Experimental Political Philosophy: Justice Judgments in the Hypothetical Society Paradigm

Gregory Mitchell* Philip E. Tetlock†

*University of Virginia School of Law, pgm6u@virginia.edu
†UCBerkeley - Haas School of Business, tetlock@haas.berkeley.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://law.bepress.com/uvalwps/uva_publiclaw/art52

Copyright ©2006 by the authors.
Experimental Political Philosophy: Justice
Judgments in the Hypothetical Society
Paradigm

Gregory Mitchell and Philip E. Tetlock

Abstract

In this draft of a chapter forthcoming in a book on political psychology, we advocate blending thought experiments with laboratory experiments via a technique we call “the hypothetical society paradigm,” which is designed to bring out the inferential advantages of both approaches while minimizing their disadvantages. We discuss the primary benefits of this technique and survey the principal empirical findings thus far obtained using this technique. We also discuss two categories of fruitful future applications of this and related techniques: (a) isolating sources of support and resistance to particular policy proposals with potentially profound societal implications; (b) helping to clarify boundary conditions for the applicability of competing and complementary psychological theories of justice.
Experimental Political Philosophy: Justice Judgments in the Hypothetical Society Paradigm

Gregory Mitchell* & Philip E. Tetlock**

Political philosophers often employ thought experiments in developing their normative accounts of justice. Most famously within modern political philosophy, Rawls (1971) develops his theory of justice by reflecting on the principles of justice to which people would consent if they made their choices behind a “veil of ignorance” that blinded them to their likely standing in a future society. Almost as famously, Nozick (1974) advances his competing, libertarian conception of justice with a hypothetical about Wilt Chamberlain’s entitlement to the fruits of his labor obtained through free exchange. Ackerman (1980) imagines a society in which a “perfect technology of justice” removes all practical impediments to achieving social justice—all for the express purpose of providing “a clean-cut thought experiment to test the claim that liberalism is bankrupt (p. 21).” Several other examples of prominent thought experiments in modern political philosophy could be given, for “[r]ecent liberal theory has been distinctive for its thought experiments” (Fishkin, 1992, p. 50).

As empiricists, social scientists tend to be skeptical of thought experiments (suspecting great potential for experimenter bias when $N = 1$—and the experimenter is the sole subject) and, as positivists, social scientists tend to be skeptical of the normative theories of justice that percolate out of thought experiments (suspecting that clashing “ought” claims boil down to pre-existing ideological preferences of the thought experimenters). Social scientists prefer to employ

* Professor of Law and E. James Kelly, Jr.-Class of 1965 Research Professor, University of Virginia School of Law, greg_mitchell@virginia.edu.
** Lorraine Tyson Mitchell Chair of Organizational Behavior, Haas School of Business, University of California, Berkeley, tetlock@haas.berkeley.edu.
their own distinctive research tools: laboratory experiments, surveys, and historical and econometric studies to develop descriptive and explanatory accounts of justice. By examining how experimental participants divide rewards among members of a small group engaged in a collective task, how survey participants respond to questions about the distribution of wealth, or how different groups react to different welfare policies or institutions distribute scarce resources, social scientists try to shift the spotlight from the unanswerable question “what is just?” to the answerable question “who finds what to be just—and why?”

Mutual suspicion is likely to persist for a long time. It is hard to persuade philosophers that empirical studies can produce findings about justice per se, due to self-interest contaminants and insufficient conceptual care in formulating public opinion questions and answers (Liebig, 2001; Swift, 1999), and it is perhaps even harder to persuade social scientists that thought experiments can produce theories of justice uncontaminated by pet theories (Barber, 1975) or cultural biases (Okin, 1989).

In this chapter, however, we argue for a partial rapprochement: we advocate blending thought experiments with laboratory experiments via a technique we call “the hypothetical society paradigm” designed to bring out the inferential advantages of both approaches while minimizing the disadvantages. We divide the chapter into three sections. In the first section, we discuss the primary benefits of this technique; in the second section, we survey the principal empirical findings thus far obtained using this technique; in the third section, we discuss two categories of fruitful future applications of this and related techniques: (a) isolating sources of support and resistance to particular policy proposals with potentially profound societal

---

1 Some political philosophers and political theorists do rely on empirical research to develop their theories of justice. For instance, Miller (1999) relies heavily on empirical studies of justice, and Elster (1992) relies heavily on actual practices of institutions.
implications; (b) helping to clarify boundary conditions for the applicability of competing and complementary psychological theories of justice.

