CHAPTER 6

Disentangling Reasons and Rationalizations: Exploring Perceived Fairness in Hypothetical Societies

Gregory Mitchell and Philip E. Tetlock

Abstract

Political psychologists often treat explicit explanations for political views as rationalizations rather than reasons and favor unconscious motives and cognitive processes as the key determinants of political ideology. We argue that "transparent-motive" theories are often dismissed too quickly in favor of "subterranean-motive" theories. We devote this chapter to finding common methodological ground for clarifying, testing, and circumscribing the claims of both the transparent-motivational theorists and the subterranean-motivational theorists, and we pose a series of empirical questions designed to explore predictions that might provide evidence that justifications are not mere by-products of the functional imperative to defend the status quo but rather functionally autonomous constellations of ideas capable of independently influencing policy.

Over the last 150 years, behavioral scientists have repeatedly revealed their deep skepticism of the reasons that ordinary mortals offer for their political views. As an epistemic community, we have shown a marked preference for "subterranean-motivational theories"—theories predicated on the assumption that people have little access to the true drivers of their judgments. Indeed, under this subterranean rubric we include a truly diverse mix of scholars, ranging from the Freudian to the evolutionary to the Marxist: psychodynamic scholars, such as Lasswell (1930) and Adorno and colleagues (1950), who view political attitudes as the product of the displacement of private motives onto public objects rationalized in terms of the common good; evolutionary and social-dominance theorists who argue that people derive psychic gratification from exercising symbolic dominance over those below them in the pecking order (Sidanius, Levin, Federico, & Platto, 2001); system justification theorists who posit a deep-rooted psychological tendency to justify existing status hierarchies (a tendency that bears a marked family resemblance to the classic Marxist notion of false consciousness—Jost & Banaji, 1994; Jost, 1995); and social identity theorists who maintain that

self-esteem needs guided by rapid-fire categorization processes are responsible for the widespread phenomenon of invidious ingroup-outgroup stereotyping (Rubin & Hewstone, 2004).

We do not doubt that good reasons often exist for doubting the reasons people offer for their policy stands—and for suspecting that these reasons do not capture the true causal dynamics behind their opinions. We readily concede that there are serious cognitive limits on our introspective access to mental processes—and powerful sources of social desirability distortion operating on what people are willing to say. But, like the plain-spoken sociologist, C. Wright Mills (1940), we worry about "motive-mongering." Indeed, if we were inclined to subterranean-motivational speculation of our own, we might suggest that subterranean motives drive the intense curiosity of social scientists in subterranean motives—be it the preventive goal of ensuring that their research conclusions not be labeled obvious or the promotional goal of being proclaimed profound. We also worry that in a discipline as ideologically lopsided as political psychology, the subterranean-motivational speculation can easily become skewed against groups in collective disfavor (Arkes & Tetlock, 2004; Haidt & Graham, 2007; Mitchell & Tetlock, 2006; Redding, 2004; Sniderman & Tetlock, 1986; Suedfeld & Tetlock, 1991).

Whatever the merits of such speculation, we are acutely aware of how difficult it is to resolve disputes over the merits of transparent versus subterranean-motivational theories—and distinguish reason from rationalization in social, personality, organizational, and political psychology. (One of us wrote many years ago on the indeterminacy problems that bedeviled far less politically charged efforts to distinguish cognitive from motivational, and "intrapsychic" from impression management, explanations in a variety of experimental paradigms; Tetlock & Levi, 1982; Tetlock & Manstead, 1985.) But we do think it vital—for reasons laid out later—to try. And we devote this chapter to finding common methodological ground for clarifying, testing, and circumscribing the claims of both the transparent-motivational theorists and the subterranean-motivational theorists.

We divide our chapter into three sections. In the first, we make our case for an underutilized methodology: transforming political-philosophical thought experiments into psychological experiments. In the second section, we describe a series of hypothetical-society laboratory studies that we have conducted over the last 15 years to explore the value judgments that guide people when they make "macro-distributive" judgment calls about the fairness of resource allocations on a societal and even global scale. These studies allow us to compare how closely the belief and value systems of actual human beings resemble a host of conceptual ideal types, including intuitive Rawlsians (who give priority to raising the guaranteed safety net income),

intuitive libertarians (who give priority to minimizing redistribution and maximizing aggregate wealth), intuitive Marxists (who reject all forms of class subjugation), intuitive Durkheimians (who place a premium on the solidarity-expressive functions of punishment), and value-pluralist pragmatists (who strike varying compromises between equality and efficiency—and other values). In the third section, we pose a series of questions designed to explore what, if any, predictions can be derived from system justification theory (SJT) and kindred subterranean formulations in the hypothetical-society context—and to determine the types of evidence necessary to induce advocates of such theories to change their minds: to view justifications not as by-products of the functional imperative to defend the status quo but rather as functionally autonomous constellations of ideas capable of independently influencing policy. The theoretical debate is as old as that between Marx and Weber: How do interests (traditionally stressed by Marxists) and ideas (traditionally stressed by Weberians) interact to shape our vision of who we collectively are and what we should collectively aspire to achieve?

TURNING THOUGHT EXPERIMENTS INTO LAB EXPERIMENTS

Carefully conducted thought experiments help philosophers clarify the role of competing principles and assumptions in their normative arguments, much like laboratory experiments help psychologists clarify the role of different variables in cause–effect relationships. In the mind of a philosopher committed to working out the logical implications of propositions in alternative worlds, the thought experiment can be a rigorous means to an end: "She follows through all the relevant implications of altering one part of her worldview and attempts to construct a coherent model of the situation she is imagining. The rigor with which thought experimenters attempt to answer 'what if' questions is what differentiates thought experiments from daydreams and much fiction. . . . The thought experimenter is committed to rigorously considering all relevant consequences in answering the 'what if' questions" (Cooper, 2005, p. 337).

Thought experiments, however, even when done carefully and with a mind open to possibilities rather than searching for confirmation, lack the transparency and replicability deemed essential to scientific research (Bunge, 1961). These weaknesses lead many to dismiss the thought experiment as a path to reliable knowledge (see Sorenson, 1992, Chapter 2). Thus, when scientists successfully employ thought experiments—Galileo, Newton, and Einstein come quickly to mind—the resulting theories must be couched in publicly testable terms to qualify as scientific (Dennett, 2003).

Thought experiments also present serious external validity concerns. Whereas laboratory researchers can make some claim that their findings represent the views of a cross-section of college students reacting to real, if simulated, situations, thought experimenters can make no claim that their findings represent the views of people in general, or even philosophers specifically, reacting to realistic simulations. Indeed, many philosophical debates persist because philosophers reach different conclusions about hypothetical cases or the validity of background assumptions in these cases (e.g., Coleman, 2000), and the very purpose of many thought experiments is to create *unreal* situations that can exist only in the imagination (Souder, 2003).

For the empiricist who finds a thought experiment interesting but doubts the reliability and generalizability of its product, a simple solution exists: reduce the thought experiment to concrete terms that can be reproduced as written scenarios and ask subjects to react to the scenarios to see what trends emerge (e.g., Machery, Mallon, Nichols, & Stich, 2004). The emerging field of experimental philosophy seeks to do just this with a variety of conundrums (Knobe, in press). But that view emphasizes what laboratory studies can do for thought experiments and philosophical explorations. In our view, thought experiments can do much for laboratory studies and the social-psychological explorations of a variety of topics, including the psychological foundations of lay conceptions of justice.

Empirical studies into the perceived justice of real-world outcomes and procedures confront difficulties that may be partially remedied by incorporating elements of thought experiments into these studies. First, and almost impossible to control in empirical studies of public reasoning on current controversies, is the problem that public opinion often depends on mixtures of emotionally charged political values (such as liberty, equality, religious purity, and national sovereignty) and technically complex matters of fact (such as whether individual or societal conditions are greater determinants of economic outcomes or whether tying welfare benefits to work requirements will encourage self-sufficiency). When causal relations and policy effects are difficult to determine, a powerful temptation exists to arrange one's beliefs about the facts in convenient ways that minimize dissonance and mental strain (e.g., Herrmann, Tetlock, & Diascro, 2001; Mitchell, Tetlock, Mellers, & Ordóñez, 1993; Skitka, 1999). For instance, Skitka and Tetlock (1992, 1993) found that liberals and conservatives held different preexisting beliefs about the causes of public assistance and, as a result, made different trade-offs in a mock public aid allocation task. Thus, surveys that find different views about distributive justice between liberals and conservatives, but fail to check for differences in background beliefs, may mistakenly attribute response differences to value differences. Conversely, surveys that find agreement across

groups regarding distributive justice and the propriety of redistribution may simply reflect widespread mistaken beliefs about underlying facts, such as the degree of economic mobility in a society (see Ferrie, 2005; Fong, 2005) or the proportion of families in different socio-economic categories (see Kluegel, Csepeli, Kolosi, Orkeny, & Nemenyi, 1995). These problems become particularly acute when one studies the impact of macroeconomic variables and system-level conditions on individual judgments of justice, but informational problems may arise whenever key facts are vague or disputed (e.g., the bargaining studies of Babcock, Loewenstein, Issacharoff, & Camerer, 1995).

