"Nonracial" base rate			
4. Racial liberals	$4.8_{1,3}$	$4.7_{1,3}$	$27.0_{1,3,6}$
5. Racial moderates	4.5_{1}	$5.0_{1,3,6}$	$26.6_{1,3,6}$
6. Racial conservatives	4.3_{1}	$3.7_{1,5}$	$21.0_{1,4,5,7}$
"White-tainted" base rate			
7. Racial liberals	4.5_{1}	4.1_{1}	$27.3_{3,6}$
8. Racial moderates	5.1_{1}	3.9_{1}	$24.8_{1,3}$
9. Racial conservatives	3.8_{1}	3.8_{1}	20.8_{1}
Total			
Racial liberals	5.4	5.2	28.8
Racial moderates	4.7	4.6	25.0
Racial conservatives	3.9	3.6	20.0

Note. Judgments on the first two measures were made on 9-point scales. Higher values indicate greater agreement with egalitarian policies. Subscripts for each mean indicate the row numbers of those means that are significantly different from that mean (LSD test, $p < .05$).

(Cronbach's $\alpha = 0.86$). To simplify exposition and to tease apart negative reactions to Blacks among conservatives and positive reactions to Blacks among liberals, we trichotomized the sample into low, moderate, and high scores.

Moral Outrage Scale. Again, maximum-likelihood factor analysis (Browne, Cudeck, Tateneni, & Mels, 1998) revealed a generic moral-outrage factor. A direct Quartimin rotation yielded good fit for a three-factor solution, with RMSEA = .012, p(close fit) = .84, and $\chi^2(18, N = 196) = 18.54$, $p = .42$. Each item that loaded .2 or higher on the first and most interpretable factor was summed to create the moral-outrage index. These 7 items possessed good internal consistency ($\alpha = .85$) and tapped anger, sadness, outrage, criticism of the decision, and beliefs that the profit-maximizing decision was immoral, foolish, and contemptible.

Testing the key hypotheses. A 2 (nonracial vs. racial base rate) × 3 (levels of racial liberalism) analysis of variance assessed impact on the perceived appropriateness of various sales policies. As Table 2 indicates, liberals most strongly endorsed the idea that the executive should sell home insurance for the same price across zones ($M = 5.43$), followed by moderates ($M = 4.72$) and conservatives ($M = 3.91$), $F(2, 163) = 9.39$, $p < .01$. Liberalism also interacted with the type-of-base-rate information, $F(4, 163) = 5.05$, $p < .01$. Liberals exposed to the Black racial base rate agreed most strongly that the executive should sell policies for the same price across zones ($M = 6.60$). This mean differed significantly from the next highest mean (moderates exposed to the Black racial base rate; $M = 4.94$), $F(1, 163) = 7.54$, $p = .01$, and all other means.

Examining the effects of base rate and liberalism on the belief that the executive should sell insurance policies in all zones but charge higher premiums in high-risk zones revealed a main effect of liberalism, $F(2, 163) = 13.12$, $p < .01$. Liberals disagreed most with this statement ($M = 5.2$), followed by moderates ($M = 4.6$) and conservatives ($M = 3.6$). However, liberalism interacted with base-rate information, $F(2, 163) = 4.3$, $p < .05$. Liberals exposed to a Black racial base rate disagreed most strongly with this statement ($M = 6.2$). This mean differed significantly from the next highest mean for moderates in the nonracial base-rate condition ($M = 5.0$), $F(1, 163) = 3.99$, $p < .05$, and all other means.

To examine the impact of the "White-tainted" base rate, an ANOVA contrasted that condition against the "Black-tainted" condition. As predicted by the symbolic antiracism hypothesis, liberals exposed to the Black-tainted as opposed to the White-tainted base rate were more likely to agree that the executive should sell insurance for the same price across zones, $F(1, 37) = 5.88$, $p < .05$. In addition, liberals exposed to the Black-tainted base rate were less likely to agree that the executive should charge higher premiums in the high-risk zones, $F(1, 37) = 7.42$, $p = .01$. To test the blatant-racism hypothesis (that conservatives would support more egalitarian pricing when the high-risk zones turn out to be populated by whites) the same contrasts were performed, but they revealed no effects on any dependent measure.

Additional analyses capitalized on the high interitem correlations among the five policy questions ($r = .41$, $\alpha = .78$) and collapsed them into a single index. Liberals were significantly more egalitarian than conservatives on the composite "policy" dependent measure, $F(1, 142) = 34.81$, $p < .01$. Consistent with the symbolic antiracism hypothesis, the liberal–conservative difference in egalitarianism was also more pronounced in the Black-tainted than in the race-neutral or White-tainted base-rate conditions $F(1, 61) = 26.02$, $p < .001$, an effect due to liberals'

becoming more egalitarian in the Black-tainted base-rate conditions, not to conservatives' becoming less egalitarian (as the blatant-racism hypothesis predicted).

Reactions to insurance executive. A 2 (nonracial vs. racial base rate) × 2 (profit maximizing vs. egalitarian decision) × 3 (levels of egalitarianism) ANOVA assessed moral outrage triggered by different sales policies. Overall, participants were more outraged by the profit-maximizing than the egalitarian decision (Ms = 32.0 vs. M = 17.7), $F(1, 157) = 56.79$, $p < .01$. A second-order interaction indicated that, as the symbolic antiracism hypothesis predicted, liberals especially harshly condemned the executive who refused to sell to high-risk neighborhoods that were disproportionately Black (M = 44.6), $F(2, 157) = 4.18$, $p < .05$. A simple main-effects analysis indicated that this peak–outrage mean differed at borderline significance from the next highest mean of 35.7 for racial moderates exposed to the nonracial base rate and the profit-maximizing executive, $F(1, 157) = 3.37$, $p < .07$, and was clearly significantly different from all other means.

