

## Proscribed Forms of Social Cognition: Taboo Trade-offs, Blocked Exchanges, Forbidden Base Rates, and Heretical Counterfactuals

Philip E. Tetlock  
A. Peter McGraw  
Orie V. Kristel

Most psychological theories of judgment and choice work from the premise that people make judgments and decisions in their functional capacities as intuitive scientists or economists. As intuitive scientists, we are hypothesized to be on an epistemological mission of achieving as accurate a mental representation of the external world, and the causal relations that hold that world together as possible (Fiske & Taylor, 1991). As intuitive economists, we are posited to be on a utilitarian mission organized around the maximization of subjective expected utility (Kagel & Roth, 1995). One advantage of adopting these functionalist frameworks has been a ready supply of elegant normative models for assessing judgmental biases and errors, and for justifying attributions of irrationality that have the net effect of making researchers look smart and research participants look, well, not so smart.

This chapter works from the premise that the aforementioned perspectives are far too restrictive (Tetlock, 2002). Ordinary people often resist the normative prescriptions of models anchored in the intuitive scientist–economist frameworks, even when these prescriptions have been carefully spelled out for them. We argue here that this resistance should not always be written off as incorrigible closed-mindedness. In some situations, people are better thought of not as intuitive scientists or economists but rather as intuitive theologians who are trying to protect sacred values from secular encroachments. We also propose a testable model of how people function as intuitive theologians. According to this sacred-value-protection model

(SVPM), certain categories of mental operations are anathema and off-limits because they require us to assign explicit finite valuations to values that our moral communities tell us should be treated as absolute, unquestionable, and unqualified commitments (Tetlock et al., 2000). People are just not supposed to think in certain ways, and as soon as one senses that others have broken the taboo, they become targets of moral outrage (as, indeed, so do those members of one's moral community who fail to censure those who violate the taboo—there is a meta-norm to enforce norms; Fiske & Tetlock, 1997). And, as soon as you sense that you yourself have started to stray down a proscribed mental pathway and begun to think the unthinkable, you are supposed not only to stop but to morally cleanse yourself: to reaffirm your commitment to shared values and your status as a member in good standing of the moral standing of the community.

Of course, it can be objected that all this is not very parsimonious. It is simpler to work with a few functionalist frameworks that provide well-defined normative benchmarks for gauging rationality. The critics are right that there is a loss in parsimony. But the loss can be minimized by advancing reasonably precise and testable models that specify the conditions under which moral outrage and moral cleansing responses are likely to be activated. One of the best ways of doing that is to link the SVPM to a comprehensive theory and taxonomy of the relational schemata that people use to judge the acceptability of thought, feeling, and action in a wide range of cultural settings. Alan Fiske's (1991) relational theory fits this bill very well indeed, and we return to it in the closing section of this chapter to organize and, in certain senses, explain the sharp, qualitative content boundaries that people often place on the acceptability of certain forms of social cognition.

We focus on four specific examples of such sharp qualitative breakpoints, contexts within which the SVPM predicts that patterns of thinking that, from an intuitive scientist or economist perspective, should be unproblematic but will provoke both moral outrage and moral cleansing. The four forms of proscribed social cognition are taboo trade-offs, blocked exchanges, forbidden base rates, and heretical counterfactuals.

### **Taboo Trade-Offs**

Trade-off reasoning is fundamental to micro-economic and decision-theoretic conceptions of rationality. We live in a world in which we cannot always get what we want so we must strike compromises among our objectives: consumption versus saving, quality versus price, and work versus family.

The SVPM posits, however, that people consider a willingness to entertain certain types of trade-offs, known as taboo trade-offs, not as evidence of cognitive sophistication, but rather as evidence of moral depravity

(Tetlock, Peterson, & Lerner, 1996). A taboo trade-off can be said to exist whenever entering one value into the trade-off calculus subverts or undercuts the other value. To compare is to destroy. The mental process of comparison can itself be morally corrosive. The longer one spends contemplating the dollar value of one's children, or friendships, or loyalty to one's country—values and obligations that are supposed to be sacred and categorically nonfungible—the more one reveals oneself to be the type of person who simply does not understand the meaning of parenthood, or friendship, or citizenship.