To overcome such confusion, we took a page from the philosopher's book on thought experiments and developed a "hypothetical-society paradigm," in which experimental participants judge the justice of different economic and legal arrangements in hypothetical societies (Mitchell et al., 1993).[1] This paradigm turns the classic weakness of thought experiments, their unreality and subjectivity, into a strength: because the experimenter is the creator of the hypothetical societies, the experimenter controls the structure of these societies down to the tiniest technical details, including the location of the poverty line and percentage of persons below it, mean income and income variance within the society, levels of redistribution and welfare services, the level of meritocracy (i.e., the degree to which individual merit versus other factors determine economic outcomes), and whether the hypothetical society is in the "original position" or considering changes to existing procedures and distributions. Using the hypothetical-society approach, an investigator can examine which features of societies are most important to people's judgments of social justice and determine how these judgments change as features of the societies change. In short, the paradigm allows

[1] The inspiration for the hypothetical-society paradigm was Rawls' impartial reasoning device, the "veil of ignorance," which seeks to "nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage" (Rawls, 1971, p. 136). Behind the veil, "no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like" (Rawls, 1971, p. 137). Because we cannot divest participants of self-knowledge as required by a true veil of ignorance, we chose instead to remove narrow self-interest as an influence on judgments by having participants disinterestedly evaluate hypothetical societies. Our efforts to approximate Rawls' original position were predated by Brickman (1977) and Frohlich and Oppenheimer (1992). Appendices in Mitchell et al. (1993) and Mitchell, Tetlock, Newman, and Lerner (2003) provide detailed descriptions of the hypothetical societies, the instructions given to participants, and the participants' tasks.

researchers to unconfound the influence of factual beliefs from that of value orientations in judgments of justice. Because individuals tend to avoid value trade-offs, often by interpreting ambiguous or disputed facts in a favorable light (e.g., Tetlock & McGuire, 1986), this ability to manipulate value conflict confers considerable experimental advantages.

One key benefit of importing hypothetical societies into the laboratory is the control one gains over otherwise complex and sharply contested matters of fact. A second, arguably equally important, benefit involves the control one gains over the influence of selfish interests. A common problem in empirical studies of justice is that of distinguishing biased from unbiased judgments (see Fong et al., 2006; Konow, 2005; Liebig, 2001). The hypothetical-society paradigm allows researchers to place participants in the position of impartial spectator: researchers who want to eliminate or minimize the role of material self-interest and social influence on judgments ask participants to make anonymous judgments about hypothetical societies with no material implications for themselves. Alternatively, researchers interested in the role of social influences can ask participants to explain or justify their judgments under various accountability conditions, or can manipulate the group identities involved, whereas researchers interested in the influence of material self-interest can alter the method to have participants imagine themselves inside the society or ask them to allocate resources within the society (using either hypothetical or real pay-offs).

In our hypothetical-society studies, we have favored experimental manipulations that place the participant in the role of impartial spectator, in order to capture unbiased judgments of justice. As a number of studies have shown, when participants have a stake in the distribution at hand, egocentric and ingroup biases will often influence participants' judgments about the fairness of these distributions (Bar-Hillel & Yaari, 1993; Epley & Caruso, 2004; Frohlich & Oppenheimer, 1997, 2000; Greenberg, 1983; Konow, 2005; Messick & Sentis, 1983; Pillutla & Murnighan, 2003). We cannot trust that unbiased judgments of justice will be given when individuals judge their own situations, and so, if we seek to know what people believe justice ideally requires, "thought experiments trump real experiments (Cooper, 2005, p. 344)."[2]

[2] Cooper (2005) makes this point in the context of thought experiments involving trade-offs between avoiding torture to oneself versus avoiding harm to others, where what we seek to know is not what the tortured person would actually do but what a rational person should do in such a situation: "The judgments of people contemplating what should be done under torture are more reliable than the judgments of people actually being tortured (p. 344)."

That said, judgments about justice by detached observers of hypothetical societies may still be useful guides about judgments of justice in real societies. Most obviously, to the extent that hypothetical societies and real societies possess common features important to lay conceptions of justice, judgments about justice in the hypothetical societies may generalize to real societies. Even with highly artificial scenarios, judgments about hypothetical societies can identify pivotal points of agreement and disagreement and explain how factual beliefs and value differences combine to produce either ideological convergence or divergence. For instance, in our first set of hypothetical-society studies (Mitchell et al., 1993), we found surprisingly wide agreement on the importance of minimum safety nets, even in perfect meritocracies. Hypothetical-society studies also shed light on which social arrangements may have the greatest "psychological stability" (see Elster, 1995). Our studies have found, for example, that conservatives are more sensitive to waste in income redistribution policies ("leaky buckets") than liberals when the redistribution was meant for deserving recipients (Mitchell et al., 2003), suggesting that the psychological stability of policy arrangements depends on the mix of liberal and conservative decision makers, the perceived deservingness of would-be recipients in the applicant pool, and the leakiness of the income transfer process (Skitka & Tetlock, 1992, 1993).

More ambitiously, to the extent that the judgments individuals reach as impartial spectators cause individuals to reflect on just distributions in their own societies, the hypothetical-society paradigm could be used as a device to foster deliberation about social policy (e.g., Fishkin, 1992). If used in this sense, the hypothetical-society paradigm performs a "reflective equilibrium" function (Rawls, 1971; see Daniels, 1996), possibly leading persons to abandon their initial intuitions or change their views about what justice requires once they are compelled to work their way through a series of controlled thought experiments.

In sum, the hypothetical-society paradigm can be a powerful tool for overcoming the limitations of alternative methods, including the problems of replication and "idiosyncratic intuition" that plague philosophical thought experiments on justice, and the problems of partiality—with respect both to facts and motivations—that plague lab and field studies of justice.[3]

[3] A closely related device for studying justice judgments is the vignette study (e.g., Bukszar & Knetsch, 1997; Konow, 2003). Vignette studies typically ask experimental or survey participants to judge whether justice occurred in some realistic but imaginary event (e.g., pay distribution in a hypothetical work setting). The advantage of a vignette study over a hypothetical-society study is that the former possesses greater external validity. The

TAKING STOCK OF THE CURRENT EMPIRICAL YIELD FROM HYPOTHETICAL-SOCIETY STUDIES

Most studies using the hypothetical-society paradigm examine the perceived justice of societal-level patterns of distribution or rules for distributing resources within a society, and so we begin with findings from these studies on social justice. We first utilized the paradigm to examine how people make macro-level trade-offs between equality and efficiency. Specifically, we described for participants three different societies that differed in their levels of meritocracy, with the correlation between effort and outcome being high (a correlation of 0.9), medium (a correlation of 0.5), or low (a correlation of 0.1), and we displayed income distributions within each society that varied in terms of their equality (income variance) and efficiency (average income). (For a full description of the hypothetical-society instructions and stimuli, see the Appendix to Mitchell et al., 1993.) Participants were asked to imagine themselves as outside observers of the societies and to make pair-wise comparisons of all possible income distributions for one of the societies, choosing which distribution in each pair was fairer, so that a fairness ranking of income distributions could be derived for each individual within a society and for groups of individuals across all three hypothetical societies. These fairness rankings were then compared to a variety of ideal-type fairness rankings for the income distributions derived from competing theories of distributive justice, namely, egalitarianism (emphasizing equality), utilitarianism (emphasizing efficiency), a Rawlsian maximin principle (emphasizing quality subject to efficiency constraints), and Boulding's (1962) compromise theory (emphasizing efficiency subject to equality constraints—in which minimum equality is required by the government ensuring a safety net for the poor, but the goal of prosperity is encouraged by rewarding individual effort above this social safety net).

Consistent with Boulding's (1962) compromise theory, as well as with later value-pluralism ideas (Tetlock, 1984, 1986), both liberals and conservatives were willing to accept considerable inequality of wealth in high-meritocracy societies but with the reservation that distributions allowing people to fall below the poverty line remained unpopular for both ideo-

disadvantage of the vignette study relative to the hypothetical-society study is that, because the participant may find the vignette more realistic and familiar, the participant may find it more difficult to imagine or accept the stipulated facts and detach herself from the situation about which she is supposed to be an impartial judge, and the researcher has less freedom when creating hypothetical situations.

logical groups even in high-meritocracy societies (a finding similar to that of Frohlich and Oppenheimer, 1992, whose experimental groups favored utilitarianism above a floor constraint). However, a majority of liberals *and* conservatives favored a Rawlsian "maximin" approach (Rawls, 1971) to the distribution of wealth in low- and moderate-meritocracy societies (a finding at odds with Frohlich and Oppenheimer [1992] and one that suggests that implicit assumptions of meritocracy may have driven Frohlich and Oppenheimer's groups to favor a modified utilitarianism). Liberals and conservatives disagreed most sharply when the reward structure in the hypothetical society was most ambiguous (i.e., in the moderate-meritocracy society), with liberals tending toward greater equality and conservatives toward greater efficiency in such societies. Thus, we found that, for both ideological groups, beliefs about the level of meritocracy in the hypothetical society moderated value trade-offs, suggesting that ideological disagreements about social justice may arise just as often from different beliefs about the nature of the reward structure in society as from value differences (compare Fong, 2004, reporting that target-specific beliefs regarding individual responsibility for economic outcomes drove attitudes toward redistributive policies).