Justifications for ignoring base rates. As the defensive-overkill hypothesis predicted, liberals were most prone to invoke mutually reinforcing reasons for ignoring race-tainted base rates. The two most moralistic objections—whether or not the statistics on riskiness of neighborhoods are true, the company should not use them; and whether or not the company would make more money by charging differential prices, it should not because doing so is morally wrong—were highly correlated (r = .75) and combined into one measure. Liberals expressed more agreement on this measure than both moderates and conservatives combined (M for racial liberals = 11.4, and M for both other groups = 8.2), $F(1, 38) = 5.35$, $p < .05$. Two other strategies of resisting base rates—denying the accuracy of the statistics on the riskiness of neighborhoods and arguing that the insurance company will make more money in the long run by treating people equally—yielded no effects, both Fs < 1.

Results: Experiment 4

Policy revision. The hypothesized interaction between racial liberalism and racial significance of the base rate emerged. Consistent again with the symbolic antiracism hypothesis, liberals were especially likely to scale down their initial recommended prices for insurance policies when they discovered that the risk-status of neighborhoods correlated with the percentage of Blacks in those neighborhoods. Of the price shifters, 27 were liberals (out of 52), 9 were moderates (out of 40), and 2 were conservatives (out of 48), a significant deviation from chance, $\chi^2(2, N = 140) = 29.43$, $p < .001$. A regression analysis shows that, among those who did scale their prices down, racial liberalism predicted the magnitude of the price shift, $\beta = -.49$, $p < .001$. An examination of all participants shows that liberals lowered their average price by $190 whereas the two other groups combined lowered their price by only $22, $F(1, 150) = 23.25$, $p = .001$. Some support also materialized for the blatant-racism hypothesis—although only 11 subjects raised premiums when they learned of the population mix. That small fraction was overwhelmingly conservative (9), with one liberal and one moderate, $\chi^2(2, N = 140) = 11.62$, $p < .01$.

Moral cleansing. We constructed a composite moral-cleansing variable that aggregated responses to the cultural-activities and racial-rally questions, $r(327) = .50$. Again, the hypothesized interaction between racial liberalism and the political status of the base rate materialized. Liberals who initially set insurance premiums responsive to race-tainted base rates expressed stronger moral-cleansing intentions, $F(2, 321) = 8.51$, $p < .001$). The mean for liberals exposed to the race-tainted base rate (13.8) differed significantly from the next highest mean of 10.6 for racial liberals not exposed to the race-tainted base rate, $F(1, 107) = 21.30$, $p < .001$. Interestingly, a similar, though less pronounced, interaction emerged for the "missing student" question, $F(2, 321) = 8.07$, $p < .01$. Liberals exposed to the race-tainted base rate reported more willingness to search for the student than did the other groups (M = 6.6; next highest mean, moderates with no racial information = 6.2). The expected second-order interaction—in which the greatest moral cleansing was expected among liberals not yet given an opportunity to correct the estimates they had inadvertently based on race-tainted base rates—did not, however, emerge, $F(2, 321) < 1.0$. Indeed, the order in which the moral-cleansing– and premium-revision–dependent variables were assessed made no difference, $F(1, 154) = .45$, $ns(p < .50)$.

Discussion: Experiments 3 and 4

For many respondents, the use of base rates raised disturbing moral issues rather than tricky statistical issues. Permissible base rates in a race-neutral context were morally foreclosed in a race-contaminated context. These effects were driven largely by the insistence of liberals that base rates became "off limits" once the linkage with race was revealed. Their overriding concern was to ensure that a group that had historically suffered from discriminatory practices (and arguably may still be so suffering) would not, once again, be victimized. The opposite effect, using base rates to justify harsh reactions to Blacks, did not materialize at all in Experiment 3, even among the most conservative, and materialized only among a small minority of conservatives in Experiment 4. This "dog-that-did-not-bark" is contrary to the prediction of theories of racial policy reasoning that depict many, even most, Americans as covert or symbolic racists who are quick to seize on pretexts for denying opportunities to Blacks (cf. Sniderman & Piazza, 1993). Indeed, the pattern is more consistent with a view of liberals as "symbolic antiracists" (who change their views about the acceptability of inequality as soon as it implicates historically oppressed groups) than it is of conservatives as symbolic racists (who are always looking for justifications for thwarting the aspirations of oppressed groups).

Answers to the policy questions shed light on sources of resistance to using race-tainted base rates. The defensive-overkill hypothesis received qualified support. Liberals were more likely to argue both that, even if the information were true, it would be morally inappropriate to use it and that, even if the profit-maximizing strategy were to charge different prices across zones, it would be morally wrong to do so. But liberals did not indiscriminately embrace any justification for not using the base rates. Liberals viewed the pragmatic or empirical grounds offered for dismissing the base rates as implausible. They were not more inclined to challenge the statistics or to argue that the best long-term profit-maximizing strategy is to charge the same price. Instead, liberals invoked a straightforward moral defense against

policies that harmed the already disadvantaged. How strategic or internalized this resistance to the base rate is could be determined by the familiar battery of methodological strategies for distinguishing impression management from intrapsychic processes (cf. Tetlock & Manstead, 1985).