It is useful to contrast taboo trade-offs, which characteristically entail comparisons of the relative importance of secular values (such as money, time, and convenience) and sacred values (that are supposed, at least for rhetorical purposes to be infinitely significant), with two other broad categories of trade-offs: routine and tragic. Routine trade-offs entail secular-secular value comparisons and are integral parts of everyday life. We execute such trade-offs every time one goes to the supermarket. Tragic trade-offs entail comparisons of sacred values—for example, one group of lives versus another, or life versus honor—and arise, almost by definition, only at dramatic turning points in life. Tragic trade-offs can be excruciatingly painful but they are widely recognized as a dreadful necessity.

In one set of studies, Tetlock, Kristel, Elson, Green, and Lerner (2000) compared reactions to taboo and tragic trade-offs. In one scenario, people judged a healthcare administrator who had to make one of two choices: either between saving the life of one boy or another boy (a tragic trade-off) or between saving the life of a boy and saving the hospital a great deal of money (taboo trade-off). This experiment also manipulated two other independent variables in a  $2 \times 2 \times 2$  factorial design. The second independent variable was whether the administrator found the decision easy and made it quickly or found the decision difficult and took a long time. The third independent variable was whether the decision maker decided to save the life of a particular boy, Johnny, or he opted for the other alternative (either saving the other boy or saving money). The results revealed the predicted second-order interaction. The administrator in the taboo trade-off condition who thought the decision easy and chose Johnny quickly was judged least negatively whereas the administrator in the taboo condition who found the decision difficult and eventually chose the hospital was judged most negatively. By contrast, people judged the decision maker who worked through the tragic trade-off slowly more positively than the decision maker who made up his mind quickly, regardless of which boy he chose to save. These results show that lingering over a taboo trade-off, even if one ultimately does the "right thing" and saves the child's life, makes one a target of a significant moral outrage (relative to what one would have received if one had made the decision quickly). Lingering over a tragic trade-off simply under-

scores that one understands the gravity of the issues at stake. A quick decision would indeed be disrespectful to one of the competing sacred values (and to those who advocate those values in contexts of legal or political accountability).

Taboo trade-offs are also morally contaminating. To observe a taboo trade-off without condemning it is to become complicit in the transgression. As a result, people often try to cleanse themselves. Tetlock et al. (2000) documented the greatest willingness to volunteer to increase organ donations among participants who had thought both that the decision maker had made the "wrong" choice in a taboo trade-off and that the decision maker had a hard time making up his mind.

Thinking a long time about taboo trade-offs, even if one makes the "right" decision, taints one's social identity. It is as though observers reason "anyone who thinks that long about the dollar value of a child's life is morally suspect." By contrast, participants in the tragic trade-off condition, in which the administrator had to choose which child should live seemed to reason: "The longer the deliberation, the greater the respect shown for the solemnity of the decision."

A big question, that economists are especially quick to raise, concerns the sometimes jarringly sharp disjunction between what people say about the importance of protecting sacred values and the compromises that decision makers must actually make in the real world of scarce resources. In the final analysis, someone must somehow set priorities, a process that—however distasteful we find it—requires attaching at least implicit monetary values to sacred values. We cannot devote 100% of GDP to taking care of the sick and elderly, or to educating children, or to uplifting the poor, or to advancing the cause of liberty and justice abroad. If these decision makers are to escape becoming targets of the moral wrath of the populace, some combination of two things must be true: Decision makers must be pretty skilled at obfuscating what they are doing or the citizenry must be willing to refrain from trying to peer through the rhetorical smokescreens that decision makers put up.

The toxic-waste experiment, reported in Tetlock (1999), addressed these issues. The specific scenario used was inspired by an actual federal law—the Superfund Act:

A government program cleans up toxic waste sites to the point where they pose zero risk to public health. Last year, the program saved an estimated 200 lives at a cost of \$200 million. The Danner Commission—whose mandate is to improve the efficiency of government—investigated the program and recommended a set of reforms that were implemented. As a result, this year the program could save an estimated 200 lives—just as many as before—but at a cost of only \$100 million, a saving of \$100 million. If the government kept funding the program at last year's budget level of \$200 million, the program could save an estimated 400 lives.