In a subsequent hypothetical-society study using similar experimental stimuli (Mitchell et al., 2003), we again found that the perceived level of meritocracy in a society greatly affected judgments about the justice of distributions in that society, with support for greater equality (and less prosperity) strongest at low levels of meritocracy and support for greater prosperity (and less equality) strongest at high levels of meritocracy. In this study, we also manipulated whether participants were judging the fairness of income distributions as if they were alternative *original distributions* for each society versus as if they were *redistributions* of income from an existing distribution in each society. When participants judged *re*distributions (i.e., when it was clear that income would be taken from one group and redistributed to another), both liberals and conservatives became more sensitive to the level of meritocracy in the society, and considered redistributions in the moderate- and high-meritocracy societies to be significantly less fair than equivalent distributions viewed as alternative starting distributions in the same societies. Further, for all three societies, including a "no-meritocracy" society with no relation between effort and outcomes, participants judged redistributions that led to losses in equality or losses in prosperity to be less fair than when they simply judged the fairness of these distributions as possible "original positions," suggesting a vicarious type of loss aversion at work even in judgments about hypothetical redistributions.

These findings highlight both the practical problems faced by advocates of redistributive policies and the conceptual problems faced by political phi-

losophers grappling with whether (or when) the distributive–redistributive distinction should count in normative theories of justice. These findings also highlight interpretive ambiguities that arise for psychological theorists in characterizing the true causes of resistance to redistribution. If people resist redistribution because they have a tendency to adopt the status quo as their reference point and to be loss-averse (directly or vicariously), as prospect theory predicts, is it accurate or fair to characterize such automatic psychophysical processes with as politically charged a label as system justification?[4]

Providing further empirical evidence against a unidimensional conception of distributive justice such as utilitarianism and in favor of a multidimensional conception such as in Boulding's compromise theory, Ordóñez and Mellers (1993) used the hypothetical-society paradigm to examine whether individuals make trade-offs when judging social fairness. They found that the great majority of participants did make trade-offs between different principles, but the principles that most concerned their participants were need and desert, with participants wanting to ensure a minimum salary for all members of the hypothetical society but also wanting to provide just deserts to those who worked hard in the society; equality and efficiency were of little concern to participants in this study. This study is also interesting because Ordóñez and Mellers asked participants to make judgments about the fairness of societies, but also to express preferences for societies as places to live. They found that most participants rated high-meritocracy societies as fair, but they preferred to live in societies with high minimum incomes (a finding that applied particularly to participants with self-reported low socio-economic status). This finding is consistent with the view that the hypothetical-society paradigm can be used to elicit both refined justice judgments and preference judgments reflecting self-interest rather than ethical concerns.

Recently, Scott and his colleagues (Scott, Matland, Michelbach, & Bornstein, 2001) employed a variant of the hypothetical-society paradigm to compare the role of equality, efficiency, merit, and need in people's judgments

[4] Although system justification theorists draw on status quo bias research to support their theory (Jost, 2001), we see nothing intrinsically system justifying about prospect theory. Prospect theory processes can just as easily fuel moral outrage as moral complacency toward the status quo (e.g., Kahneman, Knetsch, & Thaler, 1986). For instance, prospect theory identifies factors that should make it easier to mobilize the losers in an earlier "illegitimate" round of redistribution to take big risks to restore the status quo ante (McDermott, 1998). Similar processes could also be at work driving intense resistance to the impact of global capitalism on climate change or driving Islamic radicals to restore the original Islamic state. From our standpoint, the "system" in system justification is underdefined.

of distributive justice, finding that each principle proved influential to some extent, except that merit considerations only influenced women's judgments of justice in this study. In a second study, this research group (Michelbach, Scott, Matland, & Bornstein, 2003) replicated their finding that individuals try to balance equality, efficiency, need, and merit in their justice judgments, but they failed to replicate the gender gap in meritocracy concerns found in their first study. However, this second study did find a racial gap in meritocracy concerns, with the nature of equality-efficiency trade-offs by White participants dependent on their merit assumptions but not those of racial minorities. Also, Michelbach and colleagues (2003), with a refinement to the hypothetical-society paradigm that provided a cleaner test between egalitarianism and Rawls' maximin principle than that employed in our original study (Mitchell et al., 1993), found that a significant number of participants endorsed the maximin principle, but many others deemed merit an important principle and deviated from a strict adherence to the maximin principle.

These studies by Scott and others support our original finding (Mitchell et al., 1993) that impartial spectators often place considerable weight on equality and the maximin principle when making justice judgments, especially when meritocracy is lacking. However, these studies and their findings of gender and racial gaps in the weight placed on meritocracy in justice judgments also caution against generalizations about the role of meritocracy in justice judgments and suggest that White men, women, and minorities, who may have had very different experiences with meritocracy in the United States, may have difficulty divesting themselves of their life experiences and placing themselves in the position of impartial observer.

Most recently, we used the hypothetical-society paradigm to examine the longstanding debate in legal theory on the relationship between corrective justice and distributive justice (Mitchell & Tetlock, 2006).[5] Some legal philosophers claim that corrective justice is parasitic on distributive justice, with the one who has caused a harm (the "tortfeasor" in legalese) having a duty only to repair the harm imposed on another if the underlying distribution of goods disturbed was just, whereas others claim that corrective justice and distributive justice impose independent moral demands on mem-

[5] Corrective justice stipulates, roughly, that a person who wrongfully causes harm to another has a duty to repair the harm (see Forde-Mazrui, 2004). The concept of corrective justice goes back to Aristotle and his distinction between justice in transactions, or arithmetic forms of justice, and justice in overall distributions within a polity, or geometric forms justice (see Weinrib, 2002).

bers of a society that cannot be traded off against one another. To test the competing views, we constructed distributively just and unjust hypothetical societies—with distributive justice operationalized in terms of meeting needs, equality, and desert—and told participants of certain intentional and unintentional torts occurring in these societies that upset the distribution of resources in these just and unjust societies. The task for participants was to declare whether justice required the tortfeasor to make the victim of the tort whole, as a norm of corrective justice would require.

We found, somewhat to our surprise in light of much empirical research showing the context sensitivity of competing norms of justice (see Miller, 1999), that the norm of corrective justice consistently trumped distributive justice norms, even where enforcing the norm of corrective justice would lead to a more unjust distribution of resources in the community (i.e., in a society with no meritocracy, where an undeserving poor man had to compensate an undeserving rich man for harm negligently done by the poor man, leading to greater inequality and greater unmet needs). Indeed, in many conditions, there was near unanimity that the tortfeasor should make the victim whole, even when participants judged the society to be unjust and the victim had insurance that would cover the harm done.

Only under conditions of extreme injustice in the distribution of resources did most participants deem it just that tortious harm go unrepaired. Thus, in a hypothetical society in which a racial minority perpetuated its hold over power through discriminatory policies, most liberal participants and some conservative participants felt that justice did not require that an impoverished member of the oppressed majority compensate a wealthy member of the racially oppressive minority who had been harmed by the former's negligence. However, when the poor member of the racially oppressed class intentionally stole a valuable watch owned by the rich man, most participants judged this action out of bounds as a matter of justice, even though it arguably is a form of self-help that would lead to a more just distribution of wealth in this racially unjust society (with half of the liberal participants and more than half of the conservative participants judging justice to require compensation for this intentional tort).

Such findings are significant in at least two ways. First, they demonstrate the importance of adding corrective justice norms to the list of justice concerns that may be triggered by context (see Konow, 2003), and they illustrate that this norm will be potent, and likely dominant, in contexts that emphasize transactional harms. These findings emphasize the importance placed on personal responsibility for rectifying harms done, at least among our sample of Americans, and cast into doubt the popularity of social compensation schemes for accidents, such as New Zealand's taxpayer-funded,

no-fault accident fund. To date, there has been little research into corrective justice, but our findings point out the need to understand the scope, source, and function of the norm of corrective justice and its relation to retributive justice, which has received more empirical attention (e.g., Darley & Pittman, 2003; Tetlock et al., 2007), but both of which have received less attention than distributive and procedural justice.

