

Social Functionalist Frameworks for Judgment and Choice: Intuitive Politicians, Theologians, and Prosecutors

Philip E. Tetlock
The Ohio State University

Research on judgment and choice has been dominated by functionalist assumptions that depict people as either intuitive scientists animated by epistemic goals or intuitive economists animated by utilitarian ones. This article identifies 3 alternative social functionalist starting points for inquiry: people as pragmatic politicians trying to cope with accountability demands from key constituencies in their lives, principled theologians trying to protect sacred values from secular encroachments, and prudent prosecutors trying to enforce social norms. Each functionalist framework stimulates middle-range theories that specify (a) cognitive–affective–behavioral strategies of coping with adaptive challenges and (b) the implications of these coping strategies for identifying empirical and normative boundary conditions on judgmental tendencies classified as errors or biases within the dominant research programs.

Once an esoteric specialty of a small cadre of cognitive psychologists, experimental research on judgment and choice has—through citation counts—become psychology’s leading intellectual export to the social sciences as well as to a host of applied fields. The influence of this research program has spread (critics might say “metastasized”) into such diverse domains as public opinion, international relations, finance, organizational behavior, marketing, medical diagnosis, and the law (Gilovich, Griffin, & Kahneman, 2002). Scholars with little else in common share a familiarity with foundational concepts in behavioral decision theory such as heuristics and biases, framing and choice, and the psychophysics of gain and loss functions.

The current article is, in one sense, testimony to the profound success of the heuristics-and-biases research program in extending cognitivist concepts into a vast array of disciplines. By advancing a succession of ingenious demonstrations, investigators faithful to the core tenets of the program have converted many skeptics, although not all (Gigerenzer, Todd, & the ABC Research Group, 1999), to the view that limitations of humans as information processors, especially the widespread reliance on inferential shortcuts, produce systematic and often surprisingly difficult-to-correct deviations from rationality ranging from base-rate neglect to myopic loss aversion (Kahneman & Tversky, 2001). The research program thus advances a core scientific value of psychology: the reduction of superficial diversity to unifying principles. However, parsimony is not a trump value. The current article is, in another sense, a cautionary comment on the perils of success. It proceeds from the contextualist premise that the more aggressively one extends abstract principles into new domains, the greater the risk of overextension (McGuire, 1983).

This article makes the case that the risk of overextension is substantial. Once placed in historical and philosophical context, the heuristics-and-biases program is revealed to be but one (albeit far and away the most successful) of an array of possible functionalist perspectives on judgment and choice.

Key Arguments

The key arguments—each to be expanded later in the article—include the following points.

Psychology Is Inherently Functionalist

William James (1890/1983) was the first of a long succession of theorists to observe that psychology is an inherently functionalist discipline. Even the most forbiddingly formal psychological explanations ultimately rest on functionalist assumptions about the goals that people try to achieve by thinking, feeling, and acting as they do.

Influential Functionalist Starting Points

The traditional functionalist starting points for research on judgment and choice have depicted people either as intuitive scientists (who seek causal understanding and predictive leverage; Kelley, 1971) or intuitive economists (who strive to maximize subjective utility; Edwards, 1962). There has also been long-standing interest in hybrid forms of functionalism that permit the ideal-type intuitive scientist’s value-neutral pursuit of truth to be deflected by various directional goals, such as signal-detection-theory concerns about avoiding one class of inferential error more than another (Friedrich, 1993) or the pursuit of intrapsychic objectives such as protecting self-esteem, restoring cognitive consistency, or affirming belief in a controllable world (cf. Baumeister, 1998; Kruglanski, 1990; Kunda, 1999; Tesser, 2000; Tetlock & Levi, 1982).

The Pivotal Role of Functionalist Metaphors

Functionalist metaphors play the role of “hard-core” assumptions in Lakatosian philosophies of social science: They set theo-

Philip E. Tetlock, Department of Psychology, The Ohio State University.
I gratefully acknowledge financial assistance from National Science Foundation Grant SBR-9696162 and helpful critiques from Hal Arkes and the social psychology faculty dinner group at The Ohio State University.

Correspondence concerning this article should be addressed to Philip E. Tetlock, who is now at Haas School of Business, University of California, Berkeley, California 94720-1900. E-mail: tetlock@haas.berkeley.edu

retical priorities and the stage for subsequent inquiry (Lakatos, 1970; Laudan, 1997). Metaphors are the psycholinguistic methods of choice for playing this role in psychological science because of their evocative power to capture superordinate features of similarity that cut across superficially diverse spheres of human activity (cf. Glucksberg & McGlone, 1999). For example, when speakers say, “all lawyers are sharks,” they do not mean that lawyers are fast swimmers or have fins and sharp teeth. They mean—and most people quickly understand them to mean—that lawyers are vicious, predatory, and tenacious. In a similar vein, when investigators launch research from the premise that people are intuitive scientists or economists, they do so because they believe it is valuable to highlight abstract commonalities underlying these highly organized, indeed, “professionalized,” forms of human activity and the far more loosely organized thoughts, feelings, and actions of ordinary people. Even if they do not always live up to the exacting standards of professional scientists, people try to make causal sense of events, to test hypotheses against reality, and to make choices that advance their interests. A natural focus for research becomes the following question: How well do ordinary people size up against well-defined normative benchmarks of rationality such as Bayesian-belief updating and expected-utility maximization?

Theorists need not, however, agree on the answers. Indeed, there are strong professional incentives to disagree and to articulate testable theories that, although faithful to the hard-core tenets of the program, are as distinctive as empirically plausible from conceptual rivals. Over the last 3 decades, the spectrum of portraits of intuitive scientists has ranged from rigorous Bayesians to glib top-of-the-head information processors, with an eclectic assortment of dual-process models between these end points (Chaiken & Trope, 1999). Although functionalist research programs are not, strictly speaking, falsifiable (there is, in principle, no end to the conceptual tinkering in which determined defenders can engage to preserve core premises), these programs are exhaustible—a point reached when the patience of investigators and professional gatekeepers collapses.

Social Functions of Thought

The approach advocated here represents a sharp departure from psychological theories that have traditionally stressed the intrapsychic functions of judgment and choice and that have placed the isolated Cartesian thinker at the center of inquiry (Gergen, 1986). The focus shifts to the social functions of thought and to embeddedness of human beings in relations with other people, institutions, and the broader political and cultural environment. Building on the observations of classical sociologists—Durkheim, Weber, Mead, Parsons—as well as on more recent advances in cross-cultural psychology, the analytical starting point becomes the psychological prerequisites of social order, the properties that people must possess to cope with the challenges of life within intricately interdependent collectivities in which they alternate, often in rapid succession, between being targets of the accountability demands from others and being the sources of accountability demands on others (cf. Aberle, Cohen, Davis, Levy, & Sutton, 1951).

Working within this overarching framework, the current article proposes three social-functional theories especially relevant to

psychological debates on human rationality and to current classifications of error and bias in judgment and choice. Each theory is designated by a guiding metaphor that captures the essence of a particular functional orientation that the vast majority of people can, under the right activating conditions, adopt toward the social world. As objects of accountability pressures from others, people strive—like intuitive politicians—to maintain positive social identities vis-à-vis significant constituencies in their lives. A key function of thought becomes internalized dialogue in which people try to anticipate objections to possible lines of action and to craft responses. As transmitters of accountability pressures onto others, people try—like intuitive prosecutors—to detect cheaters and free riders who seek the benefits but shirk the responsibilities of membership in the collective. A key function of thought becomes closing loopholes in accountability regimes that unscrupulous intuitive politicians might otherwise exploit. Finally, as beings capable of reflecting on the accountability regimes within which they live, love, and work, people are posited to be intuitive theologians who have an existential need to believe that the rules governing their social world are not just the arbitrary preferences of currently dominant interest groups but rather are anchored in sacred values that confer legitimacy on collective practices. A key function of thought becomes protecting sacred values from secular encroachments.

Intuitive Politicians, Theologians, and Prosecutors

Just as it is possible to advance an array of psychological models of how people function as intuitive scientists or economists, the same can be said for the intuitive politician, theologian, and prosecutor. It is possible to posit intuitive politicians of varying skill in anticipating the objections of key constituencies, intuitive theologians of varying forgiveness toward secular encroachments on sacred values, and intuitive prosecutors of varying punitiveness toward norm violators. The mere possibility of doing something does not, however, demonstrate the scientific desirability of doing it. The scientific community confronts a tricky effort-allocation decision in which it must weigh the risks of excessive closed-mindedness—of failing to diversify its portfolio of guiding assumptions and of thereby neglecting key aspects of human nature that traditional assumptions slight—against the risks of excessive open-mindedness—of aiding and abetting the proliferation of frivolous frameworks that divert scarce scientific talent from proven to unproven approaches.

Payoffs of Diversification

Advocates of diversifying the discipline’s portfolio of functionalist frameworks need to reassure skeptics that diversification yields tangible empirical and conceptual payoffs. Functionalist frameworks need to move expeditiously from vague metaphorical pontificating to testable process models that specify the adaptive challenges that activate functionalist mind-sets, the goals linked to these mind-sets, and the coping strategies used to achieve these goals. The resulting social functionalist theories should identify empirical boundary conditions on judgmental tendencies classified as errors or biases within the dominant epistemic or utilitarian research programs—that is, conditions under which judgmental shortcomings are attenuated or amplified by social contextual

manipulations that create adaptive challenges that, in turn, activate sociocognitive processing goals and coping strategies. The resulting social functionalist theories should also identify normative boundary conditions that specify when it is reasonable to challenge extant classifications of judgmental tendencies as errors or biases. Social functionalism raises the possibility of normative Gestalt shifts: Effects that look dysfunctional within an intuitive-scientist or intuitive-economist framework sometimes look quite adaptive within a political, prosecutorial, or theological framework.

The remainder of this article develops the preceding arguments. The next section directly addresses skeptics who worry that the approach recommended here is too permissive of runaway functionalist speculation. It lays out the overarching themes that unify and impose common standards on the three social functionalist theories. The following three sections are organized around the politician, theologian, and prosecutor programs. Each section begins by characterizing the functionalist premises of the incipient research program: the core goals, adaptive challenges, and coping strategies attributed to human beings. Each section then delineates a conceptual range of testable theories consistent with these premises, advances a specific theory that falls within the program's tolerance zone of opinion, derives testable hypotheses, and evaluates relevant evidence. The concluding section addresses influential objections to social functionalism and stresses the need for a distinctive style of psychological theorizing pitched at a societal level of analysis that links intrapsychic processes explicitly to the political and institutional contexts within which people live and work.

Conceptual Guidelines for Constructing Social Functionalist Theories

Screening Tests

Skeptics have long suspected that functionalism is inherently vague and tautological (Boring, 1963). If a thought, feeling, or action appears maladaptive within one functionalist framework, one can always concoct an alternative functionalist rationale. Without denying the kernel of truth to this characterization of earlier forays into functionalism, the current article implements several checks against undisciplined functionalist speculation. Social functionalist research programs must initially pass three screening tests and should eventually be held to account for satisfying a fourth:

1. The program should be organized around a postulated function that highlights a fundamental adaptive challenge that arises whenever human beings are locked together in complex patterns of interdependence;
2. The program should stimulate testable theories that provide reasonably specific answers to questions concerning the adaptive challenges that activate functionalist mind-sets, the goals that people are trying to achieve, and the coping strategies that people use in pursuit of the goals;
3. The testable theories should stimulate discoveries of empirical and normative boundary conditions that would have been unlikely had theorizing been confined to the dominant functionalist frameworks;
4. The testable theories should ultimately address the outstanding problem of integrating the psychosocial functions of judgment

and choice by specifying how people resolve cross-functional conflicts.

Adaptive Challenges

The social functionalist theories advanced here depict human beings whose thought processes are organized around three adaptive imperatives of collective life. First is the necessity of coping with accountability demands from others with whom one is intertwined in reciprocal networks of interdependence. Each person is embedded in a matrix of accountability relationships that specify who must answer to whom, for what, and under what ground rules. To survive and prosper in any social unit, people (in their capacity as intuitive politicians) must possess a reasonably reliable mental compass for navigating the self through these sometimes Byzantine role–rule structures (Semin & Manstead, 1983; Stryker & Stratham, 1985).

The second imperative is social control, the need to place accountability demands on others who might be tempted to derive the benefits of collective interdependence without contributing their fair share or without respecting other aspects of the prevailing role–rule regime (Cosmides & Tooby, 1994). To survive and prosper in any social unit, intuitive prosecutors need to protect themselves from exploitation but must be careful to avoid overzealousness in the pursuit of wrongdoers by respecting culture-specific norms for defining the boundaries of self and the concept of harm (Scott & Lyman, 1968).

The third imperative is legitimation of the accountability ground rules. People need moral backstops (M. Lerner & Lerner, 1981). To stave off anomie, alienation, and even existential despair, intuitive theologians need to believe that the prevailing accountability and social control regime is not arbitrary but rather flows naturally from an authority that transcends accidents of history or whims of dominant groups (Durkheim, 1925/1976). Framed in this fashion, the social functionalist research programs are an integral part of a societal social psychology that must ultimately be consistent with, but is not reducible to, the intrapsychic levels of analysis at which most theory is now pitched. The theories can thus be viewed as distinctive members of an already large family of identity-maintenance and social control formulations in the literature (cf. Haidt, 2001; Hamilton, 1980; Schlenker, 1985; Steele, 1988; Tesser, 2000).