Now the Danner Commission recommends redirecting the saving of \$100 million to other uses, including reducing the deficit, increased funding for programs to stimulate economic growth, and lowering taxes. (p. 254)

In one experimental version, the story ends with the Commission avoiding any reference to dollar-life trade-offs and simply offering a vague deontic rationale: "Based on its analysis of these options, the commission concludes that "morally this is the right thing to do." What do you think? Should the government keep funding the program at \$200 million or should it redirect the \$100 million saved to other priorities?"

In another experimental condition, participants learned that "based on its analysis of the options, the Danner Commission has concluded that the cost of saving the additional 200 lives is about \$500,000 per life—a cost that it still considers too high and one that cannot be justified given other needs and priorities. The Commission therefore recommends redirecting the saving of \$100 million to other uses, including reducing the deficit, increased funding for programs to stimulate economic growth, and lowering taxes" (p. 255).

Support for the Danner Commission recommendation hovers at 72% when the trade-off is masked by a deontic rationale but falls to approximately 35% when the trade-off explicitly violates the normative injunction against dollar valuations of human life.

Additional experimental conditions address alternative interpretations. One argument is that people object not to attaching a dollar value to human life but rather to the particular cost estimate that the Danner Commission endorsed. There is nothing inherently wrong about pricing human life but one should get the price right. To explore this possibility, participants reacted to a Commission that engaged in cost-benefit reasoning and concluded that funding should remain at \$200 million because \$500,000 per human life is a cost worth bearing. In still other conditions, \$500,000 is a bargain; the Commission would have been willing to pay either \$1 million or \$10 million (at which price the entire federal budget for fiscal year 1996-1997, \$1.7 trillion, would allow us to save only 170,000 lives). Although people judge the Commission more favorably when it places a higher dollar value on human life, they still judge it less favorably, and substantially so, than when the Commission offers a deontic rationale either for keeping funding at the \$200 million level or for redirecting spending to other priorities.

Another possibility is that people object to utilitarian reasoning in general and not just to the violation of this specific taboo trade-off. To test this possibility, participants reacted to a Danner Commission that either redirected funding or kept funding constant and did so "after weighing all the relevant costs and benefits." Interestingly, vague utilitarian reasoning was

every bit as effective in providing political cover as vague deontic reasoning. Taboo trade-offs need to be cloaked (Calabresi & Bobbit, 1978) and either Kantian or utilitarian cloaks will do the job. People objected to the explicit spelling out of the trade-off. Doing so tarnished the image of a previously well-regarded reformist commission, rendering suspect its entire policy agenda.

These methodological variations show that previously popular politicians and acceptable policies can be transformed into objects of scorn by revealing that the politicians performed taboo mental calculations in reaching their conclusions. The damage to one's political identity can be so severe that even possessing an otherwise winning utilitarian argument is not sufficient to prevail in the court of public opinion.

The "contamination" findings have suggestive implications. As long as the public insists on treating certain values as sacred and beyond "monetization," politicians will be forced to resort to deceptive rhetorical subterfuges. To let down those subterfuges, to act as though Social Security or Jerusalem or Kashmir is just another issue to be compromised and bargained out, is to open oneself to demagogic attacks, vilification, and ridicule. A mark of a stable polity may well be the willingness to forgo the short-term advantages that come from accusing opponents of selling out secular values. Systems in which power rotates reasonably regularly among major contenders create incentives for such restraint. A mark of an unstable system may well be the escalation of charges of taboo trade-offs and the resulting "demonization" of adversaries that such charges can fuel.

#### **Blocked Exchanges and Relational Moderators of the Endowment Effect**

Blocked exchanges are a special case of taboo trade-offs in which observers express their disapproval of the "commodification" of goods and services that they deem to have special moral or emotional significance. From a micro-economic point of view, there is really only one good justification for blocking mutually consenting parties from entering into contracts: namely, negative externalities or side effects on third parties. From the standpoint of the SVPM, however, these negative externalities need not take the form of tangible material damage to identifiable individuals; they can take invisible symbolic forms and the victims can be as abstract as the community, society, the nation, the Volk, disembodied moral abstractions, or, for that matter, God.