Second, these findings further illustrate the malleability of the hypothetical-society paradigm. Outside the admittedly highly stylized hypothetical-society paradigm, it would be very difficult to disentangle competing theoretical positions on the relationship between norms of distributive and corrective justice. The simplicity of the paradigm makes it easy to eliminate confounding variables and test alternative explanations for why people view certain social arrangements to be just or unjust. We explore some of the untapped potential of the hypothetical-society paradigm in the next section.

USING HYPOTHETICAL SOCIETIES TO CLARIFY RIVAL THEORETICAL POSITIONS

The hypothetical-society paradigm arguably gives us a chance to glimpse relatively pure value judgments, undistorted by the usual real-world mix of either clashing interest groups or clashing ideological views of the magnitude and causes of social problems. We find that, although some respondents do fit sharply defined ideological ideal types—committed egalitarians and libertarians—the aggregate data are more consistent with an alternative portrait of how most people make decisions in these spectator roles: a value pluralism account (Berlin, 1990; Tetlock, 1986; Tetlock, Peterson, & Lerner, 1996). It is as if people were trying—not necessarily successfully—to balance competing values, with the relative importance of certain values holding quite firm against the counter-pressures thus far applied and the relative importance of other values showing considerable lability and context specificity.

The stablest commitments so far seem to be to a safety net and corrective justice. Like good egalitarian collectivists, people care a lot about ensuring that no one falls below a basic-need safety net across a wide range of circumstances (Frankfurt, 1987), and like good property-rights individualists (and also Durkheimians, in Tetlock et al., 2007), people care a lot about ensuring that norm violators are punished across a wide array of socio-economic background conditions. If we gave voice to these sentiments, they might sound like this: "Give us safety nets (for we know that people can fall far through no fault of their own—and in any event, it pains us to see others suffer), but hold all norm violators, even the poor, accountable to the precepts of corrective justice, lest we revert to the law of the jungle."

By contrast, other values oscillate more in importance across background societal conditions. Like good egalitarian collectivists, people give heaviest weight to equality when they think the society has deviated from the ideals of meritocracy, but like good capitalist individualists, people give heaviest weight to efficiency and wealth maximization—and resist redistribution most intensely—when society is highly meritocratic and the wealth transfer process inefficient (a "leaky bucket" for transferring assets). Also, intriguingly, people are most likely to polarize along ideological lines when there is greatest ambiguity about meritocracy—arguably the most realistic of the conditions in hypothetical-society experiments, as our participants consistently liken American society to the moderate-meritocracy society in our studies—perhaps a sign that real-world conditions create the most room for implicit ideological values (better to err in the leftward or rightward direction) to come into play.

Skeptics of the hypothetical-society paradigm could argue, however, that it only taps into relatively superficial psychological processes to which people have ready conscious access and that people are not embarrassed about revealing. The skeptics would be correct that we have thus far tended to take the intuitive political philosophies of our respondents at face value. If respondents say that they are Rawlsian egalitarians (Rawls, 1971) or Nozickian libertarians (Nozick, 1974) or value pluralists in the mold of Isaiah Berlin (1990), and respond in that spirit to our instruments, we classify them accordingly. These ideal-type belief system models are best classified as transparent-motivational theories that make the working assumption that people are lay political philosophers struggling to make sense of the world and balance reasonable arguments against each other. From the skeptics' perspective, we have yet to explore seriously the possibility that motives to which our respondents do not have conscious access (or might be embarrassed to admit) are swaying their judgments of macro-level distributive justice. It is useful, therefore, to consider how a system justification theorist might explain our data—and explore how we might reconfigure hypothetical-society experiments to clarify and eventually disentangle the predictions we might expect from SJT and alternative accounts, such as our own.

System justification theorists could argue that our findings are simply a special case of their own demonstrations that people will accept explanations that justify the status quo, regardless of the objective accuracy of the explanation (Haines & Jost, 2000). But our finding that respondents often favored changes to a status quo that they judged unjust seems hard to square with an authoritarian–acquiescence version of SJT. Nonetheless, system justification theorists could counter that the motive to system-justify operates only when one's own status or societal hierarchy is at stake, in which case the

hypothetical-society paradigm will be dismissed as too hypothetical to be relevant.[6] However, if cognitive and motivational components of system justification are triggered automatically by status-relevant stimuli (e.g., Jost & Hunyady, 2002), if system justification processes are triggered regardless of personal responsibility for the status quo (Jost & Hunyady, 2002), and if system justification beliefs comprise an "ideology" that people rely on to interpret, respond to, and assimilate new stimuli (e.g., Blasi & Jost, 2006; Jost, Banaji, & Nosek, 2004), then the hypothetical nature of our societies—in which we can simulate inequalities in existing societies but remove all ambiguity about causation—should not be a barrier to our experiments serving as a testing ground for SJT.[7]

Alternatively, system justification theorists could argue that hypothetical-society researchers have merely reconfirmed that people have a moral preference for social orders roughly similar to the world they currently inhabit: democratic capitalist states, with safety nets of varying height, committed to individualistic norms of justice. Indeed, we would never dispute that the societal status quo is a powerful anchor for moral-political judgment (even in hypothetical societies, as our distribution/redistribution mindset manipulation showed): we strongly suspect that if we could bring the vast numbers of antebellum Americans who regarded slavery as a reasonable

[6] To address this specific concern, we note that the hypothetical-society approach could be modified to fit a number of systems about which the experimenter could credibly claim to have undisputed factual information, but that are much less hypothetical or unreal than in our studies to date. Most promising would be a "hypothetical class action" study in which the parties have stipulated to all relevant factual matters and agree on the future impact of different remedies but disagree on the desirability of, or need for, different remedies. Participants then would be tasked with setting policy for the organization going forward, with the policy options set along a continuum anchored by status quo preservation on one end and radical reform on the other.

[7] Indeed, the experimental paradigm employed by Jost and Burgess (2000) and discussed in Jost (2001) bears some resemblance to our hypothetical-society studies. In that paradigm, the experimenters manipulate participants' perceptions of the relative socioeconomic success of alumni of their own university and a competing university to examine how these perceptions affect explanations for differential success and evaluations of these groups. Studies along these lines, in which arcane matters of public policy are chosen such that participants may be led to believe that facts associated with different policies are real, may be additional good candidates for some of the "stress testing" of system justification theory that we propose in the next few subsections.

accommodation in the mid-19th century into contemporary America, those individuals would bear little psychological resemblance to whatever pathological fringe of the current population endorses race war and the oppression of minorities.[8]

We would counter that, at minimum, the hypothetical-society paradigm has already revealed a good deal about what varying viewpoints consider plausible justifications for varying social orders. For instance, it is telling that even many hard-core conservatives embrace equality when confronted with a hypothetical society in which one's socio-economic status has been determined randomly, not by skill and hard work. And even many hard-core liberals embrace efficiency when confronted with a hypothetical society in which one's socio-economic status has been determined entirely by hard individual work, with no role for chance. If even the belief systems of hard-core ideologues (who might be hypothesized to resemble in profile extreme low and high scorers on the system justification scale) acknowledge boundary conditions on their belief systems, so, too, should researchers who are trying to model the political-psychological functioning of these belief systems. Indeed, we would argue that our studies, which focus on choices between alternative social systems, provide more direct evidence on the operation of putative system justification motives than do system justification studies that focus on attitudes toward high- versus low-status groups that typically are subject to both false- and veridical-consciousness interpretations.[9] From this standpoint, the largest lacuna in system justification research is the paucity of research into the motive–behavior linkage—it is one thing to argue

[8] We acknowledge, however, that the psychological similarities may be more pronounced between support for slavery in antebellum America and support for anti-redistributive policies in the early 21st century. But we caution against the historicist fallacy that those similarities shed light on who has the normative high ground in policy debates in the early 21st century. For instance, the same integratively simple style of reasoning that led Churchill to oppose self-government for India also led him to see Nazi Germany as an existential threat to the British Empire—and the same absolutist reasoning that led fire-eater defenders of slavery to secede from the United States also led abolitionists to pressure Lincoln to define the Civil War as a war against slavery (Tetlock, Armor, & Peterson, 1994).

[9] Certainly some system justification studies employ behavioral measures (e.g., Jost, Pelham, & Carvallo, 2002) and assess preferences and beliefs potentially relevant to the social order (e.g., Kay, Jimenez, & Jost, 2002; Jost, 1997), but many examine attitudes and stereotypes about ingroups and outgroups that vary in their socioeconomic status and do not directly examine system-justifying behaviors.

that humans are adept at rationalizing outcomes and quite another to argue that these rationalizations have deleterious effects on low-status groups (as Blasi & Jost [2006] suggest is true with respect to underutilization of the legal system by disadvantaged groups; see also O'Brien & Major [2005] and Jost & Thompson [2000] for evidence on the positive and negative effects of system-justifying beliefs on psychological well-being, respectively, for high- and low-status groups).