The moral-cleansing effects in Experiment 4 on forbidden base rates roughly parallel those observed in Experiments 1 and 2 on taboo trade-offs. The manipulation in Experiment 4—convincing participants that they had inadvertently used a race-tainted base rate—was arguably stronger, however, than in the taboo–trade-off experiments (in which there was no implication that participants were guilty of taboo trade-offs). And the effect in Experiment 4 was greater (explaining 21% of the variance in moral cleansing as opposed to 7% in Experiment 2 and 8% in Experiment 1, excluding libertarians and Marxists). It is also worth noting that the predicted order effect in Experiment 4 did not arise. Moral cleansing was as intense among race-tainted participants who were immediately given the opportunity to revise premium estimates as among those who could change their premium estimates only after moral cleansing. Two possibilities emerge here: (a) The compensatory hypothesis is wrong—when the identity threat is great enough, people often use multiple identity-repair strategies (changing their minds and affirming their fair-mindedness in other ways); and (b) the compensatory hypothesis is right, and we have yet to create the necessary conditions for observing it.

Experiment 5: Heretical Counterfactuals

Heretical counterfactuals apply causal schemata that are routine in everyday life but profoundly controversial when extended to the sacred founders of religious or political movements. The extensions become controversial because they undercut the guiding assumption that the movement arose not as the result of historical accident that can be easily "mentally mutated" out of existence (Kahneman & Miller, 1986) but rather as the result of higher order forces, perhaps even divine in character, that guarantee the fundamental correctness of the creed.

Key hypotheses were that: (a) Christian fundamentalists will most emphatically reject close-call counterfactuals that imply that the life of Christ could easily have been transformed by accidental forces of human life and social circumstance; (b) Christian fundamentalists will be most outraged by these heretical counterfactuals; (c) fundamentalists will not object to the rules of causal reasoning that underlie heretical counterfactuals when those rules are applied to nonreligious content (secular counterfactuals); (d) fundamentalists will feel morally tainted by the mere contemplation of heretical counterfactuals and engage in moral cleansing.

Method: Experiment 5

A total of 225 undergraduates were randomly assigned to a 2 (secular vs. heretical counterfactuals) × 2 (order of questioning) design. Participants were told that the goal of the project was to explore the perceptions of both laypersons and clergy of the historical events surrounding the life of Jesus Christ as described in the New Testament. The focus would be on the "what-ifs" of the Biblical narrative: ways, if any, in which events might conceivably have worked out otherwise. To this end, the questionnaire would present potential choice points in the life of Christ. For each claim, respondents made the following judgments (on 9-point scales):

1. How easy or difficult is it to imagine that the starting point for the argument could have been true? Consider the argument If Joseph had not believed the message that Mary had conceived a child through the Holy Ghost and that there was no reason to fear taking Mary as his wife, then Jesus would have grown up without the influence of a father and would have formed a very different personality. Is it easy or difficult to accept the premise that Joseph could have decided not to believe the angel's message?

2. Assuming, just for sake of argument, that the starting point is reasonable (putting to the side your personal views on the subject), how easy or difficult is to imagine the consequence following?: For example, assuming that Joseph played no active role in the childhood of Jesus, does it follow in your mind that Jesus would have grown up to be a very different person?

In addition to the previous counterfactual, participants judged the following counterfactuals: "If the three wise men had not believed the warning from God (delivered in a dream) that they should not return to Herod and report the birth of Christ, Herod would have killed Christ in his infancy"; "If Jesus had given in to one of the devil's temptations during his fast of 40 days and nights in the wilderness, Jesus's mission on earth would have been hopelessly compromised"; "If Jesus had not chosen Judas as one of his 12 disciples, Jesus would not have been betrayed or crucified"; "If Pilate had persisted with his initial belief that he could find no fault in Jesus and refused to order crucifixion, Jesus would not have died on the cross"; "If Mary had given birth to more children after Jesus, she could not be portrayed as the Holy Virgin central to Christian beliefs"; "If Jesus' body was taken from the tomb by Joseph of Arimethea (who helped remove Jesus from the cross), the apostles would have falsely interpreted the empty tomb as Jesus being raised from the dead"; and "If Jesus had allowed himself to be saved by his apostles or through divine intervention, Jesus would not have died on the cross and thus would have failed in his divine mission."

Finally, participants made judgments on 9-point rating scales of the author of a book who endorsed each of the counterfactual claims: "This person is likely to admire–have contempt for the Christian faith"; "This person displays a deep ignorance–understanding of the Christian faith"; "I find this person's beliefs to be highly offensive–compatible with my own beliefs"; "My emotional reaction to this belief is anger–sorrow–disappointment–hope"; and "I would like to seek out–avoid this person's company." Respondents also answered moral-cleansing questions exploring their intentions concerning future support for religious causes (*much less than last year–about the same–much greater*). Approximately half the participants judged the book author first, while the other half responded to the moral-cleansing items first.

In addition, a control group judged a set of counterfactuals that had no religious content but applied the same causal reasoning underlying the heretical counterfactuals. These participants learned that the goal was to assess reactions to causal arguments framed in the form "If X had happened, then Y would/would not have happened. You may find certain arguments controversial or you may feel that others are obviously true." Participants then judged both the plausibility of the antecedents and antecedent–consequent linkages for a series of assertions designed to capture the abstract causal logic of corresponding heretical counterfactuals: (a) It is fair to say that, for the typical adult, if his/her father had left the family early in that person's childhood, that person would have developed a very different personality from the one he/she would have developed if the father had remained; (b) If a person who had a reputation for great integrity and morality had given in to temptation to act immorally, most people would lose faith in that individual; (c) If a group that was betrayed by a corrupt or dishonorable member had not been so betrayed, the group could have escaped the consequences of the betrayal; (d) If a judge in a criminal trial believed that he could find no fault in the defendant's behavior, he would be very unlikely to convict and punish the defendant; (e) If someone who intends to commit murder does not know the location

of his victim, then he cannot commit the murder; (f) If an object that people expect to find in a certain place is missing because someone has sneakily removed it, then people will be surprised and may often draw false conclusions about why it is missing.