I. Turning Thought Experiments into Lab Experiments

Thought experiments help philosophers isolate and clarify the role of different principles and assumptions in their normative arguments, to see what is essential and what is not to a theory, much like laboratory experiments help to identify 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 conducted by philosophers who faithfully report their mental simulations and conclusions, lack the transparency and replicability deemed essential to scientific research (Bunge, 1961), and suspicions about the accuracy and reliability of self-reports on introspection lead many to dismiss the thought experiment as a path to reliable knowledge (see Sorenson, 1992, chapter 2). Thus, when scientists employ thought experiments, as Galileo, Newton and Einstein did with some success, the resulting theories must ultimately be couched in publicly testable terms to qualify as scientific (Dennett, 2003).

Thought experiments also suffer from an extreme version of the external validity complaint often lodged against laboratory studies. 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., the philosophical study of personal identity has been particularly plagued by competing hypotheticals and competing conclusions; see 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 et al., 2004). The emerging field of experimental philosophy seeks to do just this with a variety of philosophical conundrums (see Knobe, in press). But that view emphasizes what laboratory studies can do for thought experiments. In our view, thought experiments can do much for laboratory studies of justice.

Empirical studies of justice pose their own difficulties, some of which may be alleviated by incorporating elements of thought experiments into these studies. First, and virtually impossible to control within empirical studies, is the problem that public opinion on matters of justice 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 (“workfare”) will encourage self-
sufficiency or whether intergenerational mobility is rising or declining). When causal relations and policy effects are difficult to determine, there is a powerful temptation to bring one's beliefs about the facts into convenient alignments that minimize dissonance and mental strain (e.g., Herrmann et al., 2001; Mitchell et al., 1993; Skitka, 1999). For instance, Skitka and Tetlock (1992, 1993) and Christiansen and Lavine (1997) found that liberals and conservatives held different pre-existing 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 important underlying facts, such as the degree of economic mobility present within a society (see Ferrie, 2005; Fong, 2005) or the proportion of families within different socioeconomic categories (see Kluegel et al., 1995). These problems are particularly acute when one tries to study the impact of macroeconomic variables and system-level conditions on individual judgments of justice, but informational problems can arise in any setting where important facts are vague or disputed (as, for example, in the bargaining studies conducted by Babcock et al., 1995).

In order to overcome factual disagreements and avoid factual 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 distributions of wealth within hypothetical societies (see Mitchell et al., 1993). This paradigm capitalizes on the unreality of thought experiments: because the experimenter designs the hypothetical societies, the experimenter controls the structure of these hypothetical societies, 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 available, 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, one is able to examine
which features of societies are most important to people’s judgments of social justice and able to
examine how these judgments change as features of the societies change. In short, the
hypothetical society paradigm allows researchers to unconfound the role of factual beliefs from
the influence of value orientations in judgments of justice. Because individuals tend to avoid
value trade-offs, often by arguing facts (see, e.g., Tetlock & McGuire, 1986), this ability to
manipulate value conflict provides considerable experimental advantages.

Thus, one main benefit of importing hypothetical societies into the laboratory is the
control one gains over otherwise complex and debatable matters of fact. A second benefit
involves the control one gains over the influence of selfish interests. A problem commonly
confronted in empirical studies of justice is that of disentangling biased from unbiased judgments
(see Fong et al., in press; 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

---

2 We speak here in terms of laboratory studies, but with innovations in web-based surveys, the hypothetical society
paradigm could also be employed in large-scale experimental surveys using the Internet (see Škitka & Sargis, 2006),
or in more traditional questionnaire surveys (e.g., Jasso and Opp, 1997) conducted a factorial questionnaire survey
with hypothetical vignettes). In simplified form, hypothetical society manipulations may also be candidates for
inclusion in telephone surveys.
can ask participants to explain or justify their judgments under various accountability conditions or can manipulate the group identities involved, while 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 studies using the hypothetical society paradigm, we have favored experimental manipulations that place the participant in the position of impartial spectator because arguably judgments from that position best represent 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 justice ideally requires, “thought experiments trump real experiments (Cooper, 2005, p. 344)” While the experience of injustice from the first-person perspective is an important subject of inquiry in its own right (Lerner, 2003; Miller, 2001), individualized appeals to justice always invoke moral norms or principles of general application that the individual claims are or should be endorsed by the polity or community. The

---

3 In fact, 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 the self-knowledge required by a true veil of ignorance, we chose instead to remove narrow self-interests as an influence on judgments by having participants disinterestedly evaluate hypothetical societies.

4 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).”
notion of justice principles idiosyncratic to the individual, as opposed to there being individual differences in phenomenological experiences of injustice on in subjective claims about what social justice requires in a particular case, is oxymoronic from a philosophical perspective. For the philosopher, a judgment about a distribution or procedure qualifies as a justice judgment only if the judgment is made from an impartial point of view, such as by third-party observers to hypothetical societies, and reflects the ethical preferences of those impartial spectators (Liebig, 2001; Lerner, 1974, likewise argues that non-self-involved persons are more likely to base their judgments on moral values than self-involved persons).