Adequate Specification of Hypothesized Functions

The social functionalist theories also fit a common explanatory template. Each specifies switch-on rules—the types of adaptive challenges that activate the metaphorical mind-set as well as the types of goals and coping strategies that become increasingly cognitively accessible, and are seen as increasingly applicable, under conditions of mind-set activation. Each also specifies switch-off rules—the types of negative feedback from the environment that deactivate the hypothetical mind-sets and their accompanying goals and coping strategies.

The self-regulatory innards of social functionalist theories owe an obvious debt to ideas drawn from the cybernetic control of complex systems (Higgins, 1989). Each theory presupposes an autonomous self-regulatory system that is activated by adaptive challenges and that relies on discrepancy-reducing negative feed-

back mechanisms designed to minimize the gap between current condition and desired end state.

The adaptive challenges for activating the intuitive-politician mind-set take the form of accountability pressures and demands: the knowledge that one is under the evaluative scrutiny of important constituencies in one's life who control valuable resources and who have some legitimate right to inquire into the reasons behind one's opinions or decisions. This knowledge activates the goal of establishing or preserving a desired social identity vis-à-vis these constituencies. In pursuit of this goal, people may resort to a wide range of coping strategies—from attitude shifting to preemptive self-criticism to defensive bolstering—that shut off only when people believe that they have achieved the desired end state or that the costs of pursuing that end state have become prohibitive.

The adaptive challenges for activating the intuitive-theologian mind-set take the form of perceived threats to sacred values, values that—by community consensus—are deemed beyond quantification or fungibility. One common type of threat arises whenever people believe that they or others have subjected sacred values to the same crass mental operations routinely performed on secular values (e.g., computing the dollar value of one's family or friendships or loyalty to a collective). Taboo violations of this sort activate the goal of reaffirming individual and group commitment to the sacred value. In pursuit of this goal, people may resort to a mixture of coping strategies—including moral outrage and moral cleansing—that shut off only when people are reassured that the secular encroachment has been rebuffed or when people are convinced that the costs of maintaining the social illusion of infinite importance are prohibitive.

The adaptive challenge for the intuitive prosecutor is the perception that norm violation is both common and commonly goes unpunished. This perception activates the goal of tightening standards of accountability as well as the coping strategies of looking for grounds for rejecting excuses and justifications for norm violation and for ways of closing loopholes. The process disengages only when people believe that the collectivity is no longer under siege or when people sense that the pursuit of social order has gone too far and has become oppressive.

Discovery of Empirical and Normative Boundary Conditions on the Error-and-Bias Portrait of Human Nature

From the standpoint of cognitive theories of judgment and choice, all this effort to specify the self-regulatory dynamics of social functions is of little interest if the resulting theories do not generate testable predictions that shed light on either empirical or normative boundary conditions on the error-and-bias portrait of human nature. Accordingly, the social functionalist theories advanced here are evaluated by a common conceptual standard: their capacity to shed light on social contextual conditions under which one should expect to observe the amplification or attenuation of effects labeled as biases or should question the very classification of effects as biases. For instance, the social contingency model of how people function as intuitive politicians identifies not only distinctive strategies of coping with accountability but also the implications of particular coping strategies (such as preemptive self-criticism) for when the amplification of biases (rooted in over-reliance on simple, easy-to-execute, heuristics) or the attenuation

of biases (rooted in fear of being second guessed) should be expected. The sacred-value-protection model of how people function as intuitive theologians not only identifies moral outrage and cleansing as coping responses to violations of sacred values but also examines the implications of those coping strategies for trade-off aversion and base-rate neglect. The fair-but-biased and occasionally erratic model of the intuitive prosecutor not only identifies strategies of coping with social deviance but also explores the implications of these strategies for debates over the normative status of the fundamental attribution "error" and the severity "bias."

The Challenges of Modeling Cross-Functional Conflicts

There is already considerable evidence consistent with key empirical-and-normative-boundary-condition predictions of each social functionalist theory. There is much less evidence, however, for guidance in the next phase of theory development: the integration of the social and cognitive functions of judgment and choice. Little is known about when the pursuit of one self-regulatory objective has negative externalities that impede attainment of other objectives. In political theory, negative externalities offer the classic justification for empowering government to regulate the conduct of its citizens (to paraphrase John Stuart Mill, your freedom to swing your arm ends where my nose begins). Likewise, in psychological theory, negative intrapsychic externalities have offered the classic justification—since Freud—for some system of mental governance or executive control, to arbitrate conflicts between modularized functions.

One common form of motivational conflict arises when social functions circumscribe cognitive and material-utilitarian functions. The scientist's quest for the truth is subordinated to the politician's need to craft compelling justifications; the economist's calculus of self-interest is checked by the theologian's obligations to communally shared values; the scientist's curiosity about the causes of behavior is qualified by both the prosecutor's determination to hold rule breakers responsible and the theologian's desire to block inquiry that demystifies objects of veneration. The flip side forms of conflict arise when social functions are circumscribed by cognitive and material-utilitarian functions. The intuitive theologian's effort to treat certain values as beyond compromise runs aground the intuitive economist's recognition of resource constraints. The intuitive politician's efforts to escape accountability, or the intuitive prosecutor's efforts to impose accountability, collide with the intuitive scientist's recognition of the implausibility of certain claims. The closing section of the article returns to these subtle issues of mental balancing.

Intuitive-Politician Research Program

People resemble intuitive politicians in that they are accountable to a variety of constituencies, they suffer consequences when they fail to create desired impressions on key constituencies, and their long-term success at managing impressions hinges on their skill at anticipating objections that others are likely to raise to alternative courses of action and at crafting accounts that preempt those objections. Moreover, just as there is room for diversity in how theorists characterize people as intuitive scientists or economists, there is room in the intuitive-politician research program for dif-

ferent approaches to developing that metaphor. Some theorists may opt to portray people as ruthless Machiavellian schemers (who will do whatever it takes to win), whereas others might portray people as bumbling impression managers (who frequently miscalculate social impact because of cognitive biases rooted in heuristics [Gilovich, Medvec, & Savitsky, 2000] or in transference [Andersen & Glassman, 1996]).

A Testable Model of How People Function as Intuitive Politicians

The social contingency model (SCM) of judgment and choice falls toward the midpoint of the Machiavellianism continuum (Tetlock, 1992, 1998, 1999; Tetlock & Lerner, 1999). It assumes that people seek the approval and respect of the key constituencies in their lives. But the model eschews a purely strategic conception of human nature. The SCM can be summarized in seven propositions, described below.

The universality of accountability. People do some things alone, but it is difficult to escape evaluative scrutiny in an intricately interdependent society. Escape arguably becomes impossible when self-accountability—the obligations that most human beings, excluding psychopaths, feel to internalized representations of significant others who keep watch over them when no one else is looking—is taken into account (Mead, 1934). The SCM adopts the symbolic-interactionist tenet that accountability is a ubiquitous feature of everyday life that links individuals to institutions by reminding people of the need to be cognitively equipped with good reasons for their opinions in the event that observers request or demand accounts.

The audience-approval motive. Whether it is a function of natural and sexual selection or of relentless social reinforcement from infancy onward, human beings have a deep-rooted need for social approval (Baumeister & Leary, 1995). Moreover, people are generally risk averse and more concerned with falling in the esteem of others than with missing opportunities to rise. Failure to create desired impressions on a target audience triggers an aversive state of emotional arousal—embarrassment or shame—that people try to reduce by means of a host of impression-management tactics (cf. Keltner & Buswell, 1997; Schlenker, 1985).

Motive competition. Although the SCM assigns a central role to the approval motive, it does not elevate that motive to sovereign regulator of all conduct. Social psychology has had too many disappointing flirtations with monistic theories that anointed master motives (Allport, 1968/1985), and there is no point in replicating the mistakes of early role theorists who treated approval maximization as the sociological equivalent of utility maximization. The SCM identifies four classes of motives that circumscribe the approval motive, including the following:

1. Achieving cognitive mastery of the causal structure of the environment (emphasized by classic attribution theory, which portrayed people as cognitively diligent intuitive scientists; Kelley, 1971);
2. Minimizing mental effort and achieving rapid closure (emphasized by later theories of social cognition, which portrayed people as lazy or cognitively miserly intuitive scientists; Kruglanski & Webster, 1996);

3. Maximizing the benefits and minimizing the costs of relationships (emphasized by evolutionary and exchange theories; Kelley & Thibaut, 1978; Ridley, 1994);

4. Asserting personal autonomy and integrity by reaffirming private convictions (emphasized by theories of ego and moral development, by self-affirmation variants of dissonance and social identity theory, and by reactance theory; Schlenker, 1985).

Generative rules for predicting coping strategies. A laundry list of motives does not a theory make. It is necessary to specify how motives can be amplified or attenuated by the interpersonal context. The conceptual formula for generating predictions fits the classic template for expectancy-value theories in psychology: Identify situational and dispositional moderators of the perceived importance of identity objectives and the perceived feasibility of achieving those objectives. The SCM posits that the likelihood of coping strategies should wax or wane as a multiplicative function of independent variables that influence (a) the cognitive accessibility and perceived applicability of dimensions of social identity and (b) the perceived feasibility, costs, and benefits of claiming a given identity.

Coping-strategy predictions. Every request for justification raises the question of how it will be received by either external constituencies (that one sees as separate from one's self) or internalized ones (familiar voices inside the head that one accepts as integral parts of oneself). In each case, the attributions that observers make define one's reputation: If I do or say X, will others view me as cooperative or confrontational, craven or candid, flexible or rigid, principled or opportunistic . . . ? The SCM distinguishes an array of identity-defining choices that arise in accountability predicaments, but two stand out as most relevant to theories of judgment and choice: preemptive self-criticism and defensive bolstering. Preemptive self-criticism is the predecisional strategy of anticipating plausible objections of would-be critics, factoring those objections into one's mental representation of the problem, and reaching a complex synthesis that specifies how to deal with trade-offs. The SCM asserts that people will perform this cognitive work only when low-effort strategies have been systematically cut off by the logic of the situation. Thus, preemptive self-criticism should be especially likely when people are unconstrained by past commitments, the views of the audience are unknown or known to conflict, the audience is knowledgeable and powerful, and the audience possesses a legitimate right to inquire into the reasons behind opinions or decisions (J. S. Lerner & Tetlock, 1999). Defensive positioning is the postdecisional mirror image of preemptive self-criticism. Accountability still motivates thought, but that thought now takes self-justifying rather than self-critical forms. People devote mental energy to generating reasons why they are right and would-be critics wrong. Self-justification should be most pronounced when people are accountable for prior public commitments, when people fear that previous decisions cast doubt on their competence or morality, and when people know that they cannot plausibly deny responsibility but suspect that they can put a more positive spin on earlier decisions (Tetlock, Skitka, & Boettger, 1989).

Empirical boundary conditions on the error-and-bias portrait of human nature. Skeptics could concede the previous predictions but still insist that none bears on the foveal concerns of the traditional research programs: Namely, when and why do judgment and choice deviate from economic and scientific standards of

rationality? Skeptics could just dust off the disciplinary division of labor between psychology and the social sciences proposed by N. E. Miller and Dollard (1941): The mission of cognitive research programs is to shed light on how people think, whereas the politician research program sheds light merely on what people think about and on when they will say what is on their minds.

To be sure, certain strategies of coping with accountability, such as attitude shifting, can be socially significant but cognitively trivial. The SCM implies, however, that the time-honored process-content division of labor crumbles in many settings. Self-presentation triggered by accountability can interact in complex ways with putatively basic cognitive processes. A central function of private thought is preparation for public performances—a proposition that echoes early 20th-century theorists such as Vygotsky (1934/1978) and Mead (1934). Thought frequently takes the form of internalized dialogues in which people gauge the justifiability of options by imagining conversations in which accounts are exchanged, debated, revised, and evaluated. Whether these imaginary conversations amplify or attenuate judgmental tendencies deemed by the expert community to be biases depends on both the substantive content of the internal dialogues and the cognitive-affective sources of the alleged biases.

This, the sixth proposition of SCM, asserts that the effect of accountability hinges on its power to prime metacognitive ideals that people in various subcultures have, to varying degrees, internalized—ideals such as “be self-critical,” “attend to all conceivably relevant information,” “stay the course,” or “don’t cry over spilt milk.” Certain prescriptive ideals direct people to think in ways that, given the problem at hand, correct judgmental tendencies widely viewed in the expert community as biases or errors. For example, forms of accountability that activate the metacognitive maxim “anticipate plausible objections” should attenuate biases that are rooted in overreliance on simple, easy-to-execute heuristics. Following this maxim should also, however, amplify biases rooted in indecisiveness, risk aversion, or fear of being second guessed. Conversely, forms of accountability that activate the flip-side maxims—stay the course and defend yourself—should have mirror-image effects, amplifying biases such as sunk-cost justification but attenuating biases linked to premature abandonment of sound policies (such as myopic loss aversion; Benartzi & Thaler, 1995). Finally, for certain inferential problems, people may either have no pertinent prescriptive ideals or contradictory ones. Here the SCM predicts null effects.

Normative boundary conditions on the error-and-bias portrait. Inherent in the politician metaphor is the notion of contending factions with opposing conceptions of rationality and morality. The SCM is cognitively egalitarian: It treats academic commentators on rationality as but one of a number of constituencies that have staked out positions on how people should think, feel, and act in particular settings. Disagreements will arise over the classification of judgmental tendencies as errors or biases whenever observers hold conflicting beliefs about either (a) the degree to which judgmental tendencies that deviate from scientific or economic standards of rationality have offsetting social or political benefits or (b) the usefulness of judgmental tendencies in facilitating attainment of agreed-on goals.