In late 20th century America, there is a great deal of consensus about what types of social exchanges should be blocked. However, Tetlock et al. (2000) found that sharp, ideologically charged disagreements do occasion-

ally erupt. There is deep disagreement, for example, over the legitimacy of buying and selling medical care or access to high quality legal representation in court. For many on the left, these transactions come perilously close to buying and selling lives and justice. For those on the right, such transactions are not morally distinguishable from buying a house or food. Both left and right do, however, largely agree—in late 20th century American political culture—that buying and selling adoption rights for orphans or buying and selling draft obligations to perform military service are unacceptable (only within the rarefied ideological subculture of libertarianism to objections to these types of exchanges or trade-offs disappear).

Tetlock et al. (2000) surveyed reactions to routine and blocked exchanges across a broad spectrum of student political activists at Berkeley in the early 1990s. Everyone but libertarians was outraged by efforts to “monetize” babies, body parts, and basic rights and responsibilities of democratic citizenship. There was far less outrage toward routine trade-offs but outrage gradually built as we moved leftward on the political spectrum toward socialists who not only found controversial topics like surrogate motherhood unacceptable but who also had problems with buying and selling uncontroversial commodities such as houses and food or borderline controversial commodities such as medical care and legal representation. Devout egalitarians tend to see such exchanges as inherently inequitable because they put the poor at a profound disadvantage (and because they seem to carry the implication that the lives and rights of the poor are worth less than those who can pay large sums for doctors and lawyers).

A high profile example of a social exchange that people were unsure how to evaluate arose during the Clinton presidency when questions were raised about the large number of big campaign donors who were sleeping in the Lincoln bedroom (a vivid image of a sacred site subjected to secular contamination; McGraw & Tetlock, 2004). Reaction to the disclosures was initially negative among both supporters and detractors of the Clinton presidency. The unadorned facts elicited moderate outrage from both camps. But at least among Clinton supporters, the reaction was subject to quite dramatic political manipulation as a function of the semantic or relational framing of the exchange. Outrage among Clinton supporters was substantially attenuated when the facts were accompanied by a rationale—the one, incidentally promoted by the White House—that invoked a friendship as opposed to economic norm, a rationale that affirmed the right of friends to do and return favors for friends. Invoking this equality-matching or reciprocity rule reduced outrage to politically tolerable levels. Outrage among both supporters and detractors went up substantially, however, when the facts were accompanied by an explicit market-pricing rationale for the decisions as to whom to invite to stay in the Lincoln bedroom: Why not let the laws of supply and demand work their will?

We have also done some small-scale experiments that explore the political-psychological roots of objections to blocked exchanges. A controversial issue concerns exactly what taboo blocked exchanges and "taboo trade-offs" actually are. Are they taboo in the primal Polynesian meaning of the term, absolute, categorical, and unreasoned aversions that have the power to contaminate those who transgress boundaries that demarcate the sacred from the secular? Are sacred values better viewed as merely pseudo-sacred, as polite social fictions to which we pay lip service and honor at ceremonial occasions, but that we abandon as soon as upholding the sacredness of the value becomes burdensome and a convenient justification for jettisoning the value comes along?

In the policy-revision paradigm, we have explored how easy or difficult it is to "talk people out of" their aversion to secular encroachments on sacred values (see also Tetlock, Peterson, & Lerner, 1996). The first step was to document, once again, that the overwhelming majority of respondents object, often emphatically, to proposals that would allow people to buy and sell body organs and adoption rights for children. The second step was to assess whether these taboo trade-offs are truly taboo in the primal Polynesian sense of the term, prohibitions, like that on incest, that require no further explanations and that permit no exceptions. We found that approximately 60% of those who object to "marketizing" babies and body organs saw no need for a reason—even when held accountable and pressed for one—beyond the blanket condemnation "the policy is degrading, dehumanizing, and unacceptable." We take this assertion to be a simple reassertion of the taboo—an "ugh" reaction that people find it odd to be asked to justify. It is as though we had reached the backstop of these belief systems—the *de gustibus* line of defense beyond which people find further queries to be absurd. But what about the 40% of respondents who offered rather specific reasons for judging the proposal unacceptable? Informal content analysis of these reasons revealed that many objections had an egalitarian character. One common concern about medical transplants was the theory that the poor would be driven into deals of desperation. They will need money so badly that they will feel compelled to submit to dangerous surgical interventions—donating part of one's liver or a whole kidney—that the well-off would never seriously consider. A widely held fear about baby auctions was that only healthy and attractive babies would draw bidders. A related concern was that the price of organs of babies would sky rocket and the market would cater only to the well-off. Another class of objections was more pragmatic. Some people were not convinced that all other options have been investigated for increasing the supply of body organs for desperately needed transplants or the supply of parents for babies in need of families. These respondents believed that "more civilized" solutions to the policy problem could be found.