Judgments about justice made by detached observers of hypothetical societies may nevertheless be useful guides about justice in real societies. Most obviously, to the extent hypothetical societies and real societies possess common features important to participants’ judgments of justice, judgments about justice in the hypothetical societies may generalize to real societies. But even with highly artificial scenarios, judgments about hypothetical societies can identify areas of agreement and disagreement on topics of economic policy and social justice and explain how factual beliefs and value differences combine to produce either consensus or dissension. For instance, in our first hypothetical society studies (Mitchell et al., 1993), we found a surprising degree of agreement across political groups on the importance of providing a minimum safety net, even in perfect meritocracies in which the poor controlled their economic fates. And perhaps most 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, 1993). If used in this sense, the hypothetical society paradigm may perform a “reflective equilibrium” function (Rawls, 1971; see Daniels, 1996), possibly leading persons to

http://law.bepress.com/uvalwps/uva_publiclaw/art52
give up their initial intuitions or change their views about what justice requires once they are confronted with a series of controlled thought experiments.

Although our focus has been on how thought experiments can aid empirical studies of justice, it is worth pausing to note that the hypothetical society paradigm may also aid philosophical studies of justice. Elster (1995) notes that one of the problems for Rawls’ (1971) theory of justice is deciding what arrangements agreed to in the original position are likely to have the greatest “psychological stability.” For instance, when will people be more concerned about wastefulness or unfairness in the original position? The hypothetical society paradigm is well-positioned to answer such robustness questions, because different hypothetical policies varying in efficiency and fairness can be systematically compared and evaluated in this paradigm. In fact, we conducted a comparison of this sort in one of our studies, and we found that conservatives were more sensitive to waste in redistributive policies than liberals when the redistribution was meant for deserving recipients (Mitchell et al., 2003), suggesting that the psychological stability of arrangements depends on the mix of liberals and conservatives among respondents and the mix of deserving and undeserving recipients in the applicant pool. Further, philosophers may use empirical data on fairness intuitions about hypothetical societies “as a heuristic in selecting which intuitions to scrutinize most intensely (Elster, 1995, p. 94),” or as a disturbance to the philosopher’s reflective equilibrium that can lead to a wider reflective equilibrium. For instance, if the great majority of considered judgments of well-educated participants in a controlled thought experiment indicate that corrective justice operates independently of distributive justice under most conditions (Mitchell & Tetlock, in press), is this not a finding that philosophers pressing the view that corrective justice depends on distributive justice should take into account?
Our hypothetical society paradigm is not the first attempt to wed political philosophy and experimental design. Most notably, Brickman (1977) and Frohlich and Oppenheimer (1992) approximated Rawls’ original position in small-group experiments, with Brickman finding support for Rawls’ maximin principle and Frohlich and Oppenheimer finding support for a modified utilitarianism in which utility was maximized above a floor constraint that the worst-off receive a guaranteed minimum income. But we do believe that the hypothetical society paradigm weds political philosophy and experimentation in a unique way that makes the experimental study of interactions between macro- and micro-level variables more manageable and opens to direct testing a range of questions that would otherwise be subject only to indirect testing. In the next section we describe some of the discoveries made using this unique approach, and in the third section we describe a sampling of controversial questions that the hypothetical society paradigm is uniquely situated to answer.

In sum, an empirical perspective on thought experiments suggests ways to overcome the replication and “idiosyncratic intuition” problems that plague many philosophical thought experiments on justice. A thought-experiment perspective on empirical studies suggests ways to overcome the partiality problems, both with respect to facts and motivations, that plague many empirical studies of justice. The hypothetical society paradigm combines elements of thought experiments and laboratory experiments in an effort to avoid both sets of problems.5

II. Justice in Hypothetical Societies

5 A closely related device for studying justice judgments is the vignette study (e.g., Bukszpare & 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 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.
The benefits of the hypothetical society paradigm are better appreciated by examining the empirical pay-offs of this approach to date. The general descriptive message from the hypothetical society studies is that, above a safety-net floor constraint, normative agreement about distributive justice varies greatly, depending on participants’ political ideology and how ambiguous the relationship is between merit and outcomes, but agreement about corrective justice is quite context-insensitive and appears to be less ideology-sensitive than agreement about distributive justice. The general methodological message from these studies is that the hypothetical society paradigm can easily be adapted to address a range of foundational questions.