Evidence Bearing on Key Predictions of the Model

When does the “political” process of computing the justifiability of response options begin to impinge on the putatively more basic cognitive processes of maintaining a stable, internally consistent world view at acceptable cost in mental effort? Experimental evidence bears out the fifth proposition of the SCM (see *Empirical boundary conditions on the error-and-bias portrait of human nature* above), that the goals of minimizing cognitive effort, dissonance, and ambiguity clash most directly with the political goals of social identity protection when decision makers confront accountability demands for which they do not have a ready-at-hand or dominant response—that is, when decision makers have not made any prior commitments to a position and are accountable either to a single audience with unknown views (Tetlock, 1983a) or to multiple audiences who hold contradictory views and who possess strong arguments in defense of their positions (Green, Visser, & Tetlock, 2000). In such cases, people respond by engaging in more reflective and self-critical patterns of thinking in which they attempt to anticipate potential objections of reasonable critics and to construct defensible integrative positions (Tetlock, 1992). Content analysis of thought protocols reveals not only that people generate more thoughts but also that thinking tends to take a dialectical “on the one hand” and “on the other hand” rhythm, as though people were sending the implicit message: “I may believe X, but I’m no fool, and I recognize counterarguments Y and Z.”

Cognitive effort directed to protecting one’s social identity as rational need not, however, always promote open-mindedness. Postdecisional accountability often reminds people of metacognitive maxims such as “stay the course” and encourages self-justifying patterns of thinking that have been linked to amplification of dissonance effects such as escalating commitment to sunk costs and trade-off avoidance (Staw, 1980; Tetlock et al., 1989).

The sixth proposition of the SCM posits that whether academic observers applaud or deplore the impact of accountability will hinge on the correspondence between the informal theories that ordinary people hold about good decision making (the metacognitive maxims primed by accountability and problem context) and the more formal theories that academic observers hold about rationality. The SCM predicts that accountability of the “right” sort (which motivates people to make judgments they deem most justifiable on careful reflection) will attenuate bias when there is convergence between formal and informal theories. The best candidates for de-biasing, therefore, should be response tendencies that satisfy two criteria. First, unaccountable decision makers are making suboptimal judgments because they have not been motivated to think sufficiently carefully and self-critically about their own mental processes. Second, decision makers intuit that the normative expectation is to engage in sober second thought when a neutral authoritative figure in an academic setting requests a justification before they have taken any position. Under such conditions, people should bring their judgments into closer correspondence with those of the expert community, the academic arbiters of normative standards. Examining the evidence, J. S. Lerner and Tetlock (1999) find considerable support for this hypothesis. Accountability that encourages preemptive self-criticism tends to have the following effects.

1. It reduces the overattribution effect (Tetlock, 1985). Self-critical thinkers are more reticent about drawing dispositional

attributions from conduct constrained by context in the “low-choice” conditions of essay-attribution experiments, but they are no more reticent about drawing such conclusions in the “high-choice” conditions.

2. It reduces primacy and recency effects (Kennedy, 1993; Tetlock, 1983b). Self-critical thinkers are more cautious about drawing conclusions from incomplete evidence and more willing to change their minds in response to evidence.

3. It improves the calibration of the subjective probabilities that people attach to their predictions (Siegel-Jacobs & Yates, 1996; Tetlock and Kim, 1987). Moreover, these improvements in “calibration” can be achieved at minimal cost in “resolution” (the variance of proportion of correct predictions across confidence categories)—evidence that people are not indiscriminately scaling down all confidence estimates but rather are carefully weighing evidence.

4. It reduces the incompatibility bias—the tendency for negotiators to assume incorrectly that relationships are zero sum and that the other party’s interests are completely opposed to their own (L. Thompson, 1995).

5. It reduces numerical-anchoring effects by encouraging subjects to consider additional evidence and to revise estimates. This effect holds up as long as subjects are not under time pressure or cognitive load (Kruglanski & Freund, 1983).

6. It decreases reliance on noncompensatory decision rules in favor of compensatory ones that explicitly acknowledge trade-offs (Kahn & Baron, 1995; Tetlock, 1983a).

7. It decreases the influence of incidental affect evoked in one situation on judgments made in a completely unrelated situation (J. S. Lerner, Goldberg, & Tetlock, 1998).

8. It decreases the influence of sunk costs on future investment decisions by increasing adherence to self-imposed earlier limits on amounts to be invested and by increasing the salience of decision-making procedures (as opposed to outcomes) in evaluation (Simonson & Staw, 1992).

9. It reduces illusory correlations and improves covariation assessment, in part, by motivating more complex inferential strategies (Murphy, 1994).

10. It increases correspondence between the judgment strategies that people claim to be using and the strategies that statistical models impute to them (Hagafors & Brehmer, 1983).

The SCM predicts bias amplification whenever the implicit theories of good judgment held by research participants diverge from those held by the researchers. Here, efforts to put on the cognitive equivalent of one’s “Sunday best”—to follow the salient metacognitive maxims—lead to judgments that depart ever more markedly from those of the expert community. The following lines of evidence are consistent with this hypothesis.

Ambiguity aversion refers to the tendency of decision makers to prefer less ambiguous alternatives when they are given a choice between options that differ only in uncertainty about the probabilities of relevant outcomes (Curley, Yates, & Abrams, 1986). Expected value is held constant, so this preference is, strictly speaking, a bias from a microeconomic point of view. Thought protocols suggest that accountability increases preferences for well-defined probabilities, in part because participants feel that, in the event of failure, it would be more difficult to justify having selected the ambiguous options (people are supposed to look before they leap).

In the *compromise effect*, a product gains attractiveness simply by virtue of becoming the middle option in a choice set (Simonson, 1989). Accountable participants were more likely to select the compromise product because they believed that it was more defensible than options that were clearly superior on one dimension but clearly inferior on another.

In the *attraction effect*, a relatively inferior alternative (Brand X) is added to a set of closely competing options (Brands A and B), thereby increasing the attractiveness of the preexisting option that happens to be superior to Brand X on all evaluative dimensions (Simonson, 1989). Again, accountable subjects were drawn to dominating options, apparently because they thought that the dominating options were less vulnerable to criticism.

The *dilution effect* refers to the tendency of observers to lose confidence in the predictive power of diagnostic evidence when that evidence is mixed with irrelevant or nondiagnostic evidence (Tetlock & Boettger, 1989). Motivating self-critical thought can induce people to try too hard to discern relevance amidst irrelevance. Rather than zeroing in on the one relevant cue (e.g., number of hours studied as a predictor of grade point averages), accountable subjects often struggle to integrate irrelevant cues into their schematic representations of the target individuals (vain searches for indicators of academic success, such as, “If he’s never dated anyone longer than 2 months, he may be emotionally unstable,” or, “Tennis players are usually physically and mentally alert”). In environments with unfavorable signal-to-noise ratios, accountability can send people off on inferential wild-goose chases.

The status quo occupies a privileged position in decision making. It is possible to reverse preferences for alternatives by holding expected value constant and arbitrarily designating one option as the status quo and the other as change (Samuelson & Zeckhauser, 1988). The “change” option is held to a higher standard of proof—and accountable subjects are particularly prone to do so when abandoning the status quo creates identifiable losers who are likely to complain (Tetlock & Boettger, 1994).

In cases where accountability has had no effect, there is presumptive evidence either that people lack intuitions about how to solve the problem (the statistically untutored do not know Bayes’s theorem) or have conflicting normative intuitions (some of which coincide with formal models of procedural rationality and some of which diverge). The proverbial file drawer of null-hypothesis results contains a sizable set of studies in which accountability has had no effect on judgmental biases, including insensitivity to base-rate information, giving more weight to causal as opposed to “merely” statistical relationships between variables, preference reversals as a function of choice-versus-matching elicitation procedures, insensitivity to sample size, and the conjunction fallacy (cf. Simonson & Nye, 1992).

Normative Boundary Conditions

Turning to the normative-pluralism postulate of the SCM, there are both good arguments and suggestive evidence that what looks like an error or bias from an intuitive-scientist or intuitive-economist perspective will often look quite defensible to observers working within an intuitive-politician framework:

Ambiguity aversion. Decision makers may be right that they will incur greater blame for choosing options with ambiguous probabilities than those with well-defined probabilities (holding

expected values constant). They just need to imagine the recriminations from constituents if they choose an ambiguous gamble (with a probability of winning between 10% and 90%) over a gamble with a well-defined probability (50/50) and if the ambiguous gamble turns out to have had only a 10% chance of positive pay-off (Taylor, 1995). Unaccountable decision makers will feel “suckered”; accountable ones will feel the same way but, in addition, will feel labeled as “suckers” by others.

Attraction effect. Simonson (1989) argued that people believe they have a more persuasive justification for choosing A over B when A but not B dominates the irrelevant alternative C. Although introducing C does not alter the offsetting strengths of Options A and B (and being swayed by the introduction of C puts one in violation of the rational-choice axiom of the independence of irrelevant alternatives), invoking the argument “A dominates C but B does not” may be a quite compelling justification for many audiences.

Dilution. Decision makers infer that they will get along more smoothly with their prospective conversational partners if they make special efforts to detect relevance in the seemingly irrelevant information that the partners have bestowed on them. This interpretation (which differs from the representativeness-heuristic account that portrays people as flawed statisticians) gains credibility from evidence that accountability amplifies the dilution effect only when participants believe that the conversational norm of relevance holds and the interaction partner views the information provided as germane (Tetlock, Peterson, & Lerner, 1996). Indeed, dilution may be one of several “biases” to which conversational norms contribute. Others include underutilization of base rates and the conjunction fallacy (Schwarz, 1994).

Disjunction effect. Shafir, Simonson, and Tversky (1993) demonstrated in a series of experiments the sensitivity of decision makers to the ready availability of reasons for their choices. In the disjunction effect, people flout a basic economic axiom of rational choice, the sure-thing principle, and sacrifice resources to delay making a choice until an uncertainty is removed, notwithstanding that they believe they would choose the same option no matter which way the uncertainty was resolved. This reasoning takes the paradoxical form “I’ll go on a Hawaiian vacation if I fail the exam (to console myself) and if I pass (to celebrate), but won’t decide until I know the outcome of the exam.” Reason-based choice reaffirms the core SCM tenet that the mind is configured for coping with accountability challenges. The guiding maxim seems to be: Don’t do anything consequential unless you have a sensible-sounding reason for doing it.

Although the previous examples were not politically controversial, the politician metaphor suggests that competing ideological factions will often have different views of the benefits and costs of judgmental tendencies. Drawing on the literature on political ideology (Sniderman & Tetlock, 1986), the SCM posits that egalitarians and authoritarians will attach systematically different “normative spins” to an array of decision-making practices. In a survey of managers, Tetlock (2000a) has documented that these disagreements are closely coupled to assumptions about human nature (authoritarians were more suspicious of human nature and thought it prudent to hold others tightly accountable), about the nature of the decision-making environment (authoritarians had more faith in the efficacy of fast-and-frugal heuristics), and about effective leadership (authoritarians applauded simple managerial styles that

projected can-do confidence but were dubious about both the cognitive and impression-management value of preemptive self-criticism). Effects such as overattribution and overconfidence can polarize along parallel political and psychological lines.

Intuitive Theologians

From Aristotle to Nietzsche, philosophers have posited a deep-rooted human need to believe that the moral codes that regulate our lives are not arbitrary social constructions but rather are endowed with transcendental significance. These bedrock values provide reassuringly absolute answers to unsettling questions about the meaning of existence and the ends to which we should devote our lives. Moral communities strive to insulate these sacred values from secular contamination. What looks like rigidity within other functionalist frameworks suddenly becomes commendable: the principled defense of the sacred from encroachments by powerful societal trends toward science, technology, and the calculus of capitalism (and attendant pressures to pursue inquiry wherever it leads and to translate all values into a utility or monetary metric). This emphasis on unswerving principles and faith distinguishes the theologian research program from the economist and scientist programs (which place a premium on cognitive agility in pursuit of epistemic or market goals) as well as from pragmatic variants of the politician program (which value agility in self-presentation).

A Testable Model of How People Function as Intuitive Theologians

Whereas theories of the intuitive scientist and economist vary on a rationality dimension and theories of the intuitive politician on a social effectiveness dimension, theories within the intuitive-theologian program vary on a ferocity continuum anchored at one end by brutal theocrats fanatically devoted to sacred values and at the other end by flexible pragmatists reconciled to the necessity of compromise and perhaps even the “hypocritical” necessity of concealing those compromises. The middle-range option adopted here falls near the middle of this ferocity continuum. The sacred-value-protection model (SVPM) can be captured in four sets of propositions.

Observer-based predictions. The SVPM defines a sacred value as any value toward which a moral community proclaims, at least in rhetoric, an unbounded or infinite commitment. Building on Durkheim’s (1925/1976) observations of how people respond to affronts to the collective conscience that disrupt the normative equilibrium, the SVPM predicts that discovery that members of one’s community have compromised sacred values triggers an aversive arousal state, moral outrage, which has cognitive, affective, and behavioral components: harsh trait attributions to norm violators, anger and contempt, and enthusiastic support for both norm and meta-norm enforcement (punishing both violators and those who shirk the sometimes burdensome task of punishing violators).