Prior to judging the counterfactuals, participants responded on 5-point scales to a 9-item scale adopted from a religious fundamentalism scale developed by Martin and Westie (1959). Illustrative items included: The New Testament of the Bible is the inspired word of God; the religious idea of heaven is not much more than superstition; Christ was a mortal, historical person, but not a supernatural or divine being; Christ is a divine being, the Son of God; if more of the people in this country would turn to Christ, we would have a lot less crime and corruption.

Results: Experiment 5

Religious Fundamentalism Measure. Replicating Martin and Westie (1959), the scale possessed good internal consistency (α = .93). This measure was trichotomized into low, moderate, and high scores on fundamentalism.

Resistance-to-Counterfactual Measure. The two strategies of neutralizing counterfactuals—challenging the mutability of the antecedent and the soundness of the antecedent–consequent linkages—were sufficiently correlated (average r[97] = .55) to justify aggregation into a single index. The expected interaction then emerged. As Figure 2 indicates, resistance peaked among fundamentalists confronted by heretical counterfactuals (M = 7.4), F(2, 228) = 46.99, $p < .001$. The mean for religious fundamentalists confronting heretical counterfactuals differed significantly from the next highest mean of 5.4 (for moderate fundamentalists confronting heretical counterfactuals), F(1, 57) = 57.46, $p < .001$.

Moral-Outrage Measure. Maximum-likelihood factor analysis (Browne et al., 1998) was used to create the index of moral outrage. A direct Quartimin rotation yielded adequate fit for a four-factor solution, with RMSEA = .064, p(close fit) = .166, and χ^2(32, N = 215) = 60.14, p = .002. The 6 items, which loaded at .3 or higher, defined the first factor: Moral Outrage. These items, which possessed good internal consistency (standardized α = 0.93), tapped anger, sorrow, disappointment, outrage, finding the author's beliefs offensive, and willingness to protest. The second factor (with high-loading items such as "leaves a bad taste in my mouth," "disgusted," "queasy," and "feeling morally violated") was designated Disgust; the third factor (with high-loading items such as "like to avoid this person's company" and "angry at author") was designated Ostracism; the fourth factor (with high-loading items including "author has contempt for the Christian faith, is deeply ignorant of the Christian faith," and "has highly offensive beliefs") was designated Strained Forbearance.

Figure 2 shows the mean outrage triggered by heretical and secular counterfactuals among low, moderate, and high scorers on fundamentalism. Overall, people reported greater outrage in response to heretical than to secular counterfactuals that applied the same underlying causal logic but to ordinary mortals in routine situations, (Ms = 3.51 vs. 3.04), F(1, 221) = 3.56, p = .06. There was also a powerful interaction between type of counterfactual and religious fundamentalism, F(2, 217) = 15.46, $p < .001$. Fundamentalist Christians were most outraged by heretical counterfactuals (M = 5.40), a mean that was significantly different from all other means (the next highest mean was 3.31 for fundamentalist Christians exposed to secular counterfactuals, F[1, 83] = 25.16, $p < .01$). The more fundamentalist the respondents, the more categorically they rejected heretical counterfactuals, F(2, 87) = 37.76, $p < .001$. As Figure 2 indicates, the same patterns emerged for the Disgust, Ostracism, and Strained Forbearance factors (average r = .70). Fundamentalists were most disgusted by heretical counterfactuals, most prone to penalize those who endorse such propositions, and most pained and strained by such propositions. There was no relationship, however, between fundamentalism and reactions to secular counterfactuals.

Moral cleansing. An ANOVA revealed the predicted interaction, F(2, 219) = 24.49, $p < .001$: Fundamentalists were especially likely to engage in cleansing after contemplating heretical counterfactuals—a mean significantly different from all other means. Again, the order effect predicted by the compensatory hypothesis proved elusive, F(1, 223) = 3.00, p = .08. Moral cleansing among fundamentalists confronted by heretical counterfactuals was neither more nor less pronounced as a function of whether participants had a chance to condemn the heretical author prior to cleansing.

Discussion: Experiment 5

Heretical counterfactuals might equally aptly be called impertinent or insubordinate counterfactuals: They undermine the dignity of what Christian fundamentalists think of as the ultimate authority-ranking relationship. How can Jesus' mission be divinely planned if it could be so easily re-routed or distorted by chance contingencies? Counterfactuals that imply that such re-directions were close calls (could easily have happened) challenge the omniscience and omnipotence of the Christian God. As one fundamentalist commented: "God did not send his only Son to die for our sins in a careless or casual way that left the success of the mission to depend on chance. God foresaw and foreclosed these possibilities."

In addition to moral outrage, moral-cleansing effects materialized—the fourth demonstration in five studies. Fundamentalists were most likely to intend to expand their involvement in church activities in the next year—a result consistent with the moral reaffirmation component of the SVPM. The nonemergence of outrage and cleansing order effects, the second failure in two attempts, does not however bode well for the compensatory hypothesis that, once people have deployed one strategy of distancing themselves from proscribed forms of social cognition, they feel less need to deploy additional strategies. There was, once again, an element of overkill in sacred-value defense.