Most studies using the hypothetical society paradigm examine the 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 hypothetical society 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, medium, or low, 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 to the societies and to make pair-wise comparisons of all possible income distributions for one of the societies, choosing which distribution within each pair was more fair, 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 a quality subject to efficiency constraints), and Boulding’s (1962) compromise theory (emphasizing efficiency subject to equality constraints—in which minimum equality is required by government ensuring a safety net for the poor but the goal of prosperity is encouraged by rewarding individual effort above this social safety net).

We found that, consistent with Boulding’s compromise theory, both liberals and conservatives were willing to accept considerable inequality of wealth in high meritocracy societies but with the reservation that distributions allowing persons to fall below the poverty line remained unpopular for both ideological groups even in high meritocracy societies (a finding somewhat consistent with choices by Frohlich and Oppenheimer’s experimental groups favoring 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 which suggests that implicit assumptions of meritocracy may have driven Frohlich and Oppenheimer’s experimental groups to favor a modified utilitarianism). Liberals and conservatives disagree 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 views about 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 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 redistributions (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 where there was 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. This finding that detached observers of imaginary societies found the fact of redistribution normatively significant, and generally distasteful, points out the practical problems faced by politicians who argue for redistributive policies, and points out the need for philosophical theories of distributive justice to consider
whether (or when) the different normative concerns primed by distributive versus redistributive
mindsets should count in normative theories of justice.

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, Mellers and Ordóñez (1994) utilized 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 Mellers and Ordóñez 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 who self-reported low socioeconomic 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 that may be contaminated by self-interested
concerns.

Recently, Scott and his colleagues (Scott et al., 2001) employed a variant of the
hypothetical society paradigm to compare the role of equality, efficiency, merit, and need in
people’s judgments 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 et al., 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 for racial minorities. Also, Michelbach and colleagues (2003), with a refinement to the hypothetical society paradigm that arguably provided a cleaner test between egalitarianism and Rawls’ maximin principle than found in our original study (Mitchell et al., 1993), found that a significant number of participants endorsed the maximin principle, but many other participants deemed merit an important principle and deviated from a strict adherence to the maximin principle.

These studies by Scott and his colleagues 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 within a society. However, these studies and their findings of gender and racial gaps in the weight placed on meritocracy in justice judgments also caution against making generalizations about the role of meritocracy in justice judgments and suggest that White men, women, and minorities, who are likely to have had very different experiences with meritocracy in the United States, may have difficulty divesting themselves of their life experiences and placing themselves into the position of impartial observer.

Most recently we used the hypothetical society paradigm to examine the long-standing debate within legal theory on the relationship between corrective justice and distributive justice (Mitchell & Tetlock, in press). Some legal philosophers claim that corrective justice is parasitic

6 Corrective justice provides, 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
on distributive justice, with a tortfeasor having a duty only to repair a 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 members 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 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 determine 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 corrective justice required that the tortfeasor make the victim whole regardless of distributive justice conditions in the society and regardless of whether the victim had insurance that would otherwise cover the harm done.

Only under conditions of extreme injustice in the distribution of resources did a majority of participants deem it just that a tort go unrepaired. Thus, in a hypothetical society in which a racial minority perpetrated its hold over power and wealth through discriminatory policies that
kept the racial majority in poverty, a majority of participants felt that justice did not require that
an impoverished member of the racially-oppressed majority compensate a wealthy member of
the racially-oppressive minority for a negligent tort (and with a great majority of this majority
being liberals). However, in this same society where the tort was the intentional tort of
conversion of a valuable watch owned by the rich man, a majority of 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).

The findings of this study 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 experimental research into corrective justice, but our findings point out the need to
better understand the scope, source, and function of the norm of corrective justice and its relation
to the norm of retributive justice, which has received much more empirical attention (e.g., Darley
& Pittman, 2003), but both of which have received considerably less attention than distributive
and procedural justice.

Second, these findings further illustrate the malleability of the hypothetical society
paradigm. Outside of an experimental setting such as that supplied by the hypothetical society
paradigm, it would be very difficult to conduct an empirical test of the different theoretical positions on the relationship between norms of distributive and corrective justice. And unlike much analysis within political philosophy, which argues for what societies should do to be moral and just, conceptual analysis within the law is concerned with making sense of why legal systems operate as they do (e.g., Coleman, 2001). Thus, empirical studies such as ours may shed considerable light on the psychological forces that give rise to or sustain a norm of corrective justice that ultimately finds expression within the law.

Although we believe that there is much to gain from using the hypothetical society paradigm to study judgments of justice in all their richness, and our most recent study shows that this paradigm can shed light on more than just judgments about social justice, there are also other promising applications of this method. We discuss a sampling of these in the next section.