The SVPM also predicts derogation as a function of mere contemplation. Traditional cognitive accounts trace the difficulty people have in making trade-offs between secular values such as money and convenience, and sacred values, such as love and loyalty, to the incommensurability problem—the absence of a common metric for comparing secular and sacred values. The

SVPM insists, however, that people find such trade-offs not only cognitively confusing but morally disturbing and traces this reaction to a deeper or “constitutive” form of incommensurability (Durkheim, 1925/1976). People’s commitments to other people require them to deny that certain things are comparable. To attach a finite monetary value to one’s friendships, children, or loyalty to one’s country is to disqualify oneself from membership in the associated moral community. Constitutive incommensurability arises whenever treating values as commensurable subverts one of the values in the trade-off calculus. Taboo trade-offs are, in this sense, morally corrosive. The longer observers believe that a decision maker has contemplated an indecent proposal, the more negative their assessments of that person’s character, even if that person ultimately makes the “right” choice and affirms the sacred value.

Actor-based predictions. Resource constraints sometimes bring people into threateningly close psychological contact with proposals that compromise sacred values. The SVPM predicts that decision makers will feel tainted by merely contemplating scenarios that breach the psychic wall between secular and sacred and will engage in symbolic acts of moral cleansing that reaffirm their solidarity with their moral community. This SVPM prediction should not, however, be confused with the self-affirmation hypotheses that can be derived from variants of dissonance and social identity theory (Schlenker, 1982, 1985; Steele, 1988). The SVPM differs from these mostly complementary formulations in two 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 prerejection interval, during which one’s natural first reaction to propositions is to consent (Gilbert, 1991), can produce a subjective sense—however unjustified—that one has been contaminated and fallen from moral grace. 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, in its original form, focused solely on the intrapsychic function of maintaining mental equilibrium or which, in later revisionist forms, focused on protecting the self, the SVPM assigns double-barreled roles to outrage and cleansing: the expressive function of convincing oneself and others of one’s worthiness and the instrumental function of shoring up the moral order by punishing violators and by personifying compliance.

Ideational-content predictions. The SVPM identifies three categories of proscribed social cognition—taboo trade-offs, forbidden base rates, and heretical counterfactuals—each a potential content boundary condition on strictly cognitive theories of judgment and choice. To predict exactly where people draw the line between the thinkable and unthinkable, it is, however, necessary to link the “process” assumptions of the SVPM with auxiliary “content” assumptions drawn from Fiske’s (1991) cross-cultural taxonomy of relational schemata that specifies how people compartmentalize their social worlds. Drawing on Fiske and Tetlock (1997), the SVPM defines a taboo trade-off as any proposal that inappropriately extends a market-pricing schema (which presupposes ratio-scale fungibility of all values into a single utility metric) to spheres of activity regulated by three less metrically demanding relational schemata: equality matching (e.g., offering

to pay a friend for dinner at her house instead of reciprocating the invitation), authority ranking (e.g., attempting to bribe legitimate authority figures rather than deferring to their judgment), and communal sharing (e.g., treating loved ones as objects of monetary calculation rather than displaying unconditional commitment to their well-being).

The SVPM maintains that categorical proscriptions on cognition can also be triggered by blocking the implementation of relational schemata in sensitive domains. For example, forbidden base rates can be defined as any statistical generalization that devout Bayesians would not hesitate to insert into their likelihood computations but that deeply offends a moral community. In late 20th-century America, egalitarian movements struggled to purge racial discrimination and its residual effects from society (Sniderman & Tetlock, 1986). This goal was justified in communal-sharing terms (“we all belong to the same national family”) and in equality-matching terms (“let’s rectify an inequitable relationship”). Either way, individual or corporate actors who use statistical generalizations (about crime, academic achievement, etc.) to justify disadvantaging already disadvantaged populations are less likely to be lauded as savvy intuitive statisticians than they are to be condemned for their moral insensitivity.

Heretical counterfactuals can be defined as what-if assertions about historical causality (framed as subjunctive conditionals with false antecedents) that pass conventional tests of plausibility but that subvert core religious or political beliefs. Cultures that emphasize authority ranking are likely to treat counterfactuals as heretical that reduce the conduct of higher-spiritual-status beings—leaders, saints, deities—to explanatory generalizations that are routinely applied to lower-spiritual-status beings. For example, Christian fundamentalists see a direct authority-ranking relationship between God and humanity. Believers are supposed to defer to the Scriptures, the literal word of God. Counterfactuals that depict the life of Christ as subject to the same vagaries of chance as the lives of ordinary mortals mock God’s message. Heretical counterfactuals are deeply disrespectful and, in earlier times, would have merited painful punishment. In modern societies, dissenters can escape such Draconian sanctions, but they still must endure the scorn of the faithful.

Reality-constraint predictions. The SVPM portrays people (intuitive theologians) engaged in a delicate mental balancing act. On the one hand, people are posited to be sincere in their protestations that certain values are sacred. On the other hand, people do run into decision problems in which the costs of treating sacred values as infinitely important become prohibitive. The model predicts that, absent social pressure to confront the contradiction, people will be motivated to look away and will be easily distracted by rhetorical diversions or smoke screens. However, when the gaze-aversion coping response is not an option, people will search for ways of redefining the situation that transform taboo trade-offs into either routine trade-offs (one secular value against another, the sort of mental operation one performs every time one buys or sells goods and services in competitive markets) or tragic trade-offs (one sacred value against another, such as honor vs. life, the sort of comparison featured prominently in classic Greek tragedies). The latter types of trade-offs do not have the morally corrosive effects that taboo trade-offs have on both personal and social identity.

Evidence Bearing on Key Predictions of the Model

Tests of the SVPM presuppose a great deal of culture-specific knowledge of what people hold sacred. The focal applications have been to taboo trade-offs, forbidden base rates, and heretical counterfactuals.

What counts as a taboo trade-off hinges on one's political reference group. Tetlock et al. (1996) found considerable agreement among liberal Democrats and conservative Republicans on the boundaries of the fungible. Widely agreed-on taboo trade-offs included body organs, adoption rights for children, and basic rights and responsibilities of citizenship. Disagreement arose mostly on the ideological fringes: among libertarians who thought it a good idea to extend market-pricing norms into taboo territory and socialists who thought it a bad idea to permit market-pricing norms in currently permissible domains such as medical care, legal counsel, and housing.

Building on this earlier work reconnoitering moral boundaries, Tetlock, Kristel, Elson, Green, and Lerner (2000) explored the SVPM notion that taboo trade-offs (which pit nonfungible sacred values against fungible secular ones) possessed distinctive properties relative to routine trade-offs (which pit secular values against each other) and tragic trade-offs (which pit sacred values against each other). Comparisons of reactions to routine and taboo trade-offs yielded results consistent with the more intuitively obvious implications of the SVPM. Taboo but not routine trade-offs provoked participants. People censured, directed anger toward, and ostracized those who made taboo trade-offs. They also engaged in meta-norm enforcement, imposing sanctions on observers who fail to condemn taboo trade-offs.

Comparisons of taboo and tragic trade-offs yielded results consistent with the less intuitively obvious predictions of the SVPM. Tetlock et al. (2000) documented a "mere contemplation" effect for moral outrage that took opposite functional forms for taboo and tragic trade-offs. The longer observers believed that decision makers pondered a taboo trade-off (e.g., the dollar value of lives), the more damage done to their reputations, even if they ultimately made the "right" choice of lives over money. By contrast, the longer observers believed that decision makers thought about a tragic trade-off (e.g., which life to save?), the wiser and more judicious they were deemed to be. Dwelling on a "sleazy" taboo trade-off increases its moral corrosiveness. Dwelling on an "ennobling" tragic trade-off reinforces the impression that one has displayed due respect for the profound importance of each of the clashing sacred values. In this same vein, Tetlock et al. (2000) also documented a mere-contemplation effect on moral cleansing. Merely contemplating taboo (but not tragic) trade-offs motivated decision makers to reaffirm their moral character by expressing stronger intentions to perform good deeds.

Tetlock et al. (2000) documented similar outrage and cleansing responses to forbidden base rates and heretical counterfactuals. One study manipulated observers' beliefs about the correlation between a base-rate classification (the distribution of fires across neighborhoods) and a forbidden predictor such as race. Observers with egalitarian views reacted with outrage to executives who used race-correlated base rates in setting premiums but not to executives who used race-neutral base rates. In a follow-up experiment, participants role-played executives who discovered that they had inadvertently used a forbidden base rate. Thrust into this moral

predicament, racial egalitarians were more likely to morally cleanse themselves by volunteering for good causes.

Another study tested the SVPM prediction that moral communities will erect emotionally charged boundaries against counterfactual speculation that applies secular or scientific standards of evidence and proof to the founders of sacred movements. Tetlock et al. (2000) showed that fundamentalists were outraged by heretical counterfactuals that "undid" key decisions by Jesus Christ but reacted with equanimity to everyday counterfactuals derived from the same schemata. Fundamentalists also felt contaminated by merely contemplating such counterfactuals and reported renewed dedication to serving their church.

As noted earlier, the SVPM is but one of a range of testable theories of how people function as intuitive theologians. It is possible in any given cultural-historical context that the model either underestimates or overestimates how dogmatic, intolerant, or punitive people are prepared to be. For instance, advocates of more cognitively flexible portraits of the intuitive theologian suspect that sacred values may merely be pseudosacred (L. Thompson, 2001). Insofar as there is disagreement here, it is over how much emphasis to place on the reality-constraint postulate of the SVPM. How quickly do people start looking for good pretexts for abandoning their supposedly bedrock commitments to sacred values? In our laboratory, we do indeed find some evidence that people can be induced—through the give-and-take dynamics of political debate—to redefine taboo trade-offs as either routine or tragic in character. Tetlock (2000b) found that small but significant fractions of samples will qualify their opposition to the buying and selling of body organs for medical transplants if convinced that (a) such transactions are the only way to save lives that otherwise would have been lost and (b) steps have been taken to assist the poor in purchasing organs and to prevent the poor from selling their organs in "deals of desperation." The first type of argument effectively "sacralizes" the secular side of the trade-off (recasting the issue as one sacred value against another: lives vs. moral objections to commodifying organs); the second type of argument "secularizes" the sacred side of the trade-off (now recasting the issue as one in which it may be possible to use transfer payments to neutralize at least egalitarian objections to organ markets). A once clear-cut taboo trade-off thus blurs into either a tragic trade-off (in the first case) or a routine trade-off (in the second). Either way, the SVPM needs to acknowledge that the primal Polynesian meaning of *taboo*—absolute, automatic, unreasoned aversion to any breach of the psychic barriers separating profane from sacred (Radcliffe-Brown, 1952)—is occasionally misleading. Anathema though the idea is to hardliners, the permeability of these barriers (to use a Lewinian metaphor) may be a negotiable item for cognitively flexible theologians.

Advocates of more cognitively flexible portraits of intuitive theologians can also point to historical evidence of how once-taboo trade-offs or exchanges have become permissible (nouveau riche merchants buying the sacred land of financially strapped feudal lords) and of how routine trade-offs and exchanges have become taboo (between the U.S. Civil War and World War I, it ceased to be acceptable to pay others to perform military-service obligations). These theorists can even argue that if the SVPM is right that people have a deep-rooted need to believe that they are faithfully upholding sacred values, and if microeconomists are right that resource constraints regularly compel taboo trade-offs

anyway, then surely moral outrage should be far more pervasive than it is. Why aren't policy makers constantly vilified? One possibility is that policy makers out-fox the public by cloaking taboo trade-offs in artful rhetoric. Tetlock's (2000b) "obfuscation" experiments suggest that these cloaking exercises need not even be all that artful. Policy makers who traded off lives and money in cleaning up toxic waste sites were sharply censured when they honestly acknowledged what they had done but they escaped censure when they offered vacuous utilitarian or deontic justifications—such as "on balance" or "in principle" Policy X is the right thing to do—for the same decision. This result may resolve the paradox of how moral communities can sustain the illusion of an unbounded commitment to sacred values when econometric (revealed preference) studies indicate the community regularly trades those values for secular ones. Moral communities may be instructively modeled as systems of organized hypocrisy (Brunsson, 1989), whose members either cannot see or do not try to see through rhetorical smoke screens for taboo trade-offs and who sometimes even collude in perpetuating the deception.

It is also possible, however, to criticize the SVPM for underestimating the moral resolution of defenders of sacred values. Greenberg, Pyszczynski, Solomon, Simon, and Breus's (1994) terror management theory, for example, suggests conditions when intuitive theologians may become especially punitive. Reminding intuitive theologians of their mortality should motivate them to seek out the existential comfort of a shared worldview that transcends their life spans. Linking this inventive work on mortality salience to the SVPM leads to the hypothesis that, agnostic Bayesian libertarians excepted, people reminded of their mortality should become especially hard-line theologians who are motivated to defend their cultural worldview and are easily roused to wrath by challenges to the collective moral order.

Connecting the SVPM to work on self-presentation also suggests conditions under which intuitive theologians may become sterner defenders of sacred values. One hypothesis is rooted in the plausible assumption that people derive social identity benefits from ostentatious displays of virtue. Moral outrage and cleansing should be amplified when people feel monitored by their community of cobelievers—which may explain the extremity of punitive-damage awards levied by juries when corporations have been caught placing explicit dollar values on human lives (Sunstein, Kahneman, & Schkade, 1998). A related hypothesis is that, when people are prevented from distancing themselves from proscribed cognitions and this failure to condemn is observed by cobelievers, people will become self-dramatizing theologians who seek reintegration into the community by publicly cleansing themselves and denouncing norm violators.