General Discussion

The central predictions of the SVPM were repeatedly supported. Taboo trade-offs, forbidden base rates, and heretical counterfactuals evoke remarkably similar responses: Moral outrage and moral cleansing, especially from those whose conception of political justice or religious authority has been most directly challenged. Unparsimonious though it may strike those who aspire to create universal theories of social cognition, the current findings suggest that people place a complex host of superficially ad hoc content constraints on how they execute trade-offs, use base rates, and apply causal schemata to narratives. People who function like intuitive scientists or economists in one setting can be quickly

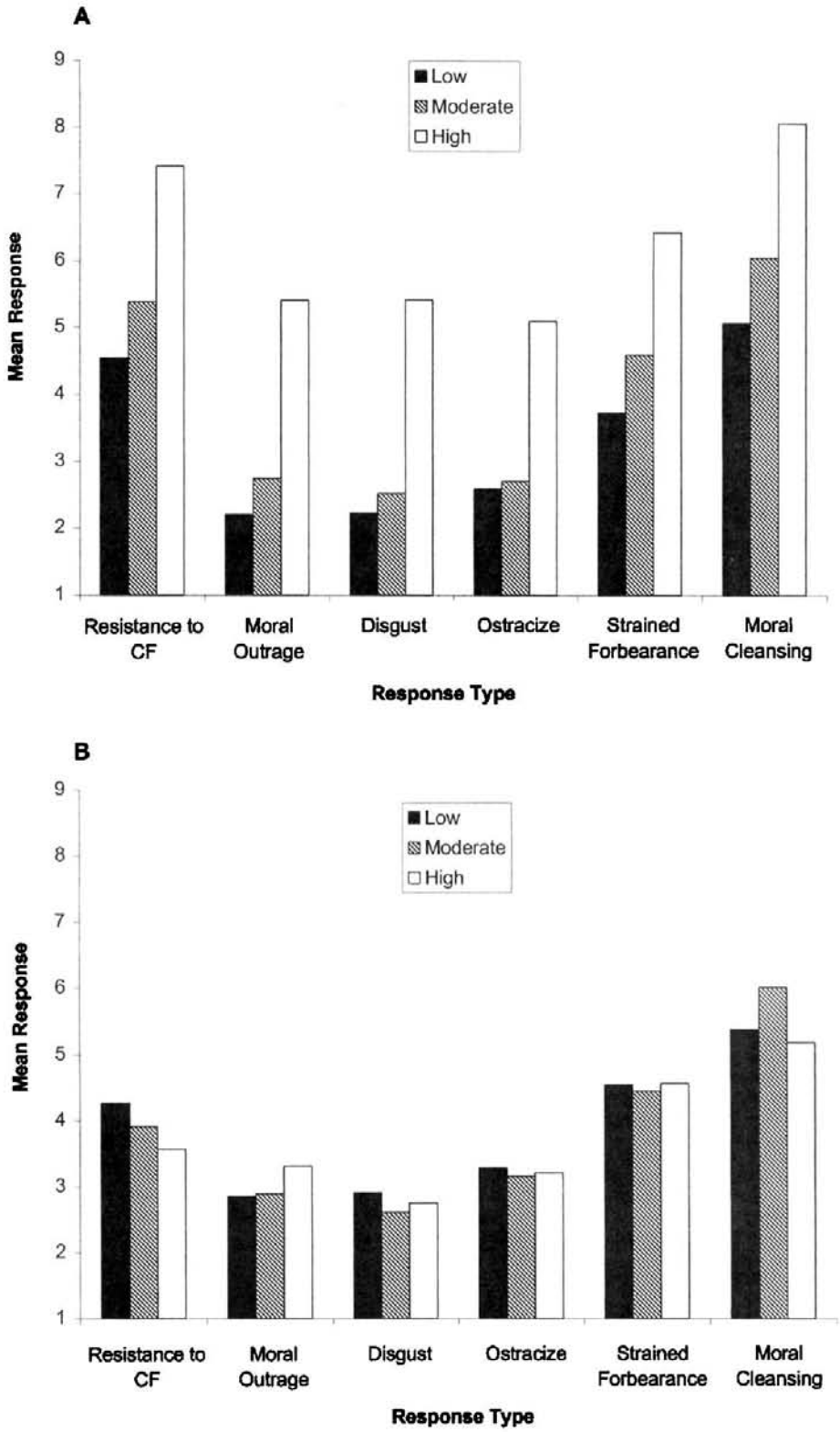

Figure 2. Mean reactions to heretical (Panel A) versus secular (Panel B) counterfactuals (CF) by level of religious fundamentalism (Experiment 5).

transformed into intuitive moralists–theologians when provoked by assaults on sacred values.

The task of general theory construction may not, however, be as hopeless as it seems if we were just to posit a never-ending series of domain-specific moralistic caveats on laws of social cognition. The solution is to link "process" frameworks such as the SVPM with "content" theories that give us explicit guidance on how people in a given culture "compartmentalize" their social world into secular and sacred domains—compartments that define the boundaries between thinkable and unthinkable. Perhaps the best off-the-shelf taxonomy of relational schemata, Fiske's (1991) model of social relations, highlights: (a) the conceptual commonalities running through taboo trade-offs, forbidden base rates, and heretical counterfactuals; and (b) the criteria that investigators can use to generate new hypotheses about other types of prescriptive and proscriptive constraints that people place on social cognition.

Turning first to taboo trade-offs, Fiske and Tetlock (1997) note that trade-offs provoke moral outrage to the degree they "inappropriately" extend a "market-pricing relational schema" (entailing ratio comparisons of absolute value) to spheres of activity regulated by the other three, less metrically onerous, schemata specified by the Fiskean model: equality matching (e.g., offering to pay one's dinner host instead of simply reciprocating the invitation), authority ranking (e.g., attempting to bribe authority figures rather than deferring to their judgment), and communal sharing (e.g., treating loved ones as objects of monetary calculation rather than honoring responsibilities to them).