**III. Future Directions**

We divide this final section into two subsections. The first focuses on the challenges of teasing apart the complex mixtures of factual and value differences underlying disagreements on foundational policy issues, including: the criteria that society should use in judging the degree to which equality of opportunity has been satisfied and in deciding what, if any, additional policy interventions are warranted to promote equality of opportunity; the criteria that society should use in judging the wisdom of entering into trade-liberalization agreements that facilitate the free flow of capital, people, goods, and services across national borders; the criteria that society should use in judging whether it has struck the right balance among competing stakeholders in formulating rules for corporate governance. In pursuing these lines of inquiry, we suspect it will also be possible to bring into sharper focus how these policy debates are likely to evolve in
response to new evidence (requiring the updating of our beliefs about the facts) and in response to shifting perceptions of societal needs (requiring updating of moral-political values).

The second subsection explores the potential of the hypothetical society paradigm to distinguish competing psychological theories of justice, in particular, those theories that we designate here as transparent-functionalist theories (which posit that the motives and goals the people openly endorse are the key drivers of their justice judgments) and subterranean-motivational theories (which posit that people are largely or entirely unaware of the functionalist motives driving their justice judgments).

A. Policy Controversies

(1) The links between intergenerational mobility and meritocracy. In the first-generation of hypothetical society research, we were content with simple operational definitions of meritocracy that highlighted the relative causal importance of hard work versus luck in determining income. But many people find it difficult to view a society as meritocratic if the children of the relatively poor have virtually no chance of rising into a higher socio-economic 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, virtually everyone moves in a more egalitarian or leftward direction on income transfers when they believe they are confronted with a low meritocracy society.

This raises an interesting question of applied cognitive-dissonance theory: what should one do if one is a conservative in early 21st century America who learns that the best scientific estimates now are that (1) inequality is growing in the United States (the distance between the economic cellar and economic penthouse); (2) it is becoming increasingly difficult for people to
rise from poverty to prosperity in one or even two generations (more difficult than it used to be and more difficult than it is in Western Europe or Canada)?

In the real world, there is an obvious and, from a conservative perspective, plausible dissonance reduction strategy, namely, to argue that richer children have better prospects than poorer children because they have genetic endowments better suited to facilitate success in competitive market economies (more intelligent, higher energy levels, optimism, etc.). The list of possible DNA suspects is long so it is difficult to rule out this hypothesis with great confidence in the real world (which may be one reason why some egalitarians simply prefer to suppress such arguments with political correctness codes). It is easy, however, to rule out this alternative explanation in the hypothetical society paradigm: we need simply to stipulate that there are no genetically heritable differences in abilities or personalities or temperaments across social strata. We suspect that this stipulation would drive many conservatives to an alternative dissonance reduction strategy: namely, to argue that richer children have better prospects because their parents do a better job bringing them up and inculcating the character traits conducive to success in competitive market economies. Again, it is difficult to rule out this hypothesis in the real world but it is easy to do so in the hypothetical society paradigm. We can simply stipulate that rich and poor parents raise their children in remarkably similar ways.

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 some hard-core conservatives will still resist egalitarian policy interventions designed to improve schooling opportunities and networking opportunities for the poor (e.g., generous vouchers and affirmative action incentives). These conservatives might argue that previous generations of parents must have worked hard to ensure
that their descendents would have advantages–and it is a bad idea to destabilize that societal
expectation. But we also suspect that many moderate conservatives at this juncture will make
some policy concessions–and accept the need for egalitarian interventions of some form.

Of course, the hypothetical society paradigm can also be used to subject liberals to
various forms of cognitive-dissonance torture. For instance, what should one do if one is a
liberal confronted with a hypothetical society in which: (1) inequality, the distance between the
economic cellar and economic penthouse, is large and growing; (2) there is, however, a very high
safety net which guarantees that even the poorest of the poor receive very high quality schooling
and medical care that is not appreciably different from that available to the wealthy; (3) there is
virtually no intergenerational mobility but the best scientific evidence is that this is because
children from wealthier families have genetic endowments better adapted for success in
competitive market economies; (4) raising taxes any more on the wealthy will have disincentive
effects that slow economic growth and, in the long run, reduce the absolute standard of living of
the poor more than that of the wealthy; (5) new recombinant DNA technologies make it possible
to level the genetic playing field for the poor. We might call this the Herrnstein scenario in
which a true meritocracy perpetuates itself genetically—and the only viable policy intervention
becomes swallowing a pair of bitter ideological pills for egalitarians: acknowledging the
importance of genetic determinants of social class and abandoning a visceral distaste for
“eugenics.”