We would also counter that existing hypothetical-society research has barely scratched the surface of the conceptual complexities of macro-level distributive justice—and of how ordinary people reason their way through these dilemmas. The more we grapple with these complexities, the more sharply we will understand both the strengths and limitations of subterranean-motivational theories, such as SJT, and more transparent-motivation theories, such as the value pluralism model. Blasi and Jost (2006, p. 1124) stake out a provocative position on the generality of the system justification motive: "Most of the time, people have a general, inherently conservative tendency to accept the legitimacy of whatever 'pecking order' is in effect and to perceive existing institutions and practices as generally reasonable and just, at least until proven otherwise." We are unsure how much we disagree with this claim, but we do believe that the hypothetical-society paradigm provides a useful vehicle for clarifying the key points of ambiguity that cause us to withhold judgment. Accordingly, we devote the remainder of this chapter to identifying how the paradigm can be used to clarify and test the predictions of the rival theoretical camps.

Clarification is the critical first step because verbal theories can often be read in many ways, and this is true both of our belief system ideal types derived from hypothetical-society work and of SJT. With that key caveat, our reading of SJT is that the ideal-type system justifier should be automatically sympathetic, across a broad range of background conditions, to any hierarchy that resembles the system onto which that individual imprinted during political socialization (Jost, Fitzsimons, & Kay, 2004), whereas the ideal-type antithesis of a system justifier in the United States should strongly prefer equality (or rebelliousness) across an equally broad range of societal background conditions. Insofar as ideologues at either end of this continuum qualify their support for, or rejection of, inequality, we have evidence either that these observers are mindlessly allowing for exceptions already permitted in their home society or that these observers are thoughtfully qualifying their original one-size-fits-all ideological templates by taking individuating information into account. This difference is, in our view, a big one. If the latter, we have evidence for what we view as value-pluralism boundary conditions on system justification: people may justify the status quo only up to

the point at which they feel the status quo is justifiable given their internalized schemata and values for judging fair play. Put differently, such data would show that the justifications in system justification theory should not be viewed as merely epiphenomenal; there may be a critical feedback linkage between the justifications that people articulate and the changes to the systemic status quo they are willing to consider.

HOW RESOLUTELY SUPPORTIVE OF INEQUALITY MUST ONE BE TO QUALIFY AS A SYSTEM JUSTIFIER?

Unless system justification theorists adopt the orthodox positivist position that system justifiers are simply high scorers on the system justification scale—a position that hobbles cross-theory dialogue—we see a need to clarify the boundary conditions for distinguishing reflexive (mindless) system justifiers from political observers whose value systems and sense of fair play lead them to approve certain types of social-systemic arrangements—and condemn others. Here, we see value in turning to the hypothetical-society paradigm, because there are many ways to adapt this paradigm to probe how far system justifiers are prepared to go in defending inequality (and the types of dissonance-reduction strategies that they are prepared to use to trivialize awkward facts and to eliminate any need to change their minds). Here, we consider the possible reactions of high system justifiers to two categories of dissonant data: (a) those on intergenerational mobility, and (b) those on the effects of free trade on national security.

"Tormenting" Conservatives with Dissonant Data on Intergenerational Mobility

In the first generation of hypothetical-society research, we were content with crude operational definitions of meritocracy that manipulated the relative importance of hard work versus luck in determining income. But many observers find it difficult to view a society as meritocratic if one's status is determined by genetic lottery—and the children of the relatively poor have virtually no chance of rising into a higher class, whereas the children of the relatively wealthy are virtually guaranteed of remaining in that class (Rawls, 1971; Fishkin, 1983). It follows that social science research on intergenerational mobility has relatively high political stakes. As we saw in the earlier hypothetical-society studies, most people move in an egalitarian or leftward direction on income transfers when they are confronted with a low-meritocracy society.

This raises the question of how high scorers on system justification, or—as we suspect they are—conservatives (for the view that political conservatism

largely is system justification, see Jost, Glaser, Kruglanski, & Sulloway, 2003), respond to hypothetical societies in which meritocracy is not specified but must be inferred from data on intergenerational mobility. We conjecture that the first cognitive reaction of high system justifiers should be to assume that the observed patterns of inequality are legitimate (or justified), and that cognitively sophisticated system justifiers should be predisposed to defend the status quo by invoking the currently politically acceptable justifications for inequality—namely, the system follows the norms of meritocracy and equality of opportunity. The hypothetical-society paradigm allows us, however, to "stress test" this belief system by manipulating key background facts on intergenerational mobility that cut off favorite conservative dissonance reduction strategies. Promising manipulations include: (a) inequality is growing (the distance between the economic cellar and economic penthouse), thus cutting off the argument that things are getting better; (b) it is becoming increasingly difficult for people to rise from poverty to prosperity in one or even two generations, thus cutting off the Horatio-Alger-style anecdotes of rags-to-riches success; (c) there is no evidence that richer children have better prospects than poorer children because they have genetic endowments better suited to facilitate success in competitive market economies or because their parents do a better job bringing them up and inculcating character traits conducive to success (more intelligent, more optimistic, higher energy levels, etc.), thus cutting off arguments of either biological or cultural superiority; and (d) there is evidence that stereotypes and prejudice are key factors restraining upward mobility among the poor.

From our value-pluralism perspective, which holds that people rely on simple modes of dissonance reduction until they are forced by circumstances to embrace more complex modes, this series of factual constraints in the hypothetical society should drive conservatives to adopt more integratively complex (and centrist) policy positions. This is so because we have now narrowed the range of plausible explanations for social inequality in the hypothetical society to two salient candidates: better schools for the rich and better networking opportunities for the rich. We suspect that when the trade-offs are made this transparent, only the hardest-core conservatives and system justifiers will still resist egalitarian policy interventions designed to improve schooling opportunities and networking opportunities for the poor (e.g., generous vouchers and affirmative action outreach—although not de facto or de jure quotas—which activate a new set of value trade-offs). These hard-right dissenters might argue—in Burkean fashion—that previous generations of parents worked hard to ensure that their descendents would have advantages, so it is a bad idea to destabilize that societal expectation. But

we also suspect that most conservatives and system justifiers will, at this juncture, make policy concessions and accept the need for egalitarian interventions of some form.

An unresolved question is how system justification theorists should react to such a result. We obviously cannot speak for them but we favor the following accommodation: people tend to be system justifiers up to the point at which they feel they can no longer justify the system because it violates an internalized ethical schema of fair play. If there remains a difference between our position and that of system justification, it is our objection to labeling any ethical schema that happens to favor the status quo as merely serving a system justification function. Here we see a classic fuzzy-set functionalist judgment call (Tetlock, 2002), with tough questions for both camps. The tough question for us is: How far must perceptions and reality diverge before we grant that the perceptions serve a system justification rather than an object appraisal function? The tough question for them is: How grounded in reality must perceptions be before they grant that perceptions serve an object appraisal as opposed to a system justification function?

"Tormenting" System Justifiers with Dissonant Data on the Effects of Trade on National Sovereignty

In the first generation of hypothetical-society research, we brought the values of economic and market efficiency into conflict with the values of social equality, but we never brought market efficiency into conflict with another value also likely to rank high in the moral-political priorities of conservatives and, by implication, high system justifiers. National sovereignty and security are promising candidates.

Consider the problems posed by international trade. For orthodox, free market theorists, the logic of comparative advantage holds that the surest method of promoting prosperity is by permitting the free flow of goods, services, capital, and human beings across borders. If only rich countries would just quit erecting protectionist barriers that prevent poor people from working their way out of poverty, there would be much less poverty in the world today. Of course, this surgically simple solution can have painful side effects—international trade can produce major dislocations within societies. American blue-collar workers accustomed to earning $25 per hour run the risk of losing their jobs to Mexican workers glad to make $5 per hour—and these Mexican workers, in turn, risk losing their jobs to Chinese workers glad to make only $2 per hour.

We suspect that conservatives, and especially libertarian conservatives, are much less worried than those on the left about the power of trade to

increase inequality within their home society (see parallel section below on "tormenting" system critics). But there may well be conditions under which conservatives do become alarmed about the effects of international trade. Consider how the following combination of facts in a hypothetical-society paradigm would become increasingly dissonant for a conservative: (a) the target society has a mutually beneficial trading relationship with another society, but the other society is reaping much larger economic growth benefits from the trade; (b) the other society is a potential military rival that is translating significant fractions of its rapidly growing economy into greater military strength; and (c) the dominant social class in the target society has a strong vested interest in the continuation of the trading relationship with the other society (a disproportionate share of the benefits of the trade flow to this elite group within the target society) (see Herrmann, Tetlock, and Diascro, 2001).