Money may be a universal solvent in economic theory, but most people manifestly want to cordon off certain spheres of human activity from its corrosive powers. Child care is a communal-sharing responsibility that is somehow tainted by adoption-rights auctions for babies (an objection that, most people insist, still stands even if auctioning proves to be an efficient mechanism for placing babies in families who most value them and can best care for them; Tetlock, 1999). Moreover, as implied by the constitutive-incommensurability postulate of the SVPM, the longer observers believe that decision makers contemplate affixing dollar values to the lives and well-being of children, the sharper the moral outrage directed at them. Shifting relational frames, citizens' obligations to perform military service or to obey court orders derive from authority-ranking relations widely perceived to possess legitimate, not just coercive, power. Shifting relational frames again, buying and selling votes undercuts the equality-matching premise of one-person, one vote in modern democracies, bringing us closer to a market-pricing variant of democracy: one share, one vote. As citizens, we are deemed equal even though, as consumers and investors, equality is a transparent sham. To synthesize across domains, taboo trade-offs undermine core assumptions underlying relationships that are central to our conceptions of our selves and our social world—a result that holds up consistently in one of the most capitalistic and secularized societies on the planet at century's close (Friedman, 1999).

Forbidden base rates and heretical counterfactuals do not involve a cross-relational violation in the Fiskean scheme, but they do undercut a central implementation rule for applying a core value (equality or religious authority). In late 20th century America, a central goal of egalitarian political movements has been eliminating racial discrimination and its residual effects (Sniderman & Piazza, 1993). This goal can be justified in communal-sharing terms ("we are all members of the same national family and hence merit equal respect and dignity") or in equality-matching terms ("African-Americans have long suffered ill treatment and the time has come to balance an historically inequitable relationship"). Either way, the prospect of a company trying to maximize profit by imposing burdensome premiums on poor Black populations triggered an especially strong outrage response from the most egalitarian respondents. Knowledge that one had inadvertently used a forbidden base rate in setting premiums also triggered an especially strong moral-cleansing response from egalitarians.

Among Christian fundamentalists, there is—in Fiskean terms—a direct authority-ranking relationship between God and humanity. Believers are supposed to defer to the Scriptures, the word of God as conveyed through His Only Son and the apostles. Counterfactuals that depict the life of Christ as highly contingent affair mock, in effect, Christ's sacrifice and God's message to humanity. Heretical counterfactuals are deeply disrespectful and, in earlier times or in other religious cultures, would have justified the infliction of corporal or capital punishment on the offender. In modern societies, dissenters do not have to endure these draconian sanctions but they do still face the moral outrage of the faithful.

As noted at the outset, the moralist–theologian metaphor is one of the least explored functionalist frameworks for social cognition. One strategy for jump-starting work within the incipient research program will be to forge stronger links with strands of social psychological work that shed light on exactly how people cope with unwanted thoughts and irritating challenges. In some cases, the connections are complementary; in other cases, we should expect explanatory turf disputes. Three points of complementarity follow.

Permeability of Secular–Sacred Boundary

Whenever a stream of thought flows into forbidden conceptual territory, paradoxes of mental self-control arise. Wegner's (1994) research suggests that the harder people try to avoid thinking about taboo topics, the more difficult it becomes to stop thinking about these topics. It is unclear whether we created such a "problem" for our participants. The moral-cleansing effects suggest so. But there is a strong counterargument. The current work differs from Wegner's in a key respect. The taboo topics in our experiments offend deeply held beliefs and values, whereas the focal topics in studies of mental self-control are typically innocuous, albeit perceptually vivid, such as dancing white bears. Many participants seemed to reach moral closure rapidly in our experiments. Their reasoning sequence often took the conscious form: "Some people certainly believe some offensive things. I reject such ideas and people categorically. Case closed."

Psychological analysis need not end, however, where introspective analysis does. If this process of reaching rapid moral closure is impeded, the mental self-control necessary for preserving taboos can become more problematic. The boundaries of the unthinkable do shift over time. Tetlock (1999) has noted historical evidence of how previously blocked exchanges can become permissible (capitalists buying and selling the sacred land of financially strapped feudal lords) and previously permissible exchanges can become taboo (between the U.S. Civil War and World War I, it ceased to be acceptable to pay others to perform one's military-service

obligations). In this vein, Tetlock (1999) has also shown experimentally that people qualify their opposition to the buying and selling of body organs for medical transplants, to the degree that they can be convinced that: (a) such transactions will save lives that otherwise would have been lost due to organ shortages; (b) the poor will be assisted in purchasing needed organs and that they will not be compelled to sell their organs in "deals of desperation." A once clear-cut example of a taboo trade-off thus blurs into either a routine or tragic trade-off, depending on whether the sacred side of the trade-off has been more thoroughly "secularized" than the secular side of the trade-off "sacralized." Either way, as this political debate unfolds, intuitive moralists–theologians should have progressively greater difficulty suppressing taboo thoughts and these thoughts will trigger less outrage. The term "taboo trade-off" is thus misleading insofar as it denotes the original Polynesian meaning to the term: absolute, automatic, unreasoned aversion to any breach of the psychic barriers separating the profane from the sacred (Radcliffe-Brown, 1952). To use a Lewinian metaphor, the permeability of the secular-sacred boundary is not a constant.

Connections to Terror Management

Greenberg, Pyszczynski, Solomon, Simon, and Breus's (1994) terror management theory posits that people who are reminded of their mortality seek out the existential comfort of a collectively shared worldview that transcends their mortal life spans and endows their lives with moral significance. Linking this alternative theory of people as intuitive theologians to the SVPM leads to the hypothesis that, agnostic Bayesian libertarians excepted, people reminded of their mortality should be especially outraged by taboo trade-offs, forbidden base rates, and heretical counterfactuals that destabilize their worldview, and should be especially inclined to moral cleansing.

Qualitative Distinctions Among Emotions

Rozin, Lowery, Imada, and Haidt (1999) identify three basic emotional responses (anger, contempt, disgust) to three basic types of moral violations (individual rights, communal obligations, and divinity and/or purity). Their analysis maps imperfectly onto our tricomponent conception of moral outrage in which affect is coequal with cognition (dispositional attributions) and action (imposing sanctions) and imperfectly onto the Fiskean taxonomy of relational schemata. Our measures were not however designed to test the Rozin et al. framework, so it would be wrong to read deep significance into our factor-analytic procedures failing to reproduced their conceptual distinctions. As the varying factor-analytic solutions we obtain suggest, it is an open question as to when moral outrage is unitary or fractionates into qualitatively distinct forms.