The potential experimental manipulations of "fact situations" are obviously numerous. If
we put aside manipulations of the heritability of intelligence and character traits, and of the
availability of currently-science-fiction technologies to equalize opportunity at the level of DNA,
we suspect that much of the policy tension will ultimately focus on the value trade-offs between
family autonomy and social equality. Egalitarian philosophers—from Rousseau to Marx to Cohen—have long recognized that, as long as the family is the social unit primarily responsible for socializing children and as long as some families are (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 the task of socializing children becomes an attractive option from this 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 suspect that transferring the task of socializing children to the state is a dangerous step toward totalitarianism and collective mind control. Rejecting a prominent state role in childcare is an equally easy choice from these points of view.

We suspect that most people are deeply torn by this value conflict and will oscillate somewhat 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). Building on earlier work on the value pluralism model (Tetlock, 1986; Tetlock et al., 1996), our prediction would be that people will develop integratively complex policy solutions to the dilemma (solutions that acknowledge and reason through trade-offs) only to the degree that simple modes of dissonance resolution have been blocked off in the hypothetical societies. These simpler modes of dissonance reduction include challenging the “fact situation” posited in the hypothetical society (e.g., “that is just too absurd to contemplate”) and the classic spreading of the alternatives response (denial of weaker political value and bolstering of the stronger value). Previous research also suggests the importance of blocking off decision evasion tactics.
such as obfuscation, buck-passing, and procrastination by communicating clearly to subjects that they are accountable for taking a stand. If integratively complex policy reasoning is one's ideal outcome from a process of deliberative democracy (Fishkin, 1992), this would be how to achieve it via the hypothetical society paradigm.

(2) The links between absolute and relative wealth within and across societies. 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 what poverty was and that poverty was bad. What counts as poor, however, in one society at one point in history may count as quite wealthy for that same society at previous point in history or for other societies at the same point in history. Middle-class societies in 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.

For orthodox free-market theorists, the solution is surgically simple: the logic of comparative advantage in international trade holds that 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. If rich countries would 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 be painful. International trade can produce major dislocations within societies. American blue-collar workers who were accustomed to earning $25 per hour run the risk of losing their jobs to Mexican workers who are rapidly becoming as efficient and were glad to make five dollars per hour—and these Mexican workers,
in turn, run the risk of losing their jobs to Chinese workers were glad to make two dollars per hour.

How should one respond if one is an egalitarian who is asked to judge the acceptability of a trade agreement that will increase inequality within one's own wealthy society (because the pay checks of one’s “own” working-class are in decline) and even increase inequality within poor societies (because the well-off in those societies gain more rapidly than do the poor) but 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)? Much depends on the degree to which one is a cosmopolitan egalitarian, who is concerned more with inequality on a global scale, versus a parochial egalitarian, who is concerned solely with inequality within one's own society. Much also depends on the escape routes one is offered from this dissonance-inducing problem (escape routes such as reserving some of the wealth generated by free trade for transfer payments to help those in one’s own wealthy society who are most adversely affected by free trade).

The maximum-value-conflict dilemmas probably take somewhat different forms for more conservative observers--those who are less concerned about equality (either within or across borders) and more concerned about national sovereignty and relative national power. For instance, how should one respond if one is a conservative confronted by a trade agreement that will increase the GDP of one's own society by 2% but increase the GDP of the other society by 5%? Much may depend here on whether that other society is perceived to be a potential geopolitical threat (Herrmann et al, 2001). Or how should conservatives respond when confronted by patterns of labor migration into one's own country that increase net wealth but threaten the traditional ethnic or racial or religious or linguistic identity of the society?
Again, it is worth emphasizing that the hypothetical society paradigm is uniquely well-suited to disentangle causality in these types of controversies. It is too easy for policy advocates in the real world to make up facts that obscure the underlying trade-offs: “no, those migrants are not making our society wealthier” or “no, a trade agreement will not really make our society wealthier,” and so on. It is also worth emphasizing that we would expect people to engage in the hard mental work of crafting integratively complex trade-offs only when we have successfully foreclosed all of the simpler dissonance-reduction options, such as redefining the facts, spreading the alternatives by bolstering and denial, and slipping into decision-evasion mode.

(3) The links between corporate governance, profits, and other social objectives. Corporations can be viewed as mini-societies of their own. Corporations also raise unique challenges of governance because of the complex principal-agency relationships connecting the formal owners of the corporation, the principals or stockholders, and the agents responsible for the daily operations of the corporation, the top executive team and the employees. The shareholders obviously want to maximize their return on their investments whereas the top executive team obviously wants as large a fraction of the profits as possible for themselves (either directly via salary and stock options or indirectly via managerial perks such as lavish offices, chauffeurs, and private planes).