Normative Challenges to Prevailing Taxonomies of Error and Bias

True believers dismiss the opportunity costs of shunning taboo trade-offs or the distortions in market pricing from ignoring forbidden base rates—although intuitive economists and statisticians are vexed by these deviations from rationality narrowly construed. True believers halt attributional inquiry as soon as it trespasses into the domain of the sacred—although intuitive scientists are disturbed by the abrupt curtailment of curiosity about heretical counterfactuals. True believers deplore the belief-bending ways of

pragmatic intuitive politicians. When sacred values are at stake, attitude shifting smacks of sellout; self-criticism blurs into genuine doubt on articles of faith, and decision avoidance looks like cowardice.

Intuitive theologians tend to be suspicious of the classic Enlightenment values of open-minded inquiry, free markets, and of tolerant pluralism. Some questions should never be asked, some transactions never proposed, and some compromises never struck. How can it be determined, though, whether the coping strategies of intuitive theologians serve their intended function? The new normative benchmarks are mostly subjective: Do people who aim moral outrage at norm violators feel vindicated by the opportunity to do so (and frustrated when thwarted)? Do moral cleansers follow through on their announced intentions to do good deeds that then, by a dissonance or self-perception process, further solidify their faith? In addition, however, there may be communitarian benchmarks of effectiveness. Durkheim (1925/1976) may have been right. Group expressions of outrage and cleansing may increase the cohesiveness of communities of cobelievers. Communities that censure together may stay together.

Intuitive Prosecutors

The core functionalist premise is that people seek to defend rules and regimes that they endow with legitimacy. Again, as with any functionalist metaphor, there is a conceptual range of permissible theories, ranging from the conscientious jurist (who draws conclusions about culpability only after dispassionately weighing the evidence) to the capricious, vengeful, or opportunistic prosecutor (who permits mean-spirited or self-interested motives to taint evaluations). But there are empirical constraints. Theorists will ultimately be compelled to incorporate personality variation (some people are more punitive than others), situational variation (certain contexts promote leniency or punitiveness) and personality-by-situation interactions (certain contexts move some but not others to be punitive). These concessions to complexity are required by salient features of the social landscape: (a) by the bitter partisan divisions between those who seek procedural safeguards that shield individuals from arbitrary exercises of state power and those who seek to empower society to deter misconduct—even if at the expense of individual rights (cf. Sniderman, Fletcher, Russell, & Tetlock, 1996); (b) by the deep philosophical divisions over the appropriate goals of punishment, ranging from rehabilitation to specific and general deterrence to retribution (D. T. Miller & Vidmar, 1981); (c) by the vast variation in how blatantly biased in-groups are toward out-groups—an historical continuum anchored at one end by state-sponsored genocide and at the other by systematic exercises in constitutional engineering to promote tolerance and pluralism (Horowitz, 1985).

A Testable Model of How People Function as Intuitive Prosecutors

The theory advanced here strikes a compromise among these competing positions: the fair-but-biased-and-occasionally-erratic model of the intuitive prosecutor. This "fair-biased-erratic" (FBE) model can be captured in five sets of propositions.

Commitment, albeit imperfect, to procedural fairness. People think of themselves as fair and pay homage to basic norms of

procedural justice that stipulate they should weigh evidence impartially, accord equal respect to each side, and respect fundamental human rights, even if doing so impedes apprehending violators (Tyler & Smith, 1998). The FBE does not challenge the sincerity of this commitment to procedural norms, but it does challenge the consistency with which people apply abstract norms to specific cases and it predicts considerable principle-policy slippage for reasons that follow.

Mechanisms for switching the prosecutorial mind-set “on” or “off.” The prosecutorial mind-set can be activated by experimental manipulations of the perceived fragility of the social order as well as by chronic personality variations in these perceptions. It makes sense to become warier of situational justifications and excuses for violating society’s rules to the degree one has been induced to believe, or is already predisposed to believe, that misconduct is on the rise, that growing percentages of misconduct are going unpunished, and that growing numbers of citizens are tempted, by the collapse of general deterrence, to behave badly. Once this prosecutorial mind-set is activated, it seems only prudent to place greater weight on punishing norm violators (minimizing Type 2 errors of acquitting the guilty) and less weight on protecting the innocent from the wrath of the collective (minimizing Type 1 errors of convicting the innocent). This policy preference is rarely stated so baldly (bordering as it does on a taboo trade-off). Intuitive prosecutors should, however, feel more justified in adopting an across-the-board hard-line stance to the degree they believe that Type II errors are both more likely and more serious than Type I errors—beliefs that, in turn, hinge on assumptions about human nature as well as the benevolence and competence of authority figures.

Of course, even the most judgmental judges do not relish living under an oppressive regime that renders them constantly vulnerable to false accusations. Intuitive prosecutors must strike a balance between upholding the social order and neutralizing defensive strategies that they themselves might want to deploy when they are in the metaphorical docket. The FBE therefore posits an exculpation gradient for the accounts offered by those accused of wrongdoing. This gradient is anchored at one end by diminished-capacity excuses that invoke well-established biological causes outside volitional control (e.g., epilepsy) that virtually no one challenges. At the other end, it is anchored by frivolous justifications or excuses that virtually no one accepts and that may well backfire (e.g., “I was too drunk to see the pedestrian”). Toward the middle of the gradient are such defenses as addiction, past victimization, and peer pressure that possess some plausibility but that intuitive prosecutors fear can readily degenerate into abuse excuses. Effects of prosecutorial-mind-set manipulations should be most pronounced in this zone of intermediate exculpatory power. This zone is not, however, fixed forever. The relative openness of intuitive prosecutors to situational defenses is a quasi-stationary equilibrium (to use a Lewinian formalism) that shifts with the relative force of the conflicting pressures to avoid Type I and II errors.

The FBE also posits that, even in the absence of pressing threats, the prosecutorial mind-set should be more chronically accessible for those who are ideologically predisposed to believe that the normative order is at risk and that it is therefore prudent to set low thresholds for attributing intentionality to norm violators. This prediction does need, however, to be qualified by ideology-by-context interactions. Intuitive prosecutors do not dispense blind

justice. The predisposition to blame should be moderated by the affect people harbor toward the act (e.g., certain acts—flag burning vs. cross burning—are more vexing to certain groups), toward the actor (e.g., is there evidence of repentance or extenuating circumstances?) and the group affiliation of the actor (e.g., does the group uphold or subvert existing social structures?). Conservatives should be maximally punitive when the consequences are severe and when the perpetrator cannot invoke extenuating circumstances, is unrepentant, and is affiliated with groups flagrantly contemptuous of the social order. Egalitarians should be maximally punitive under similar circumstances, except that the act should target a traditionally disadvantaged group and the perpetrator should hold socioeconomic underdogs in disdain (cf. Skitka & Tetlock, 1993).

There is a palpable tension between the “fair” and “biased” components of the FBE. The FBE resolves this tension by positing that intuitive prosecutors will play favorites only when they can generate convincing justifications that they have not done so. In the language of impression management, biased prosecutors need the cover of attributional ambiguity to deflect accusations of bias. Naked discrimination is unacceptable in most cultures wealthy enough to fund psychological research programs.

Intuitive prosecutors: Forward- and backward-looking mind-sets. Intuitive prosecutors can be either forward looking or backward looking when they translate their attributions of responsibility and emotional reactions into recommendations of punishment. The forward-looking mind-set should be activated in settings that legitimize the efficiency logic of *Homo economicus*: Given pressing resource constraints, the group’s goal is to identify cost-effective forms of punishment that promote specific and general deterrence (minimize the likelihood that the offender or others will act in the proscribed manner). The backward-looking mind-set should be activated in settings that legitimize the retributive quest to restore the ante of the moral status quo, settings such as cultures of honor in which agents of social control had better move quickly to reaffirm the dignity of victims lest the victims take justice into their own hands (Nisbett & Cohen, 1996).

Although retributive justice is often portrayed as a mindlessly vengeful and limbic reaction to transgressions, the FBE traces the desire for retribution to a more reasoned calculus in which people implicitly think along the following lines: The transgressors treated the victims contemptuously; the damage to the victims’ dignity is real; the transgressors see themselves as above the victims and the laws of society; the punishment should humiliate the transgressors to a degree commensurate with the humiliation inflicted on the victims; anything less fails to reaffirm the dignity of the victims and indeed adds to the original insult. The FBE predicts that retributive prosecutors will impose punishment beyond that required by deterrence whenever the deterrence prescription for punishment falls short of reaffirming the identity damage to victims. The acrimonious debate over hate crimes illustrates the point. Advocates of traditionally victimized groups often insist that, even if the conventional penalties for assault and battery are adequate for deterrence, additional penalties for hate-motivated crimes affirm the value that society now places on protecting groups wronged in the past (cf. Jacobs & Potter, 1998). There is, however, an ironic twist. Advocates of laws against hate crimes are mostly on the left, which has traditionally denied the legitimacy of retribution, whereas opponents of such laws are mostly on the

right, which has traditionally upheld retribution as a legitimate goal.

Irrelevant affective states. Intuitive prosecutors can be fickle and inconsistent, allowing irrelevant moods and emotions to taint attributions of responsibility. These irrelevant affective states fall into two categories: spillover from classic Durkheimian outrage triggered by violation of society's standards and spillover from disruptions to the classic Freudian equilibrium between unacceptable impulses and the internalized voice of conscience (disruptions triggered by violations of norms that one is tempted to violate oneself).

Flexible self-correction. When motivated and able, intuitive prosecutors do engage in flexible self-correction of their judgment processes, checking irrelevant influences and difficult-to-justify double standards (cf. Wegener & Petty, 1995). Flexible self-correction should work best when intuitive prosecutors are themselves under tight accountability scrutiny, possess a good intuitive sense of the direction and magnitude of the potential bias, and are not under cognitive load.

Evidence Bearing on Key Predictions of the Model

Tetlock et al. (2002) reported a series of experimental tests of FBE predictions. These studies manipulated independent variables designed to activate the prosecutorial mind-set, especially beliefs about the direction and rate of change of certain offenses and the proportion of offenses going unpunished. By the logic of the FBE, these manipulations should promote punitive dispositional attributions that advance the goals of specific and general deterrence and of retribution. Results indicated that people in the prosecutorial mind-set were (a) less moved to leniency by extenuating circumstances but more moved to punitiveness by exacerbating circumstances (evidence that intuitive prosecutors were not oblivious to context, just selective in which cues they used); (b) more punitive as the severity of the consequences of an act mounted, holding intentionality and foreseeability constant (extending past work on the severity effect—Burger, 1981); and (c) less interested in overturning verdicts when there had been a violation of procedural justice that could be dismissed as a “technicality.” It was difficult, however, to activate the prosecutorial mind-set when people sympathized with the violator or were reminded of the dangers of false conviction.

Tetlock et al. (2002) also found support for personality-by-context predictions of the FBE. In studies of managers, authoritarian conservatives favored a managerial style that gave little leeway for justifications or excuses for failing to achieve organizational objectives, whereas liberals preferred an empathic style that gave floundering employees the benefit of the doubt. A certain insensitivity to mitigating situational factors is, from a conservative standpoint, a useful message to send to subordinates tempted to test the tolerance zone of acceptable performance by offering exculpatory accounts. Conservatives were not always, however, the most vigilant defenders of the normative order. Conservatives were more forgiving of an overzealous police officer accused of beating a suspected drug pusher. And liberals were more punitive toward racist vandals and toward corporations that befouled the natural environment. It is also worth stressing that the prosecutorial mind-set was not inherently simplistic. Intuitive prosecutors engaged in relatively complex attributional analysis that required

greater sensitivity to contextual cues when those charged with enforcing norms were overzealous in their duties. People in the prosecutorial mind-set were more likely to let off a police officer who beat a suspected drug dealer when there was an extenuating circumstance (emotional stress). Moreover, intuitive prosecutors were not indiscriminately soft on the police. The police got no benefit from mitigating circumstances when they stood accused of corruption.

Affect played a negligible role in mediating the effects obtained in the low-emotional-involvement vignette studies of Tetlock et al. (2002). Investigators using more vivid stimulus materials have, however, found evidence for more emotion-driven portraits of the intuitive prosecutor. Goldberg, Lerner, and Tetlock (1999) examined anger carryover effects by persuading subjects that they were participating in two unrelated experiments. In the first study, people watched a videotape of a brutal assault on a young man and subsequently learned that the aggressor was or was not punished. Later, participants judged culpability in tort cases in which there was presumptive evidence of negligence. Although subjects in the “punished” and “unpunished” conditions were equally angered by the assault, only participants who believed that the perpetrator had gone unpunished translated their anger over the original offense into punitiveness toward perpetrators of torts. It was as though people needed a “Durkheimian warrant” for translating anger over unpunished breaches of norms into punitiveness in later unrelated cases. The warrant effectively tagged the anger as righteous, and subjects in that state drastically simplified their attributional analysis, moving directly from acts to dispositions, with only a perfunctory pause for extenuating circumstances.

J. S. Lerner et al. (1998) deployed the same anger-priming tape to assess the impact of case-irrelevant emotions, authoritarianism, and accountability on punitiveness in fictional tort cases. As predicted, anger simplified processing of tort information (reducing the complexity of cues utilization) and encouraged punitive attributions to defendants (“negligence” began to shade into “malevolence”). But when intuitive prosecutors were held accountable for their tort judgments, the effects of irrelevant anger on complexity of thought and punitiveness disappeared.