Our research on the cognitive substrate of blocked exchanges assessed how much people change their minds about markets for body organs and for babies when we revised the proposals to address substantive objections that people raised. Each policy revision was designed to minimize what economists might call the moral externalities of permitting a particular taboo trade-off. For instance, one set of questions included: "Would you still object to markets for body organs: (a) if you lived in a society that had generous social welfare policies and never allowed the income for a family of four to fall below \$32,000 per year (explaining the concept of inflation-adjusted 1996 dollars)?; (b) if society provided the less well-off with generous organ-purchase vouchers that increased in value as recipient income decreased (the poorer the recipient, the larger the voucher)?; (c) if they could be shown that all other methods of encouraging organ donation had failed to produce enough organs and that the only way to save large numbers of lives was to implement a market for body organs?" Another series of questions probed reactions to permitting all qualified parents to bid for adoption rights in one scenario, participants were assured that if only attractive and healthy babies attracted high bids, then the money raised through the auction would go to create incentives to encourage other parents to adopt less attractive and less healthy babies as well as to improve the conditions of institutionalized life for these children. In another scenario, respondents were assured that poor people would not be prevented from adopting children because the program would provide generous vouchers to all poor people who want to adopt children, thereby permitting them to compete with the affluent would-be adopters (again, explaining the concept of vouchers). In a third scenario, participants were assured that all less radical solutions to the problem of non-adopted children had been explored and proven inadequate.

How successful were we in eroding the taboo through this series of counterfactual thought experiments? Whereas nearly 90% of respondents initially objected to each proposal, opposition fell to approximately 60% in the hypothetical worlds with protection for the poor and unattractive and a guarantee that all other options had been thoroughly explored. For nearly half of the population, the term taboo is apparently too strong. They are willing to consider—in a quasi-utilitarian manner—conditions under which the benefits of the proposal might conceivably outweigh the costs. But for the other half of respondents, the term *taboo* still seems apt. They are prepared to ban exchanges that both donors and recipients believe would leave them better off and that, arguably would leave many third parties better off as well. In their eyes, taboo trade-offs remain an affront to the collective conscience of society. As one respondent protested: "what kind of people are we becoming?" Consequentialist arguments—whether they invoked egalitarian or efficiency concerns—missed the point: This woman's

sense of personhood required categorically rejecting the taboo trade-offs. In March and Olson's (1989) terminology, the logic of obligatory, not anticipatory, action governed her choices. She knew with existential certainty who she was and she knew that she did not fit into a social world that permitted such transactions.

These results also have suggestive political implications. Once a trade-off has become culturally defined as taboo, it is an uphill battle to render it respectable. Here is another reason why politicians in stable systems should be reluctant to initiate accusations of taboo trade-offs by opponents. Such accusations greatly complicate reaching integrative solutions to policy disputes.

Research on blocked exchanges need not be limited to big, controversial questions bearing on the limits of markets; it can also proceed at more personal and interpersonal levels of analysis that focus on the social, moral, and emotional significance people invest in personal possessions. It is a truism of microeconomic theory that, when income effects are small, there should be negligible differences between a person's maximum willingness to pay (WTP) for a good and minimum compensation demanded for giving up the same good (willingness to accept or WTA). [Markets cannot function well, they cannot clear, if people suddenly insist that what they now happen to own is worth twice what they thought it was prior to owning it.] Daniel Kahneman, Richard Thaler, and colleagues have, however, established in a series of compelling experiments that people often demand much more to give up an object that they already possess than they would be willing to acquire it, the definition of the endowment effect (Kahneman, Knetsch, & Thaler, 1990; Thaler, 1980). This effect is closely related to, and may even be a by-product of, the well known status quo bias (Samuelson & Zeckhauser, 1988), a preference for the current state of the world that biases consumers in their functional capacities as intuitive economists, against both buying and selling goods. The most influential explanation for these anomalies invokes an asymmetry of value that Kahneman and Tversky (1984) call loss aversion, the disutility of losing an object is greater than the utility of gaining it.