The now conventional solution to this principal-agent problem–formally articulated by Jensen and Meckling (1976)–is stock options. By linking managerial compensation to the price of the stock–the same index that determines returns to shareholders–the idea was to transform an adversarial relationship into a collaborative one. But, the devil as always lurks in the details and much hinges on how such stock-option agreements are written. If the agreements are written by members of Boards of Directors who are cronies of the CEO, there is a good chance that the
agreements will guarantee massive rewards for relatively mediocre performance (e.g., Khurana & Pick, 2005).

The hypothetical society paradigm can be used to explore how tolerant people are of inequality within the corporate world under various business and regulatory conditions. One libertarian reaction is that any pay package is acceptable as long as it is a contract freely entered into by the shareholders (or their representatives on the Board of Directors) and the CEO. But even some libertarians have difficulty when the corporation is functioning under a regulatory regime that makes shareholder democracy difficult and the extraction of large managerial rents easy. (Hard-line libertarians tend to be unrelenting: they argue that if someone does not like the ground rules in one capitalist economy, move on to another.)

To create maximum value conflict for libertarians and many conservatives, one would describe corporations in hypothetical societies that have created regulatory ground rules that favor managers over shareholders and one would describe pay packages that reward CEOs lavishly even for relatively poorer mediocre performance. These would be ideal conditions for pressuring anti-regulation observers to make integratively complex concessions and trade-offs in response to pro-regulation arguments. To create maximum value conflict for egalitarians, the hypothetical societies would need to stipulate that the CEOs make enormous amounts of money only when they are playing a key causal role in generating massive wealth for the shareholders and that significant amounts of that wealth trickle down to the employees, that the ratio between highest-paid and lowest paid workers is staggeringly large (say, 500:1) and growing, and that these super-incentives for top executives are indeed necessary causes for stimulating rapid-based entrepreneurial growth that makes pay increases for the lowest-paid possible. These would be
ideal conditions for pressuring critics of CEO compensation to make integratively complex concessions in response to anti-regulation arguments.\footnote{The questions we explore in the text are meant only to illustrate the types of policy controversies that may be explored using the hypothetical society paradigm. Other examples include the criteria that society should use in judging how far to take advances in biotechnology that have the potential to transform the nature of human nature and how society should deal with intergenerational externalities such as global warming and deficit spending. Although we have yet to collect data on most of these questions, one question that we have explored in some pilot research is the degree to which people do or do not look hard at surface appearances when looking harder has the potential either to reinforce or undercut their belief systems (and either make trade-offs easier or harder). For instance, if one is a libertarian who is motivated to believe that corporate capitalism is functioning as it should, then one should be quite willing to take at face value the information that CEO compensation is linked to a stock option plan whereas if one is an egalitarian who is motivated to perceive corporate capitalism as malign, then one should be more suspicious of the stock option plan (suspecting cronyism and corruption).}

**B. Exploring Further Theoretical Implications**

Skeptics of the hypothetical society paradigm could argue that it is equipped to tap into only relatively superficial psychological processes to which people have ready conscious access and which people are not embarrassed about revealing. The skeptics are correct that we have thus far tended to take the political values and intuitive philosophies of our respondents at face value. If respondents say that they are Rawlsian egalitarians or Nozickian libertarians 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 intuitive political philosophers were making good-faith efforts to understand the world around them and reach defensible conclusions about what constitutes a fair social order. From the skeptics' perspective, we have yet to seriously explore the possibility that motives to which our respondents do not have conscious access (or might be embarrassed to admit) are systematically influencing judgments of macro-social fairness.

It is not difficult, however, to adapt the hypothetical society paradigm to test less transparent, more subterranean-motivational theories that posit that people have very limited access to the true functionalist forces driving their judgments. Over the last century and a half,
behavioral and social scientists have advanced an extraordinarily ingenious array of such theories. We would include under this rubric the work of psychodynamic scholars who have argued that insecure authoritarian personalities derived psychic gratification from exercising symbolic dominance over those below them in the societal pecking order; the work of social-dominance theorists who have drawn on evolutionary and other speculation to reach a strikingly similar conclusion about high scorers on their measure of social-dominance orientation; and the work of system-justification theorists who have posited a deep-rooted psychological tendency to justify existing status hierarchies (a tendency that there's a marked family resemblance to the classic Marxist notion of false consciousness).

In our view, the hypothetical society paradigm is the optimal context for testing the relative merits of more transparent-motivational and more subterranean-motivational theories. It is instructive, for instance, to examine the deep indeterminacy problems that arise in testing a formulation such as system-justification in the real world. We repeatedly run into a problem that one person's reason for holding a belief (say, about social-class differences in intelligence) can often be viewed by others as a mere rationalization (say, as a means of justifying existing inequality—or, to flip it around, as a means of justifying proposals to redistribute income).