Turning to potential tensions between SVPM and influential theories, skeptics might argue that there is no need for littering the intellectual landscape with yet another minitheory. The moral outrage and cleansing results are more parsimoniously assimilated to existing frameworks—variants of dissonance theory or ego-defensive or self-presentational formulations—that focus on how people deflect threats to the moral integrity of the self. Given previous positions taken by the first author on the impossibility of drawing sharp behavioral (if not psychophysiological) dividing lines between explanations grounded in competing functionalist metaphors (Tetlock & Manstead, 1985), it would be odd now to insist that sharp demarcations exist between the SVPM, a middle-range theory anchored in the intuitive moralist–theologian metaphor, and middle-range theories with roots in the cognitive-consistency or psychodynamic or social identity traditions. But there are differences in explanatory emphasis. The SVPM's closest competitor, Steele's (1988) self-affirmation theory, is hard-pressed to account for several results across the five experiments:

The Mere Contemplation Effect

Why should just reading about a normative transgression—no counterattitudinal act required—trigger such concerted efforts to reaffirm one's virtue and moral standing? Are some ideas so socially toxic that to fail to register one's outrage contaminates one's self-image as a decent, norm-abiding being? To be sure, dissonance theory has undergone many conceptual mutations en route to becoming a theory of ego or self-image defense (Greenwald & Ronis, 1978), so there is no reason why it cannot undergo one more transformation and dispense altogether with the notion that counterattitudinal deeds are necessary to activate dissonance.[3] This particular mutation does, however, bring us much closer to Durkheimian ideas of maintaining social equilibrium than to Festingerian ideas of mental equilibrium. The presumption must become that people feel responsible not just for their own acts but for the acts of others. Those who shirk their share of the norm-enforcement chore become violators of the meta-norm to police norm observance (Coleman, 1991). The rupture is less intrapsychic than relational: The threat to the bond that links self to the external normative order that appears to be under siege.

Lack of Substitutability of Defensive Strategies

Here again, it is unwise to draw sharp rhetorical distinctions. Work on dissonance and self-evaluation processes typically finds compensatory relationships among threat-reduction strategies, whereas our studies yielded more evidence for defensive overkill, in which participants effectively announced: "Not only do I condemn these norm violators, I'll now show you that I personally exemplify support for the norm." With benefit of hindsight, it is possible—within the logic of the SVPM—to identify circumstances under which either compensatory or overkill relationships are more likely to hold. Overkill should occur when: (a) outrage and cleansing are not costly to express; and (b) the observed normative violation is so egregious (as ours usually were) that it severely undercuts the moral order. People should then quickly hit a ceiling effect on outrage and seek out additional symbolic

[3] Indeed, one variant of dissonance theory has already mutated in this direction. Research using the hypocrisy paradigm demonstrates that simply reminding people of occasions in which they have acted contrary to their principles can trigger threats to self-esteem (Stone et al., 1997). Here would seem to be a conceptual halfway house between dissonance theory and the SVPM. The hypocrisy paradigm does still require counterattitudinal conduct, albeit from the distant past and now encoded as an event node in autobiographical memory. But the paradigm does undeniably reveal the power of mere contemplation to activate defensive reactions.

affirmations of the threatened values. Compensatory relations should hold when: (a) either outrage or cleansing has become awkward, effortful or dangerous to express; and (b) the violation is bad enough to warrant a reaction but is not "over the top." People should then be content with a single-pronged defense of the moral order. The current studies were not designed to test these ideas, but they did generally satisfy the two preconditions for defensive overkill.

Domain-Specificity of Reactions to Threat

Steele's (1988) self-affirmation theory implies that identity repair need not focus on where the damage occurred. The SVPM implies that people are choosier and that moral outrage needs to be directed at the actual perpetrators and that moral cleansing needs to redress the specific threat to the social order—be it monetizing babies, undermining racial justice, or undercutting Christianity. Our studies shed very limited light on this controversy, with Experiment 4 favoring Steele's view that cleansing (self-affirmation) can take diverse forms. This difference between formulations is also, however, best treated as one of degree, not of kind. The SVPM posits a steep generalization gradient: The functional value of outrage and cleansing in parrying a threat declines rapidly as we move farther away in moral meaning or significance from the societal values under assault. The most direct way to rebut insinuations that one is a racist is to affirm one's commitment to civil rights causes. Participants in Experiment 4 did that, but they also showed more interest in helping to find a missing person. One way to reconcile these results with the SVPM is to argue that participants assimilated all three moral-cleansing items into a generic good-cause mental account in which the goal was to create a caring society that helps those in need. But this raises more questions than it answers: How generalizable across domains must moral cleansing be to falsify the SVPM prediction? and How domain-specific must moral cleansing be to pose a problem for self-affirmation theory? The SVPM hypothesis would be decisively falsified if the effects of sacred-value threat on moral cleansing were attenuated by personality-test feedback that participants possessed a morally neutral, but self-esteem–enhancing trait such as intelligence (in implicit-personality-theory research, the morality and competence dimensions often emerge as orthogonal factors in semantic space). Self-affirmation theory would be falsified if there were, contra the results of Experiment 4, absolute domain-specificity. The interpretation of everything between these two ideal-type contrasts, including the results of Experiment 4, depends the slope of generalization gradient for this or that dimension of social identity.