Normative Challenges to Conventional Taxonomies of Error and Bias

The distinction between the politician and prosecutor metaphors can be as murky as the distinction between politics and the law (witness the endless disputes over “politicized” judges). In the prosecutorial arena, as in the political arena, what constitutes an error or bias is often bitterly contested (Sniderman et al., 1996). In studies of managerial and political elites, Tetlock (2000a) has documented ideological fissures in judgments of the rationality of individual, small group, and institutional decision processes. Authoritarian conservatives believe that most people look for and exploit loopholes in social control systems, whereas egalitarian liberals believe that most people will refrain from exploiting loopholes as long as they feel fairly treated. These discrepant views of human trustworthiness lead to diverging assessments of the fundamental attribution error. Authoritarian managers think it prudent to communicate a low tolerance for justifications and excuses for conduct that falls short of organizational expectations and that people will try harder if they believe that failure will

automatically tarnish their reputations—a social variant of the legal doctrine of strict liability. Failing to hold people responsible for outcomes that they could have controlled is every bit as serious an error as holding people responsible for outcomes outside their control. By contrast, many liberals view the fundamental attribution error as punitive, not prudent. They argue that conservatives exaggerate how frequently people will invent specious justifications and excuses for substandard performance.

Prosecutorial priorities also collide with epistemic or economic ones in other arenas. These arenas include the following.

Severity effects. A drunk driver speeds through a red light and, in one scenario, there is a minor fender bender and, in the other, a child is killed. From a prosecutorial perspective, people should deem it justifiable to mete out harsher punishment as a function of outcome severity, holding volition, foreseeability, and intentionality constant (Tetlock 2000a). The severity effect is a sound method of sending the message to miscreants that the risks they recklessly inflict on others can randomly rebound back on them. But from an intuitive-scientist perspective, the same effect looks like an irrational intrusion of the certainty-of-hindsight bias into the judgment process. Once observers learn of the outcome, they erroneously infer that they knew what was coming all along and that the actor therefore should also have foreseen the outcome.

The ethical impasse cannot, of course, be broken empirically, but the degree to which ordinary people resonate to the functionalist defense of the severity effect by the intuitive-prosecutor camp and the functionalist critique from the intuitive-scientist camp can be gauged. Tetlock et al. (2002) assessed the degree to which liberal and conservative managers found it embarrassing to acknowledge that they impose different penalties on careless employees as a function of the severity of the consequences. Using a repeated-measures design that made it transparent which cues were being manipulated, they found that (a) liberals were more lenient than conservatives and (b) both liberals and conservatives displayed a severity effect and were harsher when they first learned that the consequence of carelessness was serious than when they first learned the consequence was trivial. However, in response to later, what-if questions, conservatives saw nothing wrong with linking penalties to severity of consequences. The severity effect was good policy. By contrast, liberals were wary of judging people on the basis of outcomes that were so under the sway of chance. Liberals were thus more reluctant to modify their initial attributions of responsibility in response to a cue that they thought should be given little weight. As a result, liberals who first received the severe-consequence and then the mild-consequence scenario were actually more punitive in reactions to the second scenario than conservatives were.

Proceduralist and retributivist challenges to Homo economicus. Economic models of legal reasoning assume that intuitive prosecutors are continually seeking the most cost-effective societal strategies to deter wrongdoing. This doctrine of “optimal deterrence” leads to a host of predictions that clash with more psychological conceptions of intuitive prosecutors. For example, Polinsky and Shavell’s (1999) analysis of punitive damages asserted that the principal function of such awards is to make up for the shortfall in compensatory damages produced by the failure of many potential plaintiffs either to detect their injuries or to secure compensation. This leads to the counterintuitive prediction that *Homo economicus* in the legal arena should multiply compensatory awards by the

inverse of the probability that any given person from the population of injured persons would receive compensation for the injury. The fewer perpetrators caught, the more harshly those who are caught should be punished. Sunstein et al. (1998) documented that people are understandably reluctant to do this (by this reasoning, why not save money by laying off a large percentage of the highway patrol and beheading the hapless few speeders who are apprehended?). From a narrowly economic standpoint, this recalcitrance should count as a judgmental defect. The FBE differs from the Polinsky and Shavell analysis in two key respects. First, it posits that people care about equitable treatment and this concern should animate opposition to concentrating punishment on the hapless few to compensate for the many scofflaws. Hence the FBE predicts less punitiveness than does the microeconomic model. Second, the FBE posits that people should punish more severely than efficiency considerations would dictate when the cost-effective punishment fails to fulfill the retributive function of humbling arrogant violators to compensate for the humiliations that they inflicted on the victims.

Available evidence supports these hypotheses. Baron and Ritov (1993) found that penalties imposed on corporations for producing defective birth control pills and vaccines were unswayed by whether the penalties had beneficial consequences (encourage the manufacture of safer products) or harmful ones (cause companies to withdraw from the market, making products less available, more expensive, and less safe). Wrongdoers must be punished—even if it is painful to society as a whole. Baron and Ritov also found that the penalties were unaffected by removing all utilitarian rationales for punishment: the penalty was secret, the company had insurance, and, in any case, the company was going out of business. In punishing companies for dumping hazardous wastes, people wanted companies to clean up their own waste, even if the waste did not threaten anyone and even if it were possible to redirect the same money to clean up the waste of a now-defunct company that posed a far greater hazard (Baron, Gowda, & Kunreuther, 1993). From a microeconomic view, these choices are perversely Pareto suboptimal. But these choices are easily reconciled with a portrait of the intuitive prosecutor as guided by the “primitive” Durkheimian intuitions that those who harm others, including nature, must restore the status quo ante.

Collective punishment. Psychologists often dismiss collective punishment as crude guilt by association (Heider, 1958). The policy can, however, be defended as a rational prosecutorial strategy to stimulate mutual accountability among group members and to forge a communitarian consciousness in which people rarely feel justified in replying “none of your business” to demands for accounts from family, friends, and coworkers (Edgerton, 1985). People who invoke the none-of-your-business defense will receive the reply. “Oh, yes, it is—for we too must answer for what you do.” This functionalist analysis bears on recent claims that the fundamental attribution error is peculiar to individualistic Western societies in which folk psychologies stress individual agency and vanishes in the Confucianist societies of east Asia in which folk psychologies stress dialectical interdependence and in which moral-legal codes have traditionally endorsed collective punishment (Peng & Nisbett, 1999). Isolating individual perpetrators as the locus of causal and moral responsibility may be a distinctively Western method of enforcing order. However, holding significant others of the perpetrator accountable as well, and yoking their

reputations to the perpetrator, appears to be equally effective, even if the locus of responsibility transcends the internal–external dichotomy.

What Then Is to Be Done?

The approach taken here is open to critique from at least three conceptual camps: relativism, reductionism, and falsificationism. This section considers these objections and sketches a metatheoretical stance toward functionalism that strikes a principled compromise among the complex scientific values at stake. This compromise position tries to avoid the excesses of relativism (collapse of all standards of rationality), the excesses of reductionism (the distortions that result from squeezing social functionalist explanations into exclusively intrapsychic templates), and the excesses of falsificationism (the arbitrariness of drawing sharp distinctions between overlapping fuzzy sets of rival functionalist theories).

Relativism

This camp argues that there is no point of diminishing returns in diversifying the portfolio of functionalist premises. The more conceptual vantage points for viewing reality, and the more faithful those vantage points are to the phenomenology of participants, the better. It is, moreover, insufferably elitist to suppose that the research community knows any better than the people they study the diverse ends that judgment and choice serve in cultural or historical context (Geertz, 2000). Imposing a priori categories of goals and functions blinds researchers to the transitory and idiosyncratic webs of meaning that people are constantly spinning around their decisions and interactions with each other.

This stance is tantamount to abandoning the nomothetic goals of science—explanation and prediction—in favor of the ideographic and hermeneutic goals of history and ethnography. It also amounts to postmodernist renunciation of rationality itself. Unconstrained shifting of functionalist premises makes it possible to concoct post hoc rationalizations for virtually anything people do, no matter how incoherent, a risk that haunts even the more restrained form of functionalist pluralism adopted here. When people appear to be defective statisticians, the functionalist can propose that they are attentive conversationalists on the lookout for predictive cues in even the most vacuous utterances (Schwarz, 1994). When people appear to be flawed psychologists, the functionalist can caution that people are shrewd prosecutors who know how to keep others in line (Hamilton, 1980). When people appear to be myopic economists, the functionalist can cast them in the role of devout theologians protecting sacred values from taboo trade-offs (Tetlock et al., 2000). The conceptual guidelines proposed here for constructing social functionalist theories do, however, place clear limits on functionalist forgiveness. The epistemological stopping rule observed in this article has been to consider only those functionalist reinterpretations of putative biases that lead to testable hypotheses and to corroborative evidence that would otherwise have gone undiscovered.

Reductionism

This camp seeks to assimilate social functionalist theories into a common conceptual language and explanatory template. The ideal

type of successful reductionism in science is thermodynamics to statistical mechanics. In Nagel's (1961) lucid exposition, Theory X is reduced to, and becomes identical to, Theory Y when the bridge laws that translate X into Y can be specified, as when the temperature of a gas became synonymous with its mean molecular kinetic energy. In an analogous vein, reductionists would insist that social functionalism is a placeholder for ignorance of underlying cognitive and affective mechanisms and that as soon as these mechanisms are understood, functionalist explanations can be replaced just as surely as statistical mechanics replaced classical thermodynamics.

There are, at present, several potentially promising conceptual strategies for pursuing the reductionist objective. In the spirit of signal detection theory, one could argue that manipulations of social functionalist mind-sets have no effect on how people think but that they do influence where people set their response thresholds (betas) for expressing various categories of judgments (Friedrich, 1993). Unaccountable and accountable decision makers think in the same basic ways, but accountable decision makers are more cautious about what they say. Or, to offer another example, people in punitive mind-sets are not any less capable of differentiating the innocent from the guilty, but they do place an unusually high value on not letting the guilty go free.

In the spirit of contemporary work on construct accessibility and spreading activation in associative networks, one could argue that there is no need to invoke mysterious functionalist distinctions to explain the effects of social functionalist manipulations when simple associationist principles of the mind can do the same job. Accountability, for example, may alter the decision calculus because it reminds people of the viewpoint of the person with whom they expect to interact, or it reminds people of past episodes in which they have been under pressure to justify their views, and the coping strategies that worked come to mind. Reinvoking N. E. Miller and Dollard (1941), social functionalist frameworks shed light on what people think about; they tell us nothing about how people think.

In the spirit of classic drive theory, one could argue that social functionalist manipulations have drive- or arousal-inducing properties, and the resulting effects on cognitive performance fit into familiar arousal-by-task-difficulty relationships that date back to Hull-Spence days. The effects of accountability on judgment and choice may be mediated by the effects of social facilitation arousal on the cognitive processes required for task performance, or the "rigidity" displayed by intuitive theologians who perceive sacred values to be under siege may be consistent with the more general literature on the effects of stress on information processing.

In the spirit of dual-process models of social cognition, one could treat social functionalist manipulations as equivalent to other incentives (such as money, consequences, involvement, or interdependence) to engage in more effort-demanding and elaborate information processing and thereby bring "functionalist" findings into the explanatory orbit of dual-process models of social cognition (Chaiken & Trope, 1999).

There are, however, good logical and empirical grounds for doubting the wisdom of full-scale reduction. On the empirical side, it can be argued that the proposed reductionist equivalence either does not hold or holds too loosely to have explanatory value. Here the argument pivots on details of evidence. Some functionalist manipulations—accountability, social threat, sacred status of values—do undoubtedly influence willingness to express certain

thoughts, but many effects cannot be so explained. In the accountability literature, there are five methodologically distinct sources of evidence that the cognitive processes, activated by self-criticism-inducing forms of accountability, are both more computationally demanding and effort taxing than a simple response-threshold-adjustment process of moving mindlessly toward the middle would be. The evidence includes the following.

1. Self-criticism-inducing accountability manipulations increase the complexity of thought as revealed by content analysis of confidential thought protocols and statistical modeling of cue utilization (Ashton, 1992; Hagofores & Brehmer, 1983; Tetlock et al., 1989).

2. Self-criticism-inducing manipulations are more effective in attenuating biases such as primacy, overattribution, and overconfidence only when participants learn of being accountable prior to viewing the evidence on which they base judgments (Tetlock, 1983b, 1985; Tetlock & Kim, 1987; E. P. Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994). Timing should not matter if people were simply shifting response thresholds; it should matter if accountability effects were mediated by the initial interpretation of evidence.

3. The power of such manipulations to attenuate bias is itself attenuated by impositions of cognitive load that disrupt effort-demanding forms of information processing (Kruglanski & Freund, 1983)—disruption that should not occur if people coped with accountability by relying exclusively on low-effort attitude shifting.

4. When such manipulations improve judgmental accuracy, the effects are often too differentiated to reproduce by response-threshold adjustment models. These effects include enhanced “differential accuracy” from a Cronbach decomposition of person-perception accuracy (Tetlock & Kim, 1987; Mero & Motowidlo, 1995), improved correspondence between subjective confidence in predictions and objective accuracy—without degradation in ability to assign differential confidence to accurate and inaccurate predictions (Siegel-Jacobs & Yates, 1996; Tetlock & Kim, 1987), and closer correspondence between subjective importance of cues in prediction tasks and “objective” weightings derived from multiple regression (Hagofores & Brehmer, 1983).

5. The effects of such manipulations on thoughts persist even after cancellation of the anticipated interview (Pennington & Schlenker, 1999).