Commenting on indeterminacy problems of this nature, Rubin and Hewstone (2004) note that system-justification theorists 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, etc., among the poor). To use their analogy to a football game, do we feel comfortable concluding that members of the losing team who attribute their defeat to their own shortcomings are guilty
of outgroup favoritism and system justification? Or should we conclude that they are engaging in highly adaptive forms of self-criticism?

The list of reason-rationalization indeterminacy problems 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? 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 my 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 Douglass North (1981, 2005)? If one believes that intelligence, the capacity to delay need gratification, and other positively charged traits conducive to economic success are found more often among the economically successful, does that count as evidence of system-justifying cognition or as evidence that I am in touch with sociological reality?

To the best of our knowledge, no social psychologist has yet discovered a method of answering these questions by asking people questions about the social world they actually inhabit. There are simply too many sources of uncertainty, too many historical confounds, to permit investigators to draw strong functionalist conclusions about whose opinions are serving subterranean versus more transparent motivational forces.

If we want to escape these indeterminacy dilemmas, we need to explore human judgment in imaginary social worlds that we can experimentally manipulate in precise, conceptually targeted ways.
Using the hypothetical society paradigm, it should be possible to answer questions of the following sorts:

(a) Are people so committed to defending traditional status hierarchies that they would rather preside over a shrinking economic pie to which their group continues to enjoy a high priority claim over expanding an economic pie to which their group enjoys a lower priority claim (but is still, in absolute terms, better off)? If so, we should find that high scorers on measures of system justification and social dominance and perhaps authoritarianism should be especially likely to reject societal reforms that simultaneously make everyone wealthier and reduce inequality and blur traditional status distinctions.

(b) Do people become more punitive toward norm violators not simply as a function of the severity of the violation of the collective conscience of society (as Durkheimian theories suggest; see Tetlock et al., in press) but also as a function of the degree to which norm violations destabilize existing patterns of social dominance? If the latter, we should expect stronger activation of the punitive prosecutorial mindset when people are led to believe (holding absolute amount of crime constant) that patterns of criminal behavior are breaking down the boundaries between the rich and poor, as opposed to having no net effect or even perpetuating them.

(c) Does tolerance of inequality reflect conscious consideration of existing societal circumstances or unconscious system-justifying motives? How much do people feel that they should change their minds when they learn not just about the distribution of objective wealth in a society but also about the distribution of subjective well-being (the extent to which people feel they are living fulfilling and happy lives)? One hypothesis—which we view as being in the spirit of social-dominance and system-justification
theories—is that observers, especially those who score highly on measures of social dominance and system justification, will latch onto evidence that the poor are as happy as the rich (or even happier) as grounds for dismissing the need for redistribution. An alternative hypothesis—which we prefer—is that support for redistribution may indeed fall when people learn that the poor are as happy as, or more happy than, the well-off but that this shift in policy preferences should not be treated as a mere rhetorical cloak for social-dominance or system-justification goals but rather as based on a reasoned reappraisal of the degree to which (1) the poor have serious unsatisfied basic human needs, (2) the rich have privileged access to resources that greatly enhance human experience, (3) the poor may have opted for lower incomes in order to gain more leisure or family time, and (4) the rich may have opted for higher income and the sacrifice of leisure and family time.

These questions are unanswerable in real-world debates because it is so easy for advocates–motivated reasoners that they are (see Kunda, 1999)—to invent facts that conceal potential trade-offs and double standards (an invention process that, if it is to serve its subterranean-motivational function, needs to occur automatically, outside of awareness). But these questions become answerable in the hypothetical society paradigm because it is so difficult for advocates to conceal the same trade-offs and double standards in a world in which all of the key factual parameters have been pre-specified by experimental fiat. The hypothetical society paradigm then becomes the platform for previously impossible conversations between theorists. For instance, if transparent-motivational theories outperform subterranean-motivational theories in the starkly simplistic hypothetical society environment, subterranean-motivational theorists can argue that more socially undesirable motivational forces only come into play when people
have rationalization covers (or attributional ambiguity). We do not automatically dismiss this argument as a desperate patch-up operation of a degenerating research program. This theoretical defense may be defensible—and the best way to determine whether this is so is by gradually adding complexity to the hypothetical society paradigm so that we can identify more precisely exactly how much attributional ambiguity we need to create for subterranean-motivational theories to gain explanatory traction.

To put our argument in a nutshell, the hypothetical society paradigm reveals a paradox: we have to step into imaginary social worlds if we want to understand how we understand our own social world.
References