In a social environment strictly regulated by market-pricing principles, it does not matter to whom you sell something or from whom you buy it. The identities of buyers and sellers are interchangeable. Of course, in much of social life, objects take on deep emotional and symbolic significance as a function of the people or institutions from whom we obtain them. A commonplace object such as a pen may take on diverse meanings depending on whether the pen was a gift from a member of a group with one has communal sharing connections (e.g., a family) or a group to which one is bound by authority-ranking connection (such as a military or other institutional hierarchy) or from individuals to whom one is connected by equality-matching connections (reciprocity networks of friends and colleagues).

In a series of studies, McGraw, Tetlock, and Kristel (2003) have experimentally manipulated the relational context within which an object was originally acquired, and shown context to be a powerful moderator of the degree of discrepancy between WTP and WTA. Whereas the customary size of the endowment effect in the laboratory is a 2:1 ratio of WTA:WTP, the ratio soared for endowments charged with communal-sharing meaning. Figure 8.1 shows, for a related experiment, the cumulative percentage distributions of buying and selling prices, which are log transformed to preserve the original ordering of dollar values but still place them on a scale that incorporates all responses (no matter how extreme) into a reasonable display (McGraw & Tetlock, 2004). As expected by the endowment effect, the data almost perfectly exemplify stochastic dominance: Selling prices exceed buying prices throughout the range of values. The difference between the distributions, which is denoted by the shaded area, represents the disparity of WTP and WTA for each relational context. The area between the curves is greatest in the CS relationship and smallest in the MP relationship.

These results indicate that people draw a sharp normative line in the sand between particularistic and universalistic relationships—that is, relationships in which people care very much about the identity of the actor from whom they receive goods and services and relationships in which people are indifferent to the identity of actors. The ever-increasing discrepancies between WTP and WTA, moreover, tell only part of the story. We also documented a sharp surge in cognitive confusion and moral outrage—as well as principled refusals even to consider assigning dollar values—when people were asked to sell an object that had been acquired in communal-sharing, authority-ranking, or equality-matching relationships. Finally, as one would expect if basic norms were being breached, anger and resentment toward proposals to treat objects acquired in particularistic relationships as fungible was especially intense when observers believed that the person making the proposal was aware of the relationship history behind the possession of the object, and was insultingly implying that the pos-

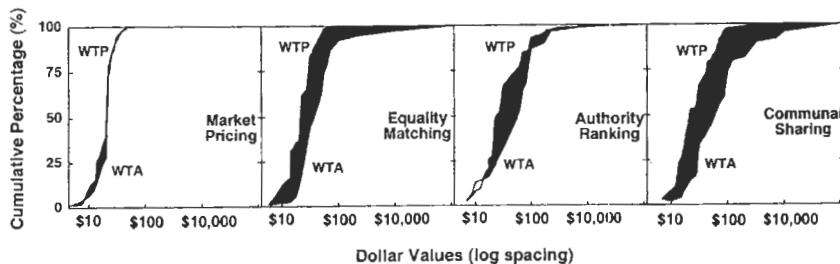


FIG. 8.1. Cumulative percentage distributions of buying prices (left edge of shaded areas) and selling prices (right edge of shaded area). The size of the shaded area represents the difference in buying and selling prices.

essor was the "sort of person" who would even think about selling a symbolically significant object.

### Forbidden Base Rates

Trade-off reasoning is widely viewed as a minimal prerequisite for economic rationality. Maximizing utility presupposes that people routinely factor reality constraints into their deliberations and explicitly weigh conflicting values.