Here, again, our suspicion would be that even high system justifiers will be hard-pressed to justify supporting the interest of the dominant class in a society so configured. There comes a point at which enough is enough: the status quo loses its legitimacy, and even those predisposed to justify the global free market status quo give up the cause. Again, although one may dismiss this stress testing of system justification theory on grounds that observers are judging a hypothetical status quo, not their own—the real—world, this approach at least promises evidence on the boundary conditions of SJT: Are system justification tendencies so automatic, and unconscious rationalization tendencies so strong, that system justification continues even when the obvious routes to rationalizing the legitimacy of the status quo have explicitly been cut off and the system in question is nominally hypothetical, or can these tendencies be overridden by cutting off normal rationalization routes at the conscious level and, if so, how easily may people be divorced from their system justifying ideologies (or, in the case of the disadvantaged, freed from the fog of false consciousness) (Jost, 1995)?

How Resolutely Opposed to Inequality Should One Be to Qualify as a System Critic?

Fair play requires subjecting those on the left to the functional equivalent of the dissonance-maximizing treatments inflicted on those on the right: How far are left-leaning respondents prepared to go in opposing all forms of inequality? And what types of dissonance-reduction strategies are they prepared to adopt to deflect bothersome facts that pressure them to change their minds? We focus on two examples: (a) reactions to increasingly dissonant data on the sources of social inequality within the home society, and (b) reactions to increasingly dissonant data on the impact of protectionist barriers

designed to protect workers in one's own society but at the price of inflicting great suffering on much poorer workers in other societies.

"Tormenting" System Critics with Dissonant Data on Social Inequality

In the first generation of hypothetical-society work, we explored the willingness of those on the left to reject increasingly meritocratic hypothetical societies by manipulating the importance of effort/ability as causes of socio-economic status. But we never subjected the left to tougher ideological tests that probed just how far they were willing to go in pursuit of equality as an end goal that trumps all other competing ends. Imagine, therefore, a hypothetical society in which we preempt arguments for a wide range of egalitarian policy interventions by stipulating that: (a) the society already rigorously enforces equality-of-opportunity laws, thus undercutting the dissonance-reduction strategy that inequality could be eliminated if only more aggressive action were taken against ongoing discrimination; (b) the society has no history of ethnic or racial prejudice, thus undercutting the strategy of arguing that inequality could be eliminated if only aggressive action were taken against the residual effects of past injustices; (c) the inequalities create powerful incentives for efficiency and economic growth from which all benefit, thus undercutting the strategy of arguing that inequality could be eliminated (without making everyone poorer) if taxation policy reallocated wealth; (d) the relatively poor are, by current objective standards of purchasing power, already very well-off, further undercutting need-based humanitarian arguments for equality; (e) the poor are satisfied with the fairness of the system or even that the poor are more satisfied with the conditions of their lives than the wealthy and are making work–leisure trade-offs in favor of leisure and less income (in other words, the poor realize that, beyond a certain point, which they feel they have reached, higher income does not buy greater happiness; Kahneman, Krueger, & Schkade, 2006); (f) scientific evidence has revealed that children from wealthier families have genetic endowments that are, on average, better adapted for success in competitive market economies and that, whenever lower-class children have the "right stuff," they do indeed rise into higher socio-economic classes (thus reaffirming that equality of opportunity does exist); (g) scientific research indicates that, short of mandating poverty for all, there are only two remaining mechanisms for breaking down social class barriers—nature or nurture—either genetic engineering designed to level the DNA playing field or socializing the task of socializing children and requiring that all children be raised in state-run institutions that prevent higher-class parents from giving special advantages to their

children (from elaborate bedtime stories to excessive homework help) and lower-class parents from teaching their children impulsive and hedonistic values detrimental to success.

Choreographing the background facts to maximize dissonance for egalitarians is obviously a complex, iterative process, best done in adversarial collaboration with rival theoretical camps. Here, though, we are most interested in the choices that egalitarians make when the only economically and technologically feasible method of achieving egalitarian goals requires acknowledging the tension between the values of social equality and family autonomy.

Radical egalitarians—from Rousseau to Marx—have long recognized this tension: as long as the family is the social unit primarily responsible for socializing children, and as long as some families are (even holding income constant) prepared to make much greater sacrifices to ensure the success of their children, it is logically impossible to achieve equality of opportunity. Socializing children is a relatively easy choice from this radically egalitarian point of view—and many socialist governments have indeed pursued this "it-takes-a-village" option (from Israeli kibbutzim to Scandinavian day care to Chinese communes). Conservative and libertarian philosophers have long resisted such arguments and warned that transferring the task of socializing children to the state is both a violation of parental rights and a dangerous step toward totalitarianism and collective mind control. Rejecting a prominent state role in childcare is a relatively easy choice from these points of view.

Our working hypothesis is, however, that, for most people, the choice is a tough one. We suspect that most people—system critics and system justifiers alike—are value pluralists who are deeply torn by this value conflict and oscillate erratically between favoring family autonomy versus equality of opportunity as a function of horror stories of child neglect and abuse (favoring the left) and horror stories of state mind control and parents losing parental rights for "trivial" reasons (favoring the right). Extrapolating from earlier work on the value pluralism model (Tetlock, 1986; Tetlock et al., 1996), we also suspect that people (especially egalitarians now) can be motivated to invest the necessary cognitive effort to generate complex compromise solutions to the dilemma only to the degree that we have systematically blocked off simple modes of dissonance resolution in the hypothetical societies. These tempting simpler modes of dissonance reduction include challenging the "fact situation" posited in the hypothetical society (such as "the poor aren't really as happy as the rich; that is just false consciousness" and "behavioral-genetics claims are just racist") and trying to find a trade-off-free solution (creating a state-funded system in which social class distinctions disappear

because everyone develops to her full potential). The key question is: What value trade-offs do egalitarians make when constrained by the factual and causal ground rules of the hypothetical society—and when they cannot make up facts of their own liking? The value pluralism model predicts that the more highly respondents value both equality and the family, the more excruciatingly complex the judgment calls will become of balancing parental control and social equality in designing exact institutional rules. If integratively complex policy reasoning is a reasonable approximation of one's ideal cognitive process outcome (and that seems to be the case for advocates of deliberative democracy; e.g., Fishkin, 1992), this would be how to achieve it via the hypothetical-society paradigm.

The process may seem torturous because the goal is to explore the conditions under which even unrelenting system critics relent. Or, framed as a question for system justification theorists, how dogmatic (principled) an opponent of inequality must system critics be to avoid reclassification as system justifiers? For instance, and we doubt that system justification theorists take this extreme a position, if the price of avoiding the label "system justifier" is compelling all families to accept a one-size-fits-all child-rearing system that guarantees equality of outcome, we suspect that 90% plus of the population will qualify as system justifiers. Simply put, would a system justification theorist consider adherence to the existing American family structure, which vests considerable autonomy and responsibility for child development in the parents and which surely breeds societal inequality, evidence of the system justification motive at work? If not, why not? In any event, if system justification is to be more than a vague expression of political disapproval, as system justification theorists surely mean it to be, we need much tighter specification of the value and policy litmus tests being used—implicitly or explicitly—by system justification theorists.

"Tormenting" System Critics with International Trade Scenarios

In the first generation of hypothetical-society research, we were content to rely on crude operational definitions of the poverty line, assuming that everyone shared an understanding of, and aversion to, poverty. What counts as poor, however, in one society at one point in history may count as wealthy for that same society at a previous point in history or for other societies at the same point in history. Upper middle class professionals in parts of sub-Saharan Africa in the early 21st century have per capita incomes substantially lower (even using a purchasing-power-parity standard) than the average factory worker in Western Europe or the United States.

In the hypothetical-society paradigm, we can require subjects to assume—as noted earlier—that the logic of comparative advantage in international trade

holds: the surest method of reducing large income gaps across societies is by promoting the free flow of goods, services, capital, and human beings across borders. How, then, should one respond if one is an egalitarian asked to judge the acceptability of a trade agreement that will increase inequality within one's own wealthy society (because the paychecks of one's "own" working class are in decline as the result of lower labor cost competition in poorer societies) but will also raise the absolute standard of living of the poorest people in poor societies, as well as decrease inequality between societies (by raising the overall per capita income of poorer societies closer to that of wealthier societies)? The predictions we can extract from SJT presumably hinge on whether we choose to define the system critics as cosmopolitan egalitarians, concerned more with inequality on a global scale, or as parochial egalitarians, concerned solely with inequality within their own society. And the data we can extract from the study will probably hinge on the escape routes that we offer respondents in hypothetical societies from this dissonance-inducing problem (escape routes such as reserving some wealth generated by free trade for transfer payments to help those in one's own society most adversely affected by free trade, the solution preferred by value-pluralistic neo-liberals such as Robert Rubin [Rubin & Weisberg, 2004] and Thomas Friedman [2005]).