Some social functionalist manipulations have drive- or arousal-inducing properties, but it is also awkward to explain existing accountability effects in this fashion. Consider the difficulties that arise in assimilating the effects of accountability on judgment and choice into a social-facilitation framework that joins assumptions about the drive-inducing properties of the mere presence of an audience with assumptions about the effects of drive arousal on performance of inferential tasks of varying difficulty (Pelham & Neter, 1995; Zajonc, 1965). From the standpoint of the SCM, this reduction is fated to fail because a unidimensional arousal metric cannot fit, even post hoc, the complex variations in accountability effects that depend on qualitative features of the relationship between decision maker and audience. Too much hinges on when people learn of being accountable, whether they know the views of the audience, and how exactly they perceive the accountability requests or demands (J. S. Lerner & Tetlock, 1999). It is also dauntingly difficult to reduce the diverse judgment-and-choice-

dependent variables to a single task-difficulty metric that permits deduction of Hull-Spence Task-Difficulty \times Arousal interactions. To fit the data, one must make so many auxiliary assumptions that the original advantage of parsimony is lost.

Some social functionalist manipulations induce mood or emotional states, and it is tempting to posit that the effects on judgment are mediated by the effects of emotion on the accessibility of cognitions and on ability or willingness to engage in sober second thought (theories that usually work from the default assumptions that people are trying to understand the world, like intuitive scientists, but are subject to certain emotional distortions; Forgas, 1995). This challenge is hard to rule out. It is difficult to conceive of any experimental induction of the prosecutorial mind-set or any other social functionalist mind-set that does not alter mood or affect and prime supportive cognitions. Even here, though, reduction is easier to announce in principle than to implement in practice. Defenders of the FBE model can argue that anger is not necessary to produce punitiveness in some studies, and it is not sufficient to produce it in others (cf. Goldberg et al., 1999; Tetlock et al., 2002).

Empirical objections to reduction can, of course, always be surmounted as new evidence and better theories become available, but there are also deeper logical objections to reduction. The most compelling of these objections is captured reasonably well by Searle’s (1998) famous Chinese-room thought experiment, which he advanced to demonstrate the logical impossibility of strong artificial intelligence and of the associated notion that appropriately programming a machine is sufficient to endow it with intentional states. The thought experiment asks readers to suppose that a monolingual English speaker is isolated in a cubicle and producing Chinese answers to Chinese questions well enough to mimic a Chinese speaker, but doing so by following an algorithm written in English. Although the occupant of the Chinese room can readily pass the Turing test, most observers balk at saying either that this person, or that a computer executing the algorithm, understands Chinese. Searle argued that, far from irrational, this resistance would be justified for any combinatorial algorithm because algorithms are “syntactically individuated” in the sense that, like generative grammar, they represent unique, internally consistent rule systems for manipulating symbols that need or have no external referents. By contrast, intentional states are “semantically individuated” in the sense that they make epistemic and other functional references to the external world from which the occupant of the Chinese room has, by definition, been isolated. Searle argued for a nonreductive biological naturalism in which intentionality, like qualia such as the liquidity of water or the beauty of the sunset, are treated as high-level features that are caused by and realized in the brain but are explicable only with reference to the interactions between the brain and the external world.

Unapologetic reductionists (such as Pinker, 1997) have countered this argument by invoking history-of-science precedents in which reductions of surface phenomena to deeper causal principles that initially struck contemporaries as counterintuitive are now universally accepted. Favorite examples include heat as nothing but the kinetic energy of molecular motion or the color red as nothing but photon emissions of 600 nanometers. Searle’s thought experiment merely shows that the underlying mechanisms of language comprehension are similarly jarringly counterintuitive. The counter to this counterargument is to marshal alternative history-

of-science precedents. Reducing language comprehension or other social functions of thought to strictly internal brain-or-mind operations, to intrapsychic combinatorics, would be every bit as misleading as reducing photosynthesis to strictly physical-chemical processes and ignoring the embeddedness of these microprocesses in macronetworks of causality represented in ecological and evolutionary systems. Causal reduction, x is a function of y , is not the same as ontological reduction, x is nothing but y , in case of systemically embedded phenomena such as consciousness or photosynthesis. There is nothing inconsistent about conceding that the psychological implementation of social functions must be reducible to strictly internal brain-or-mind operations but still insisting that social functions, connecting individual minds to social systems, are ontologically irreducible. In this view, theorists who advocate the ontological reduction of social functions (e.g., such functions are nothing but intrapsychic combinatorics) have over-extended history-of-science precedents in which causal and ontological reduction were one and the same.

This argument still leaves one big reductionist contender standing: a higher order form of functionalism. Evolutionary psychologists are not at all averse to functionalist speculation or to the notion that people might think in qualitatively different ways in different contexts (Cosmides & Tooby, 1994). Qualitative variation of this sort reflects basic properties of the human mind that are the product of millions of years of natural and sexual selection pressures. The evolutionary program requires reverse engineering thought experiments that pose the question, "What types of adaptive problems was this mental software suited to solve in our ancestral evolutionary environments?" Many implementation rules specified for functionalist theories of judgment and choice satisfy the solvability and evolvability criteria of Cosmides and Tooby (1994). People, by virtue of natural and sexual selection, should come well prepared to solve such central life tasks as competing for scarce resources (the economist), making predictively useful inferences about physical and social causation (the scientist/statistician/psychologist), getting along with key constituencies (the politician), punishing norm violators (the prosecutor), and even affirming their existential significance by protecting sacred values (the theologian). The core "if . . . , then" production rules for implementing these functions are hardly so dauntingly difficult that they could not have gradually emerged over thousands of generations in a species as intensely interactive as our own—rules such as "avoid quarreling with fellow hominids who control valuable resources," "think twice before making irreversible commitments," and "spread damaging gossip about rivals." The more elaborate coping strategies of intuitive politicians, prosecutors, and theologians may be cultural overlays on modular templates.

The ultimate relevance of this framework is not in doubt: Virtually all psychologists subscribe to some variant of evolutionary theory. The position of functionalist pluralism taken here is also strikingly compatible with modularity hypotheses advanced by theorists who challenge the once conventional wisdom that people possess general-purpose reasoning skills and who insist that the mind is compartmentalized into separate systems for solving specific adaptive challenges such as linguistic communication, cheater detection, and social perspective taking (cf. Haselton & Buss, 2000; Hirschfeld & Gelman, 1994). That said, the reductionist move still appears premature. Post hoc data fitting rightly earns a theory fewer points—by most standards of epistemological

accounting—than generating novel hypotheses of the sort that can still be derived from social functionalist theories that focus on contemporary adaptive challenges rather than on those of the Pleistocene.

Strict Falsificationists

Falsificationists who subscribe to a Popperian philosophy of science demand that new functionalist frameworks justify their existence by making predictions that clearly diverge from those of traditional frameworks (Suppe, 1974). Functionalist research programs are, however, better viewed as exhaustible than as falsifiable. Even if one testable theory grounded in a functionalist framework is falsified, it remains possible to amend the theory or to advance alternative theories (Laudan, 1997). Moreover, the track record for drawing sharp distinctions between functionalist research programs has not been too inspiring in social psychology. Tetlock and Levi (1982) documented how an eclectic array of motivational theories of attribution bias (which stressed protecting self-image and social image) was once regularly pitted against cold cognitive explanations organized around the scientist metaphor. Tetlock and Manstead (1985) documented how conceptual offshoots from the intrapsychic cognitive-consistency tradition—theories such as cognitive dissonance, reactance, and equity—were once pitted against impression-management theories organized around the dramaturgical metaphor. Although these empirical confrontations did stimulate methodological innovations and noteworthy discoveries that shed light on boundary conditions for the applicability of rival theories, after each round of hypothetico-deductive confrontations, each side could still claim vindication by fine-tuning auxiliary assumptions. For instance, some dissonance theorists maintained that evaluative audiences heighten self-awareness and thus magnify dissonance ("stealing" the impression management prediction that accountability to an audience magnifies social identity-protection motives) and some impression-management theorists argued that people try to please internalized audiences ("stealing" the intrapsychic prediction that defensiveness should disappear when people were unaccountable). Strict scorekeepers are disinclined to permit such borrowing, sometimes disparaging the practice as a form of cheating that violates sacrosanct conceptual boundaries. Tetlock and Manstead (1985) replied, however, that investigators will just have to increase their tolerance of ambiguity: There are no sharp boundaries in the fuzzy-set world of functionalist frameworks, and what looks like a transparently defensive maneuver from one point of view may later generate a string of discoveries that will be credited to the vibrant problem-solving power of the research program.

Boundary blurring of this sort is already apparent in reductionist challenges to social functionalism that treat accountability effects as merely response threshold adjustments or by-products of drive or arousal states. It has already been shown that considerable conceptual stretching was required to assimilate social functionalist frameworks into mainstream ones. The analysis also suggests that observers will often disagree on when stretching has been taken to the breaking point—disagree for the simple reason that, for any given guiding functionalist metaphor, there is a substantial range of permissible theories specifying how well or poorly people cope with the designated adaptive challenges. The resulting fuzziness of category boundaries guarantees that, in the domain of

judgment and choice, most stretching exercises will neither clearly succeed nor fail but rather fall in the gray zone of indefinite intellectual stalemate. Moreover, the ultimate arbitrariness of these boundaries cannot be concealed by quasi-formal algorithms for calculating whether a research program has become “degenerative” by, say, dividing the frequency with which the “negative heuristic” of the program is invoked to deal with anomalies by the frequency with which the positive heuristic makes boldly precise predictions that turn out to be accurate (cf. Meehl, 1990).

A Compromise Position: Functionalist Pluralism

The argument thus far can be briefly summarized: (a) Waiting for the grand reduction of research programs to a common conceptual language will turn out to be a waiting-for-Godot exercise; (b) the undisciplined proliferation of research programs is an invitation to intellectual anarchy (a conclusion that thoughtful relativists never denied; Feyerabend, 1993); and (c) the quest for crucial experiments is tempting given the professional incentives confronting young scholars but is unlikely to yield anything much more conclusive than has already been accomplished by traversing this path. The metatheoretical position taken here, functionalist pluralism, can be viewed as a form of compromise among these options. It concedes to relativists the need to acknowledge multiple possible benchmarks of rationality (although it insists on constraints on the number of permissible standards), it concedes to reductionists the usefulness of documenting the precise psychological and eventually neurological mechanisms underlying the activation and implementation of functionalist mind-sets (although it insists on the inevitable incompleteness of accounts that do not recognize the external adaptive challenges that functionalist mind-sets are directed to meet), and it concedes to falsificationists the usefulness of documenting boundary conditions on the applicability of rival functionalist accounts (although it insists on the inevitability of boundary blurring given the variety of theoretical instantiations that can be advanced for each guiding functionalist metaphor).

From the standpoint of functionalist pluralism, mixed metaphors may be ugly but they are the next step in theory building. The looming challenge will be to develop viable mixed-metaphor models of how people resolve conflicts among functionalist imperatives. A useful first step is to specify more precisely the properties of the self-regulatory “machinery” to be integrated.

One promising approach, drawing on work on category accessibility (Higgins, 1996) and person-situation interactionism (Mischel & Shoda, 1995), is to treat functionalist orientations as emotionally charged sets of production rules that specify the contextual preconditions that must be satisfied to activate particular coping strategies and goal states. The relative importance of functionalist orientations should, in this view, hinge on the momentary salience of situational cues for priming mind-sets as well as on the chronic accessibility of these mind-sets. Cross-functional conflicts should be especially intense when two functions are roughly equally accessible and activate processing rules that lead to contradictory conclusions.

Even at this early stage of articulating social functionalist theories, it is possible to identify numerous contexts in which cross-functional conflicts seem to arise. Intuitive statisticians seek predictively useful cues but stop when doing so undercuts their

self-image as racially egalitarian. Intuitive prosecutors consider extenuating circumstances but stop when the social order seems to be under siege. People defend the normative order but stop when the rules start to feel intolerably oppressive. Intuitive economists seek to maximize expected utility but stop when utility maximization requires monetizing sacred values that they are supposed to treat as infinitely important. People want the approval of key constituencies, but some draw the line at compromising basic convictions. The general theme seems to be cybernetic: People discover they have had enough of a functionalist mind-set only after they have had more than enough.

Kunda (1999) offered one framework for modeling how people deal with such conflicts. Like other theorists (e.g., Kruglanski, 1990), she has focused on the tension between nondirectional accuracy goals (the quest for the truth that directs the thought processes of ideal-type intuitive scientist) and directional goals (the desire to reach a foreordained conclusion). Directional goals can bias encoding, retrieval, and subsequent interpretation, but there are plausibility constraints on the ability of noninstitutionalized adults to draw whatever conclusions they want. Even when they are swept along by a motivated current of thought, people keep in mind the countervailing need to reach conclusions that they could justify to what Kunda called a “dispassionate observer” (the homunculus who personifies the accuracy motive). Kunda’s proposal assigned a mental-watchdog status to the intuitive scientist: prevent people under the sway of directional goals from slipping into self-serving delusional worlds.

An alternative framework, however, turns this argument on its head. Personifications of social directional goals—the theologian, prosecutor, and politician—might just as easily serve as watchdogs on the mental operations of the intuitive scientist and economist, checking lines of thought that undercut social values. For example, one could even go so far as to assign metacognitive primacy to the intuitive politician. People do not care about accuracy per se; they care only about justifiability, and justifiability is a profoundly relational construct that hinges on the identity of the audience and its evaluative standards. Kunda’s dispassionate observer is a rare special case of justifiability that arises when the evaluative audience cares solely about procedural fairness or rationality.