We find just as solid a normative consensus that good intuitive scientists and or statisticians should use base rates as we do that good intuitive economists should confront trade-offs. Decision theorists routinely invoke Bayesian probability theory as the appropriate set of principles for aggregating base rate and case-specific information (Koehler, 1996; Tversky & Kahneman, 1982). And researchers routinely find that people often deviate from Bayesian prescriptions and fail to use base rates as much as they should. The most influential explanation for this underuse of base rates is that people over-rely on error-prone heuristics such as representativeness (Kahneman & Tversky, 1972).

The SVPM reminds us that over-relying on error-prone heuristics is not the only pathway to base-rate neglect. In a society committed to racial, ethnic, religious, and gender egalitarianism, forbidden base rates could include observations bearing on the disproportionately high crime rates and lower educational test scores of certain categories of human beings. People who use such 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 moral insensitivity.

Tetlock et al. (2000) report a set of experiments that demonstrate this point. In the red-lining experiment, an insurance executive must decide whether to sell policies in neighborhoods designated as high risk or as low risk. Sometimes the low risk towns have a disproportionately large African-American population; sometimes they do not; sometimes there is no reference whatsoever to race. The results indicated that respondents who thought of themselves as racially liberal were especially likely to favor selling policies across neighborhoods for the same price (ignoring the base rate) when the base rate was correlated with racial composition of neighborhood. In other words, a base rate that was acceptable to use when race was not an issue became unacceptable when race became an issue.

In a follow-up experiment, we examined what happens when people use a base rate to recommend a pricing policy for insurance policies and then later discover that the base rate either is or is not correlated with racial composition of neighborhood. It was hypothesized that racial liberals should find it especially morally threatening and dissonant to discover that they have recommended a pricing policy that works to the disadvantage of

a traditionally disadvantaged group. And this is exactly what happens. Racial liberals are more likely to change their pricing recommendations when they discover that the base rate that influenced those recommendations was correlated with race; and they were also more likely to engage in moral cleansing (express greater interest in volunteering for a civil rights cause as well as a cause unrelated to civil rights, helping to find a missing person).

### Heretical Counterfactuals

Good intuitive scientists are supposed to use the covariation principle in gauging the plausibility of causal hypotheses, and be on the lookout for causal variables that regularly covary with effects of interest (Tetlock & Belkin, 1996). When full covariation information is not available, and in life it rarely is, they should be willing to engage in counterfactual thought exercises in which they try to figure out what would have happened if a cost that had been present had been absent or vice versa.

The SVPM posits that cognitive theories of counterfactual reasoning need to acknowledge the emotionally charged normative boundaries that religious and political groups erect against what-if speculation. Particularly irksome are counterfactuals that apply the normal laws of human nature and of physical causality to heroic, larger than human, founders of the movement. Consider the reaction of the Ayatollah Khomeini to Salman Rushdie's heretical counterfactual that invited readers to imagine that the prophet Muhammed kept a company of prostitutes. For this transgression, the theocratic regime in Iran sentenced Rushdie to death (the ultimate expression of moral outrage).

In one experiment, Tetlock et al. (2000) examined the reactions of fundamentalist Christians to counterfactual conjectures that undermine faith in the "unique historical status" of Jesus Christ—the view that Jesus was God made man, divine yet also human, and that the events of his life as revealed in the New Testament gospels were the product of a divine plan and 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 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-intuitive scientist point of view, such counterfactuals are reasonable. They merely introduce causal schemata that virtually all of us apply reflexively in everyday life to a text that many people deem to be divinely inspired.

Tetlock et al. (2000) compared the reactions of low, moderate, or high scorers on the religious fundamentalism scale both to heretical counterfactuals (that apply regular causal principles to exceptional figures) and to

secular counterfactuals (which apply regular causal principles to regular people). They found virtually no connection between fundamentalism and reactions to secular counterfactuals; but powerful tendencies for fundamentalists to resist heretical counterfactuals, to express moral outrage toward them, to express emotions such as disgust, to ostracize members of their religious in-group who endorse such counterfactuals, and to engage in moral cleansing by expressing renewed intentions to attend church regularly.