Again, the "system" in system justification theory is underdefined.[10] The theory offers little guidance on how to apply it to complex debates that activate clashing values—and on which reasonable people disagree. We see roughly equally strong arguments for classifying "egalitarian" protectionists in wealthy countries as either system justifiers or system critics—and no good reason to suppose that psychologists deserve any special deference in the answers they might give as to which systems should count, except to the extent that their answers are founded on empirical data. If conservatives become system critics and liberals become system justifiers in the "America becoming more open to international trade" scenario, and if other similar reversals can be identified, then it becomes difficult to argue that the perpetuation of economic inequality or the defense of the status quo per se generally triggers system-justifying tendencies in those deemed high system justifiers in the United States, namely, conservatives (Jost, Blount, Pfeffer, & Hunyady, 2003; Jost, Glaser, Kruglanski, & Sulloway, 2003). In such a case, we see the benefit of the hypothetical-society approach as pushing toward a more con-

[10] Blasi and Jost (2006) recognize this problem and note the need for studies to determine when one system will prevail over another in cases of system conflict, but to our knowledge, little or no research addresses this question.

textualized theory about the conditions under which system-justifying, and system critical, tendencies should occur.

CONCLUSION

The hypothetical-society paradigm may well be the best of the many imperfect methodological means at our disposal for testing the relative merits of more transparent-motivational and more subterranean-motivational theories of public policy reasoning. Here, it is instructive to recall just how deep the indeterminacy problems are in testing a theory such as system justification in the real world. We repeatedly run into variations on C. Wright Mills' vocabulary-of-motives problem: one person's reason for holding a belief (say, about social class differences in achievement values or about the wisdom of the market) can typically be dismissed by others as a mere rationalization (say, as a means of justifying existing inequality or as evidence of insensitivity to the residual effects of past and current discrimination). Rubin and Hewstone (2004) make a somewhat analogous point when they argue that system justification theory should not get explanatory credit for phenomena, such as attributional favoritism toward higher-status groups, that could simply be the result of people observing depressing patterns of covariation between group membership and outcomes in society at large (e.g., the higher levels of crime, family breakdown, drug abuse, school failure, and so on among the poor). To use their analogy to a football game, should we conclude that members of the losing team who attribute their defeat to their own shortcomings are, ipso facto, guilty of outgroup favoritism and system justification? Or, should we conclude that they are engaging in highly adaptive forms of self-criticism? Indeed, it is worth asking what happens to disadvantaged groups that develop political cultures that censure all self-critical commentary as evidence that the commentator has been co-opted by the oppressors. Do they not risk trapping themselves in an ideology of victimology?

The list of Millsian reason-rationalization riddles is a long one. For instance, if one believes that prosperity and economic efficiency require creating incentives for hard work and risk-taking (incentives that inevitably create inequality), does that belief count as evidence for the operation of a system justification motive (one's belief that the wealthy are being rewarded for merit) or as evidence simply that one understands a fundamental scientific principle of economics (for the former view, see Jost et al., 2003)? If one believes that a social system with stable, secure property rights is essential for promoting prosperity and economic efficiency, does that count as evidence of a desire for unequal relations among social groups, or does it count as evidence that one

has drawn correct lessons from history—at least according to one influential school of economic history (North, 1981, 2005)? If one believes that ego resilience, intelligence, the capacity to delay gratification, and a strong work ethic are found more often among the economically successful, does that count as evidence that one has been gulled into accepting system-justifying Horatio Alger stories (cf. Wakslak, Jost, Tyler, & Chen, 2007), as evidence that one is in touch with sociological reality (as Herrnstein & Murray, 1994, suggest), or as evidence that one has embraced an adaptive illusion (Taylor & Brown, 1988)?

These questions are unanswerable in real-world debates because it is so easy for advocates—motivated reasoners that we all are to some degree (Kunda, 1999)—to invent facts and double standards that conceal potential trade-offs (an invention process that, if it is to serve its subterranean-motivational function, should occur out of awareness and be invisible to others). But these questions become answerable in the hypothetical-society paradigm because it is so difficult for advocates to conceal the same trade-offs in a world in which all of the key factual parameters have been specified by experimental fiat. The hypothetical-society paradigm then becomes the platform for previously impossible conversations between theorists. For instance, even if we are right and if transparent-motivational theories can outmaneuver subterranean-motivational theories in carefully choreographed hypothetical societies that compel conscious acknowledgment of complex value trade-offs, subterranean-motivational theorists still have a number of reasonable counterarguments. They can posit that socially undesirable motivational forces only come into play when enough attributional ambiguity exists to permit rationalization covers—or that such motives only come into play in settings that better simulate real-world status relationships. We do not dismiss such arguments as patch-up operations of a degenerating research program. Such defenses may well be defensible, and the best way to tell is by gradually adding the requisite complexity and realism to hypothetical-society studies.

In brief, if we want to escape otherwise intractable disputes over political motive attribution, we need to explore human judgment in imaginary social worlds that we can experimentally manipulate in precisely targeted ways that reflect the key conceptual parameters of real-world political debates.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper and Row.

Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?," *Psychological Inquiry*, *15*, 257–278.

Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased judgments of fairness in bargaining. *American Economic Review, 85,* 1337–1343.

Bar-Hillel, M., & Yaari, M. (1993). Judgments of distributive justice. In B. A. Mellers & J. Baron (Eds.), *Psychological perspectives on justice.* Cambridge: Cambridge University Press.

Berlin, I. (1990). *The crooked timber of humanity.* London: John Murray.

Blasi, G., & Jost, J. T. (2006). System justification theory and research: Implications for law, legal advocacy, and social justice. *California Law Review, 94,* 1119–1168.

Boulding, K. E. (1962). Social justice in social dynamics. In R. B. Brandt (Ed.), *Social justice.* Englewood Cliffs, NJ: Prentice Hall.

Brickman, P. (1977). Preference for inequality. *Social Psychology Quarterly*, *40*, 303–310.

Bukszar, E., & Knetsch, J. L. (1997). Fragile redistribution choices behind a veil of ignorance. *Journal of Risk and Uncertainty, 14,* 63–74.

Bunge, M. (1961). The weight of simplicity in the construction and assaying of scientific theories. *Philosophy of Science, 28,* 120–149.

Coleman, S. (2000). Thought experiments and personal identity. *Philosophical Studies*, *98,* 53–69.

Cooper, R. (2005). Thought experiments. *Metaphilosophy, 36,* 328–347.

Daniels, N. (1996). *Justice and justification.* Cambridge: Cambridge University Press.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review, 7,* 324–336.

Dennett, D. C. (2003). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies, 10,* 19–30.

Elster, J. (1995). The empirical study of justice. In D. Miller & M. Walzer (Eds.), *Pluralism, justice, and equality.* Oxford: Oxford University Press.

Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research, 17,* 171–187.

Ferrie, J. P. (2005). The end of American exceptionalism? Mobility in the United States since 1850. *Journal of Economic Perspectives, 19,* 199–215.

Fishkin, J. S. (1983). *Justice, equal opportunity, and the family.* New Haven, CT: Yale University Press.

Fishkin, J. S. (1992). *The dialogue of justice.* New Haven, CT: Yale University Press.

Fong, C. M. (2004). *Which beliefs matter for redistributive politics? Target-specific versus general beliefs about the causes of income.* Unpublished manuscript.

Fong, C. M. (2005). *Prospective mobility, fairness, and the demand for redistribution.* Unpublished manuscript.

Fong, C. M., Bowles, S., & Gintis, H. (2006). Strong reciprocity and the welfare state. In J. Mercier-Ythier, S. Kolm, & L. A. Gerard-Varet, *Handbook on the economics of giving, reciprocity and altruism* (Vol. 2, pp. 1439–1464). Amsterdam: Elsevier.

Forde-Mazrui, K. (2004). Taking conservatives seriously: A moral justification for affirmative action and reparations. *California Law Review, 92,* 683–754.

Frankfurt, H. (1987). Equality as a moral ideal. *Ethics, 98,* 21–43.

Friedman, T. H. (2005). *The world is flat: A brief history of the twenty-first century.* New York: Farrar, Straus and Giroux.

Frohlich, N., & Oppenheimer, J. A. (1992). *Choosing justice: An experimental approach to ethical theory.* Berkeley, CA: University of California Press.

Frohlich, N., & Oppenheimer, J. A. (1997). A role for structured observations in ethics. *Social Justice Research, 10,* 1–21.

Frohlich, N., & Oppenheimer, J. A. (2000). How people reason about ethics. In A. Lupia, M. D. McCubbins, & S. L. Popkin, *Elements of reason.* Cambridge: Cambridge University Press.

Greenberg, J. (1983). Overcoming egocentric bias in perceived fairness through self-awareness. *Social Psychology Quarterly, 46,* 152–156.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20,* 98–116.

Haines, E. L., & Jost, J. T. (2000). Placating the powerless: Effects of legitimate and illegitimate explanation on affect, memory, and stereotyping. *Social Justice Research, 13,* 219–236.