Of course, the choice is not either-or. One could argue that both monitoring processes need to operate concurrently for people to adjust reasonably to the surrounding social world. Pressures to balance conflicting functionalist imperatives could be formally captured by parallel-constraint-satisfaction models in which (a) each functionalist orientation is a network structure with informational elements (e.g., beliefs, values) and relations of support or conflict between elements that can have varying intensity and (b) the psycho-logic of constraint satisfaction allows each element to simultaneously influence and be influenced by all other elements to which it is directly or indirectly connected. These simulations often converge on stable asymptotic balances that maximize the consistency of the entire network (cf. Read & Marcus-Newhall, 1993).

The battle for explanatory primacy will obviously not suddenly end with the adoption of an integrative agenda. Multifunctional frameworks must still make potentially tendentious priority assignments to functions and their modes of implementation. Functionalist pluralists insist, however, that these disputes are best viewed as matters of degree, not of kind. The relative importance of

mind-sets varies across individuals and contexts, and research effort is better devoted to (a) documenting these moderators and (b) shedding light on the self-regulatory dynamics of shifting from one mind-set to another and of striking compromises between mind-sets. This analysis fits with yet another, albeit higher order, political metaphor. Functionalist pluralists model the mind as a polyglot polity populated by semiautonomous functional fiefdoms, each with its own operating principles, in uneasy coexistence. They tend to be suspicious of monistic proposals, such as Greenwald's (1980) totalitarian ego, viewing them as special cases in which a particular set of functions—those stressing cognitive order, continuity, and self-image protection—have achieved self-regulatory primacy. For most people, most of the time, the human psyche is a patchwork quilt of ever-shifting functionalist compromises.

Skeptics might counter that mixed-metaphor theories take on the Promethean aspirations of classic theories of personality. Not enough is known about the psychological implementation of specific functions to merge these diverse facets of human nature. However, the counterclaim is that these functionalist facets blur into each other. Unifunctionalist tunnel vision blinds the research community to empirical and normative boundary conditions on basic effects. Inconvenient though it is, people are multifunctional entities that demand cumbersomely complex explanations.

References

- Aberle, D. F., Cohen, A. K., Davis, A. K., Levy, M. J., Jr., & Sutton, F. X. (1951). The functional prerequisites of a society. *Ethics*, 3, 60–111.
- Allport, G. W. (1985). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1, pp. 3–56). Cambridge, MA: Addison-Wesley. (Original work published 1968)
- Andersen, S. M., & Glassman, N. S. (1996). Responding to significant others when they are not there: Effects on interpersonal inference, motivation, and affect. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 263–321). New York: Guilford.
- Ashton, H. R. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, 52, 292–306.
- Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it. *Risk Analysis*, 13, 183–201.
- Baron, J., & Ritov, I. (1993). Intuitions about penalties and compensation in the context of tort law. *Journal of Risk and Uncertainty*, 7, 17–35.
- Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 1, pp. 680–740). New York: McGraw-Hill.
- Baumeister, R. F., & Leary, M. F. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motive. *Psychological Bulletin*, 117, 497–529.
- Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, 110, 73–92.
- Boring, E. G. (1963). *History, psychology, and science: Selected papers*. New York: Wiley.
- Brunsson, N. (1989). *The organization of hypocrisy: Talk, decisions, and actions in organizations*. New York: Wiley.
- Burger, J. M. (1981). Motivational biases in the attribution of responsibility: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin*, 90, 496–512.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York: Cambridge University Press.
- Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, 38, 230–256.
- Durkheim, E. (1976). *The elementary forms of the religious life* (2nd ed.). London: Allen & Unwin. (Original work published 1925)
- Edgerton, R. B. (1985). *Rules, exceptions, and social order*. Berkeley: University of California Press.
- Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, 69, 109–135.
- Feyerabend, P. (1993). *Against method*. New York: Verso.
- Fiske, A. P. (1991). *Structures of social life: The four elementary forms of social relations*. New York: Free Press.
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress spheres of justice. *Political Psychology*, 18, 255–297.
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model. *Psychological Bulletin*, 117, 39–66.
- Friedrich, J. (1993). Primary error detection and minimization strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, 100, 289–319.
- Geertz, C. (2000). *Local knowledge: Further essays in interpretive anthropology*. New York: Basic Books.
- Gergen, K. J. (1986). Correspondence versus autonomy in the language of human understanding. In D. W. Fiske & R. Schweder (Eds.), *Metatheory in social sciences* (pp. 136–162). Chicago: University of Chicago Press.
- Gigerenzer, G., Todd, P., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46, 107–119.
- Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Inferences, heuristics and biases: New directions in judgment under uncertainty*. New York: Cambridge University Press.
- Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, 78, 211–222.
- Glucksberg, S., & McGlone, M. S. (1999). When love is not a journey: What metaphors mean. *Journal of Pragmatics*, 31, 1541–1558.
- Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology*, 29, 781–795.
- Green, M., Visser, P., & Tetlock, P. E. (2000). Coping with accountability cross-pressures: Low-effort evasive tactics and high-effort quests for complex compromises. *Personality and Social Psychology Bulletin*, 26, 1380–1391.
- 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. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603–618.
- Hagafors, R., & Brehmer, B. (1983). Does having to justify one's judgments change the nature of the judgment process? *Organizational Behavior and Human Performance*, 31, 223–232.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.
- Hamilton, V. L. (1980). Intuitive psychologist or intuitive lawyer: Alternative models of the attribution process. *Journal of Personality and Social Psychology*, 39, 767–773.
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new

- perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78, 81–91.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Higgins, E. T. (1989). Self-discrepancy theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 93–136). New York: Academic Press.
- Higgins, E. T. (1996). Knowledge activation. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford Press.
- Hirschfeld, L. A., & Gelman, S. A. (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
- Horowitz, D. L. (1985). *Ethnic groups in conflict*. Berkeley: University of California Press.
- Jacobs, J., & Potter, K. (1998). *Hate crimes: Criminal law and identity politics*. New York: Oxford University Press.
- James, W. (1983). *The principles of psychology*. Cambridge, MA: Harvard University Press. (Original work published 1890)
- Kahn, B. F., & Baron, J. (1995). An exploratory study of choice rules favored for high-stakes decisions. *Journal of Consumer Psychology*, 4, 305–328.
- Kahneman, D., & Tversky, A. (2001). *Choice, values, and frames*. New York: Cambridge University Press.
- Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. W. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–21). Morristown, NJ: General Learning Press.
- Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.
- Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin*, 122, 250–270.
- Kennedy, J. (1993). Debiasing audit judgment and accountability: A framework and experimental results. *Journal of Accounting Research*, 31, 231–245.
- Kruglanski, A. W. (1990). *Lay epistemics and human knowledge*. New York: Plenum.
- Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, 19, 448–468.
- Kruglanski, A. W., & Webster, D. (1996). Motivated closing of the mind: Seizing and freezing. *Psychological Review*, 103, 263–268.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 8–101). Cambridge, England: Cambridge University Press.
- Laudan, L. (1997). *Progress and its problems*. Berkeley: University of California Press.
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, 24, 563–574.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255–275.
- Lerner, M. J., & Lerner, S. C. (1981). *The justice motive in social behavior: Adapting to times of scarcity and change*. New York: Plenum Press.
- McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1–47). New York: Academic Press.
- Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
- Meehl, P. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108–141.
- Mero, N., & Motowidlo, S. (1995). Effects of rater accountability on the accuracy and favorability of performance ratings. *Journal of Applied Psychology*, 80, 517–524.
- Miller, D. T., & Vidmar, N. (1981). The social psychology of punishment reactions. In M. J. Lerner & S. C. Lerner (Eds.), *The justice motive in social behavior* (pp. 145–168). New York: Plenum Press.
- Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102, 246–268.
- Murphy, R. (1994). The effects of task characteristics on covariation assessment: The impact of accountability and judgment frame. *Organizational Behavior and Human Decision Processes*, 60, 139–155.
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace, & World.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview Press.
- Pelham, B. W., & Neter, E. (1995). The effect of motivation of judgment depends on the difficulty of the judgment. *Journal of Personality and Social Psychology*, 68, 581–594.
- Peng, K., & Nisbett, R. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, 51, 252–264.
- Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin*, 25, 1067–1077.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Polinsky, A. M., & Shavell, S. (1999). Punitive damages: An economic analysis. *Harvard Law Review*, 111, 869–895.
- Radcliffe-Brown, A. (1952). *Structure and function in primitive society*. London: Cohen and West.
- Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, 65, 429–447.
- Ridley, M. (1994). *The Red Queen: Sex and the evolution of human nature*. New York: Macmillan.
- Samuelson, W., & Zeckhauser, R. (1988) Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1, 7–59.
- 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.
- Schlenker, B. R. (Ed.). (1985). *The self and social life*. New York: McGraw-Hill.
- Schwarz, N. (1994). Judgment in a social context. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 123–162). San Diego, CA: Academic Press.
- Scott, M. B., & Lyman, S. (1968). Accounts. *American Sociological Review*, 33, 46–62.
- Searle, J. (1998). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Semin, G. R., & Manstead, A. S. R. (1983). *The accountability of conduct: A social psychology analysis*. London: Academic Press.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49, 11–36.
- Siegel-Jacobs, K., & Yates, F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, 1, 1–17.

- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16, 158–174.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51, 416–446.
- Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology*, 77, 419–426.
- Skitka, L., & Tetlock, P. E. (1993). Providing public assistance: Cognitive and motivational processes underlying liberal and conservative policy preferences. *Journal of Personality and Social Psychology*, 65, 1205–1224.
- Sniderman, P., Fletcher, J., Russell, P., & Tetlock, P. E. (1996). *The clash of rights: Liberty, equality, and legitimacy in liberal democracy*. New Haven, CT: Yale University Press.
- Sniderman, P., & Tetlock, P. E. (1986). Interrelationship of political ideology and public opinion. In M. Hermann (Ed.), *Handbook of political psychology* (pp. 232–260). San Francisco: Jossey-Bass.
- Staw, B. (1980). Rationality and justification in organizational life. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 2, pp. 1–57). Greenwich, CT: JAI.
- Steele, C. M. (1988). The psychology of self-affirmation. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). San Diego, CA: Academic Press.
- Stryker, S., & Statham, A. (1985). Symbolic interaction and role theory. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1, pp. 311–378). New York: Random House.
- Sunstein, C. R., Kahneman, D., & Schkade, D. (1998). Assessing punitive damages. *Yale Law Journal*, 107, 2071–2125.
- Suppe, F. (1974). *The structure of scientific theories*. Urbana: University of Illinois Press.
- Taylor, K. (1975). Testing credit and blame attributions for choices under ambiguity. *Organizational Behavior and Human Performance*, 64, 128–137.
- Tesser, A. (2000). On the confluence of self-esteem maintenance mechanisms. *Personality and Social Psychology Review*, 4, 290–299.
- Tetlock, P. E. (1983a). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, 45, 74–83.
- Tetlock, P. E. (1983b). Accountability and perseverance of first impressions. *Social Psychology Quarterly*, 46, 285–292.
- Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, 48, 227–236.
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 331–376). New York: Academic Press.
- Tetlock, P. E. (1998). Losing our religion: On the collapse of precise normative standards in complex accountability systems. In R. Kramer & M. Neale (Eds.), *Influence processes in organizations: Emerging themes in theory and research* (pp. 168–189). Thousand Oaks, CA: Sage.
- Tetlock, P. E. (1999). Accountability theory: Mixing properties of human agents with properties of social systems. In J. Levine, L. Thompson, & D. Messick (Eds.), *Shared cognition in organizations: The management of knowledge* (pp. 255–278). Hillsdale, NJ: Erlbaum.
- Tetlock, P. E. (2000a). Cognitive biases and organizational correctives: Do both disease and cure depend on the political beholder? *Administrative Science Quarterly*, 45, 293–326.
- Tetlock, P. E. (2000b). Coping with trade-offs. In S. Lupia, M. McCubbins, & S. Popkin (Eds.), *Political reasoning and choice* (pp. 237–262). Berkeley: University of California Press.
- Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, 57, 388–398.
- Tetlock, P. E., & Boettger, R. (1994). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making*, 7, 1–23.
- Tetlock, P. E., & Kim, J. (1987). Accountability and judgment in a personality prediction task. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, 52, 700–709.
- 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., & Lerner, J. (1999). The social contingency model: Identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In S. Chaiken & Y. Trope (Eds.), *Dual process models in social psychology* (pp. 542–551). New York: Guilford Press.
- Tetlock, P. E., & Levi, A. (1982). Attribution bias: On the inconclusiveness of the cognition–motivation debate. *Journal of Experimental Social Psychology*, 18, 68–88.
- 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). Revisiting 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* (pp. 37–64). Hillsdale, NJ: Erlbaum.
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, 57, 632–640.
- Tetlock, P. E., Visser, P., Singh, R., Polifroni, M., Scott, A., Elson, B., & Mazzocco, P. (2002). *People as intuitive prosecutors: The fair-but-biased-and-occasionally-erratic model*. Unpublished manuscript, University of California, Berkeley.
- Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology*, 66, 474–489.
- Thompson, L. (1995). They saw a negotiation: Partisanship and involvement. *Journal of Personality and Social Psychology*, 68, 839–853.
- Thompson, L. (2001). *The mind and heart of the negotiator*. Englewood Cliffs, NJ: Prentice-Hall.
- Tyler, T. R., & Smith, H. J. (1998). Social justice and social movements. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 595–629). New York: McGraw-Hill.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press. (Original work published 1934)
- Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment. *Journal of Personality and Social Psychology*, 68, 36–51.
- Zajonc, R. B. (1965, July 16). Social facilitation. *Science*, 149, 269–149.

Received March 9, 2000

Revision received March 19, 2001

Accepted July 31, 2001 ■