It is also not hard to identify examples of heretical (or, at least, intensely controversial) counterfactuals that fall outside the religious domain. The historian John Charmley (1993) found one when he argued that Britain *would* have been better off *if* Lord Halifax rather than Winston Churchill had been Prime Minister in the summer of 1940 when Hitler was expressing an interest in a peace settlement that, according to Charmley, would have demanded only fairly modest concessions and allowed Britain to remain more viable as a great power than it was in the world in which Churchill defied Hitler. The economist Walter Williams may have generated a heretical counterfactual when he argued against reparations for the enslavement of American blacks on the grounds that ancestors of Africans who were enslaved and brought to America have far higher per capita incomes than they would have had if they had not been enslaved and had remained in Africa. The novelist Stephen Fry generated an heretical counterfactual when he concocted a science-fiction scenario in which the best way for the heroes to save lives in the 20th century is to go back in time and *prevent* the death of Adolf Hitler because, it turns out, in the absence of Hitler, a far shrewder Nazi takes power in Germany and succeeds in accomplishing what Hitler failed to accomplish.

### Closing Thoughts

There do indeed appear to be sharp qualitative breakpoints in social cognition, and conceding this point does not mean conceding that the quest for parsimonious theory is hopeless. The breakpoints are not arbitrary. They become predictable when we join a psychological theory such as the SVPM to Alan Fiske's (1991) cross-cultural taxonomy of relational schemata. Taboo tradeoffs become taboo when they violate cultural conventions on norms that place limits of market-pricing rules; base rates become forbidden when they impede implementation of egalitarian efforts to repair past wrongs, thereby thwarting an equality-matching reformist agenda; heretical counterfactuals become heretical when they show disrespect for established religious or political authority, mocking authority-ranking prerogatives and sensitivities. These proscriptions make good relational and normative sense even if they look bizarre from a strictly intuitive-scientist or

economist framework. Here we are reminded of a quotation from Immanuel Kant that Isaiah Berlin (1991) adopted as the title for one of his books: "From the crooked timber of humanity no straight thing can be built." Grand reformist plans to straighten out thought, to induce all of us to apply the same normative guidelines across content domains and social context, are fated to run into sharp resistance from ordinary people who insist on thinking, feeling, and acting in seemingly incommensurable ways across spheres of activity and who appear content with ad hoc "compartmentalizing" solutions when conflicts are called to their attention.

## REFERENCES

- Berlin, I. (1991). *The crooked timber of humanity: Chapters in the history of ideas*. New York: Alfred A. Knopf.
- Calabresi, G., & Bobbit, P. (1978). *Tragic choices*. New York: Norton.
- Charmley, J. (1993). *Churchill, the end of glory: A political biography*. London: Hodder & Stoughton.
- 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.
- Fiske, S., & Taylor, S. (1991). *Social psychology*. New York: McGraw Hill.
- Kagel, J., & Roth, A. (1995). *Handbook of experimental economics*. New Jersey: Princeton University Press.
- Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*, 1325-1348.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3*, 430-453.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341-350.
- Koehler, J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1-53.
- March, J., & Olson, J. (1989). *Rediscovering institutions*. New York: The Free Press.
- McGraw, A. P., & Tetlock, P. E. (2004). *Relational framing: The power to determine the acceptability of changes*. Working paper, Princeton University, Princeton, NJ.
- McGraw, A. P., Tetlock, P. E., & Kristel, O. V. (2003). The limits of fungibility: Relational schemata and the value of things. *Journal of Consumer Research, 30*, 219-229.
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*, 7-59.
- Tetlock, P. E. (1999). Theory-driven reasoning about possible pasts and probable futures: Are we prisoners of our preconceptions? *American Journal of Political Science, 43*, 335-366.
- Tetlock, P. E. (2000). Coping with trade-offs: Psychological constraints and political implications. In S. Lupia, M. McCubbins, & S. Popkin (Eds.), *Political reasoning and choice* (pp. 239-263). Berkeley: University of California Press.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: People as intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*, 451-471.
- Tetlock, P. E., & Belkin, A. (1996). *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives*. New Jersey: Princeton University Press.

- Tetlock, P. E., Kristel, O., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853-870.
- 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* (Vol. 8, pp. 25-51). Mahwah, NJ: Lawrence Erlbaum Associates.
- Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*, 39-60.
- 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.