Herrmann, R., Tetlock, P. E., & Diascro, M. (2001). How Americans think about trade: Resolving conflicts among money, power, and principles. *International Studies Quarterly, 45,* 191–218.

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Jost, J. T. (1995). Negative illusions: Conceptual clarification and psychological evidence concerning false consciousness. *Political Psychology, 16,* 397–424.

Jost, J. T. (1997). An experimental replication of the depressed entitlement effect among women. *Psychology of Women Quarterly, 21,* 387–393.

Jost, J. T. (2001). Outgroup favoritism and the theory of system justification: An experimental paradigm for investigating the effects of socio-economic success on stereotype content. In G. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 89–102). Mahwah, NJ: Erlbaum.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33,* 1–27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25,* 881–919.

Jost, J. T., Blount, S., Pfeffer, J., & Hunyady, G. (2003). Fair market ideology: Its cognitive-motivational underpinnings. *Research in Organizational Behavior, 25,* 53–91.

Jost, J. T., & Burgess, D. (2000). Attitudinal ambivalence and the conflict between group and system justification motives in low status groups. *Personality and Social Psychology Bulletin, 26,* 293–305.

Jost, J. T., Fitzsimons, G., & Kay, A. C. (2004). The ideological animal: A system justification view. In J. Greenberg, S. L. Koole, & T. Pyszczynski (Eds.), *Handbook of experimental existential psychology* (pp. 263–282). New York: Guilford Press.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129,* 339–375.

Jost, J. T., & Hunyady, O. (2002). The psychology of system justification and the palliative function of ideology. *European Review of Social Psychology, 13,* 111–153.

Jost, J. T., Pelham, B. W., & Carvallo, M. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology, 38,* 586–602.

Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology, 36,* 209–232.

Kahneman, D., Knetsch, J., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review, 76,* 728–741.

Kahneman, D., Krueger, A. B., & Schkade, D. (2006). Would you be happier if you were richer? A focusing illusion. *Science, 312,* 1908–1910.

Kay, A., Jimenez, M. C., & Jost, J. T. (2002). Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo. *Personality and Social Psychology Bulletin, 28,* 1300–1312.

Kluegel, J. R., Csepeli, G., Kolosi, T., Orkeny, A., & Nemenyi, M. (1995). Accounting for the rich and the poor: Existential justice in comparative perspective. In J. R. Kluegel, D. S. Mason, & B. Wegener (Eds.), *Social justice and political change: Public opinion in capitalist and post-communist states.* New York: Aldine de Gruyter.

Knobe, J. (in press). What is experimental philosophy? *The Philosophers' Magazine.*

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature, 41,* 1188–1239.

Konow, J. (2005). Blind spots: The effects of information and stakes on fairness bias and dispersion. *Social Justice Research, 18,* 349–390.

Kunda, Z. (1999). *Social cognition.* Cambridge, MA: MIT Press.

Laswell, H. D. (1930). *Psychopathology and politics.* Chicago, IL: University of Chicago Press.

Liebig, S. (2001). Lessons from philosophy? Interdisciplinary justice research and two classes of justice judgments. *Social Justice Research, 14,* 265–287.

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition, 92,* B1–B12.

McDermott, R. (1998). *Risk-taking in international politics: Prospect theory in American foreign policy.* Ann Arbor: The University of Michigan Press.

Messick, D. M., & Sentis, K. P. (1983). Fairness, preference, and fairness biases. In D. M. Messick & K. S. Cook (Eds.), *Equity theory.* New York: Praeger.

Michelbach, P. A., Scott, J. T., Matland, R. E., & Bornstein, B. H. (2003). Doing Rawls justice: An experimental study of income distribution norms. *American Journal of Political Science, 47,* 523–539.

Miller, D. (1999). *Principles of social justice.* Cambridge, MA: Harvard University Press.

Mills, C. W. (1940). Situated action and vocabularies of motives. *American Sociological Review, 5,* 904–913.

Mitchell, G., & Tetlock, P. E. (2006). An empirical inquiry into the relation of corrective justice to distributive justice. *Journal of Empirical Legal Studies, 3.*

Mitchell, G., Tetlock, P. E., Mellers, B. A., & Ordóñez, L. D. (1993). Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology, 65,* 629–639.

Mitchell, G., Tetlock, P. E., Newman, D. G., & Lerner, J. S. (2003). Experiments behind the veil: Structural influences on judgments of social justice. *Political Psychology, 24,* 519–547.

North, D. C. (1981). *Structure and change in economic history.* New York: Norton.

North, D. C. (2005). *Understanding the process of economic change.* Princeton, NJ: Princeton University Press.

Nozick, R. (1974). *Anarchy, state, and utopia.* New York: Basic Books.

O'Brien, L. T., & Major, B. (2005). System-justifying beliefs and psychological well-being: The roles of group status and identity. *Personality and Social Psychology Bulletin, 31,* 1718–1729.

Ordónez, L., & Mellers, B. A. (1993). Tradeoffs in fairness and preference judgments. In B.A. Mellers & J. Baron, J. (Eds.), *Psychological perspectives on justice: Theory and applications* (pp. 138–154). New York: Cambridge University Press.

Pillutla, M. M., & Murnighan, J. K. (2003). Fairness in bargaining. *Social Justice Research, 16,* 241–262.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Belknap.

Redding, R. E. (2004). Bias on prejudice? The politics of research on racial prejudice. *Psychological Inquiry, 15,* 289–293.

Rubin, M., & Hewstone, M. (2004). Social identity, system justification, and social dominance: Commentary on Reicher, Jost et al., and Sidanius et al. *Political Psychology, 25,* 823–844.

Rubin, R. E., & Weisberg, J. (2004). *In an uncertain world: Tough choices from Wall Street to Washington.* New York: Random House.

Scott, J. T., Matland, R. E., Michelbach, P. A., & Bornstein, B. H. (2001). Just deserts: An experimental study of distributive justice norms. *American Journal of Political Science, 45,* 749–767.

Sidanius, J., Levin, S., Federico, C., & Pratto, F. (2001). Legitimizing ideologies: The social dominance approach. In J. Jost and B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (pp. 307–331). Cambridge: Cambridge University Press.

Skitka, L. J. (1999). Ideological and attributional boundaries on public compassion: Reactions to individuals and communities affected by a natural disaster. *Personality and Social Psychology Bulletin, 25,* 792–808.

Skitka, L. J., & Tetlock, P. E. (1992). Allocating scarce resources: A contingency model of distributive justice. *Journal of Experimental Social Psychology, 28,* 491–522.

Skitka, L. J., & Tetlock, P. E. (1993). Of ants and grasshoppers: The political psychology of allocating public assistance. In B. A. Mellers & J. Baron (Eds.), *Psychological perspectives on justice.* Cambridge: Cambridge University Press.

Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues, 42,* 129–150.

Souder, L. (2003). What are we to think about thought experiments? *Argumentation, 17,* 203–217.

Suedfeld, P., & Tetlock, P. E. (Eds.) (1991). *Psychology and social policy.* Washington, DC: Hemisphere.

Taylor, S. E., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.

Tetlock, P. E. (1984). Content and structure in political belief systems. In S. Chan & D. Sylvan (Eds.), *Foreign policy decision-making: Perception, cognition, and artificial intelligence.* Boulder, CO: Westview Press.

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

Tetlock, P. E. (2002). Social-functionalist frameworks for judgment and choice: The intuitive politician, theologian, and prosecutor. *Psychological Review, 109,* 451–472.

Tetlock, P. E., Armor, D., & Peterson, R. (1994). The slavery debate in antebellum America: Cognitive style, value conflict, and the limits of compromise. *Journal of Personality and Social Psychology, 66,* 115–126.

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., & McGuire, C. (1986). Cognitive perspectives on foreign policy. In R. White (Ed.), *Psychology and the prevention of nuclear war.* New York: Free Press.

Tetlock, P. E., Peterson, R., & Lerner, J. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *Ontario symposium on social and personality psychology: Values* (pp. 25–52). Hillsdale, NJ: Erlbaum.

Tetlock, P. E., & Tyler, A. (1996). Winston Churchill's cognitive and rhetorical style: The debates over Nazi intentions and self-government for India. *Political Psychology, 17,* 149–170.

Tetlock, P. E., Visser, P., Singh, R., Polifroni, M., Elson, B., Mazzocco, P., & Rescober, P. (2007). People as intuitive prosecutors: The impact of social control motives on attributions of responsibility. *Journal of Experimental Social Psychology.*

Wakslak, C., Jost, J. T., Tyler, T. R., & Chen, E. (2007). Moral outrage mediates the dampening effect of system justification on support for redistributive social policies. *Psychological Science, 18,* 267–274.

Weinrib, E. J. (2002) Corrective justice in a nutshell. *University of Toronto Law Journal, 52,* 349–356.