Another possible challenge to the SVPM comes from advocates of self-presentational theories who might posit that participants were feigning outrage and cleansing intentions for public consumption (Schlenker, 1982). The strong form of this argument clearly contradicts the SVPM, which treats outrage and cleansing as both sincere and internalized. However, a weaker form of the self-presentational argument, which asserts that people will vent more outrage and engage in more cleansing when under the scrutiny of their community of cobelievers, is deeply compatible with the as-yet-untested SVPM hypothesis that outrage and cleansing serve instrumental interpersonal functions (norm-enforcement) as well as intrapsychic purification functions. Pace Durkheim (1925/1976), people should seek to affirm, as publicly as possible, their moral solidarity with the community. This analysis leads to testable hypotheses, including: (a) outrage and cleansing should be most pronounced when observers feel accountable for their judgments to their community of co-believers (an audience that will enforce the meta-norm that no one shirk his or her share of the task of enforcing norms); (b) observers who are under scrutiny by cobelievers but who have been prevented from directing outrage at norm violators should try to compensate for the damage to their moral identities via conspicuous forms of moral cleansing.

Ultimately, functionalist metaphors are not testable. But metaphor-inspired research programs are exhaustible. Investigators should tire quickly of sterile metaphors that bear neither conceptual nor empirical fruit. The moralist–theologian metaphor has a justifiable claim on scientific resources to the degree it stimulates testable hypotheses that generate novel discoveries and to the degree we can eventually reconcile these discoveries with reasonably well-established knowledge. There is thus an optimal level of metaphorical novelty: novel enough to lure investigators into terra incognita but not so novel as to be unassimilable into established explanatory frames of reference. On both counts, the theologian metaphor passes—at least for now.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715–729.

Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology, 68,* 986–996.

Becker, G. (1981). *A treatise on the family.* Cambridge, MA: Harvard University Press.

Belk, R. W., Wallendorf, M., & Sherry, J. F., Jr. (1989). The sacred and the profane in consumer behavior: Theodicy on the Odyssey. *Journal of Consumer Research, 16,* 1–37.

Bell, D. (1976). *The cultural contradictions of capitalism.* New York: Basic Books.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (1998). CEFA: Comprehensive Exploratory Factor Analysis. [Computer program]. Available on the World Wide Web: http://quantrm2.psy.ohio-state.edu/browne/

Buckley, W. F. (1997). *Nearer, my God: An autobiography of faith.* New York: Doubleday.

Carlsmith, J. M., & Gross, A. E. (1969). Some effects of guilt on compliance. *Journal of Personality and Social Psychology, 11,* 232–239.

Coleman, J. (1991). *Foundations of social theory.* Cambridge, MA: Harvard University Press.

Durkheim, E. (1976). *The elementary forms of the religious life* (2nd ed.). London: Allen and Unwin. (Original work published 1925)

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90,* 239–60.

Fiske, A. P. (1991). *Structures of social life: The four elementary forms of social relations.* New York: Free Press.

Fiske, A., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress spheres of justice. *Political Psychology, 18,* 255–297.

Friedman, T. (1999). *The Lexus and the olive tree.* New York: Farrar, Straus, & Giroux.

Gilbert, D. (1991). How mental systems believe. *American Psychologist, 46,* 107–119.

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts

in morality salience effects. *Journal of Personality and Social Psychology, 67,* 627–637.

Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of dissonance: Case study of the evolution of a theory. *Psychological Review, 85,* 53–57.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–453.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

Koehler, J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–53.

Martin, J., & Westie, F. (1959). The tolerant personality. *American Sociological Review, 24,* 521–528.

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology, 49,* 447–77.

Myers, D. (1993). *Social psychology* (4th ed.). New York: McGraw-Hill.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology, 43,* 87–131.

Radcliffe-Brown, A. (1952). *Structure and function in primitive society.* London: Cohen and West.

Raz, J. (1986). *The morality of freedom.* New York: Clarendon Press, Oxford University Press.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76,* 574–586.

Rozin, P., & Nemeroff, C. (1995). The borders of the self: Contamination sensitivity and potency of the body apertures and other body parts. *Journal of Research in Personality, 29,* 318–340.

Schlenker, B. R. (1982). Translating actions into attitudes: An identity-analytic approach to the explanation of social conduct. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 15, pp. 194–248). New York: Academic Press.

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology, 68,* 247–260.

Sniderman, P. M., & Piazza, T. (1993). *The scar of race.* Cambridge, MA: Harvard University Press.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). San Diego, CA: Academic Press.

Stone, J., Wiegand, A. W., Cooper, J., & Aronson, E. (1997). When exemplification fails: Hypocrisy and the motive for self-integrity. *Journal of Personality and Social Psychology, 72,* 54–65.

Tesser, A., & Cornell, D. P. (1991). On the confluence of self processes. *Journal of Experimental Social Psychology, 27,* 501–526.

Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology, 50,* 819–827.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in Experimental Social Psychology, 25,* 331–376.

Tetlock, P. E. (1999). Coping with trade-offs: Psychological constraints and political implications. In S. Lupia, M. McCubbins, & S. Popkin (Eds.), *Political reasoning and choice.* Berkeley: University of California Press.

Tetlock, P. E., & Belkin, A. (1996). *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives.* Princeton, NJ: Princeton University Press.

Tetlock, P. E., & Manstead, A. S. R. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review, 92,* 59–77.

Tetlock, P. E., Peterson, R., & Lerner, J. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *Ontario Symposium on Social and Personality Psychology: Values.* Hillsdale, NJ: Erlbaum.

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84,* 1329–1393.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160), New York: Cambridge University Press.

Walzer, M. (1983). *Spheres of justice.* New York: Basic Books.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34–52.

Received May 27, 1999
Revision received November 1, 1999
Accepted November 2, 1999 ■