The Promise (and Limits) of Neuroeconomics

Jedediah Purdy♥

Abstract:

Neuroeconomics – the study of brain activity in people engaged in tasks of reasoning and choice – looks set to be the next behavioral economics: a set of findings about how people make decisions that casts doubt on widely accepted premises about rationality and social life. This essay explains what is most exciting about the new field and lays out some specific research tasks for it.

By enabling researchers to view the mind at work, neuroeconomics puts in question the most basic premise of twentieth-century empiricism, sometimes called positivism or behaviorism: that people are black boxes to one another, and scientific social inquiry must observe only their objective behavior, what they say and do. This premise came to the center of neoclassical economic method via the 1930s work of the hugely important economist Lionel Robbins, and it occasioned a methodological split in social inquiry. Positivists (most importantly, economists) follow the strictures of studying observable behavior, while interpretivists insist that we cannot understand social life without interpreting the minds and intentions of others, even though we cannot view them directly.

The limits of these two methods have restricted progress in understanding three critical issues for legal scholarship: how people solve collective-action problems, why some people are more susceptible than others to extremist political appeals, and whether "commodification" creates a conflict between economic rationality and other values. I show how the progress already made in neuroeconomics could make each of these questions more answerable than it has recently seemed, with potentially significant payoffs.

♥ Assistant Professor of Law, Duke University School of Law. A.B., Social Studies, Harvard College, J.D. Yale Law School. Thanks to Jeff Powell for penetrating comments on an earlier draft. I am indebted to David Grewal for directing my attention to the relationship between the history of economics and the philosophical problem of interpersonal intelligibility.

Table of Contents:

What is the promise of neuroeconomics? This nascent field has converted a few laboratory results into great excitement.[1] Using magnetic resonance imaging (MRI) technology, neuroscientists observe blood flow in the brains of people engaged in familiar tasks of reasoning and choice. The results, while few and crude, already suggest a lot about which regions of the brain engage in which dimensions of reasoning, choice, and moral judgment. Like alchemists who dreamed of finding an elixir of life and mystics who labored to see the face of God, investigators speak of "seeing utility."[2] What we have seen only in a mirror darkly, we might now meet vividly and immediately. And in an age when economics has vast reach and authority both in the academy and in the larger culture, the analogy to encountering life's essence or God's person might not be altogether far-fetched. For the first time, we could unveil and examine the hidden logic of our lives. This is the promise.

There is a deflationist response, which sees neuroeconomics as the false promise of an academic culture too enamored of laboratory science and ephemeral trends.[3] Even

---

[1] *See*, *e.g.*, Owen D. Jones & Timothy H. Goldsmith, *Law and Behavioral Biology*, 105 COLUM. L. REV. 405, 424 (2005) (identifying neuroeconomics as part of a cluster of rapidly advancing inquiries in the biology of human behavior that might significantly change our estimation of how law interacts with individuals' reasoning and decisions); Troy A. Paredes, *Too Much Pay, Too Much Deference: Behavioral Corporate Finance, CEOs, and Corporate Governance*, 32 FLA. ST. U. L. REV. 637, 715 n. 179 (2005) (noting that neuroeconomics promises to "reduce the guesswork" about the effects of incentives on decision-making that been so important to shaping corporate governance); William Bradford, *In the Minds of Men: A Theory of Compliance with the Laws of War*, 36 ARIZ. ST. L. J. 1243, 1420-21 (2004) (arguing that advances in neuroeconomics may make possible the explanation and prediction of compliance with and violations of the international humanitarian law); Paul Steinberg & Gerald Lescatre, *Beguiling Heresy: Regulating the Franchise Relationship*, 109 PENN. ST. L. REV 105, 129-30 (2004) (arguing that neuroeconomics' revelation of the limits of the rational-actor model of human behavior can improve franchise regulation); Sandra Blakeslee, *Brain Experts Now Follow the Money*, N.Y. TIMES at F1 (June 17, 2003) (noting the growing importance of neural imaging in research into behavioral economics).

[2] Presentation of Colin Camerer, Duke University, Dec. 8 2005.

[3] *See*, *e.g.*, Terence Chorvat & Kevin McCabe, *Neuroeconomics and Rationality*, 80 CHI.-KENT L. REV. 1325, 1237-38 (2005) (noting the argument that neural correlates of already observable behavior give us no information that we do not already have).

at its most sophisticated, brain imaging can only give us a map of correlations, physical events in the brain that correspond to the activity of the mind. But neoclassical economics rests on the axiom that utility functions reveal themselves through actual choice – revealed preference – and thus that utility maximization literally *just is* what we do. Thus it gains nothing from peeking into the black box of the brain. As for the heterodox proposals of behavioral economics, with its observations about systematic non-maximizing behavior, there too investigators are concerned with the observable behavior of human beings, with their implication for the design of institutions and the achievement of cooperation. Correlations add little or nothing.

In this Essay, I argue that the promise of neuroeconomics is considerable – even greater, in fact, than the claims its advocates have made for it. To understand why this is so, it is necessary to address a basic methodological division in law and social science. For much of the last century, social inquiry, including legal inquiry, has been marked by a schism between two methodological schools, neither capable of giving a fully satisfactory account of social life. One camp, which we might call *positivist*, takes observation, prediction, and, ultimately, falsifiability as the elements of its gold standard. The other, the *interpretivist* camp, insists that these criteria misapprehend the nature of human activity – both individual action and social life – because whatever people do, they do as self-conscious and self-interpreting actors. On this account, giving a description of human activity that omits the pervasive and perennial activity of interpretation is like dancing about architecture: it is a category mistake. Whoever makes the mistake of trying to describe human activity in positivist terms ensures that the vocabulary of his description cannot capture essential features of object being described.

Neuroeconomics matters to this schism because the schism arose from a very particular episode in the history of twentieth-century thought: a deep methodological embarrassment about the invisibility of any mind to any other mind. In this view, we are black boxes to one another, knowable only through observable behavior that may or may not correspond to any particular internal state. Even the conviction that others have internal states like those we experience in ourselves, rather than something completely different, is a point of faith rather than knowledge. As I will show later, this difficulty motivated the shift to the methodology of contemporary microeconomics and economically informed social inquiry. The promise of neuroeconomics is to bring other minds one giant step nearer visibility. The effect of this change would be to soften the methodological opposition between positivist prediction and falsification on the one hand and interpretativist effort on the other.

Diminishing the methodological opposition matters because the respective limits of the two approaches have rendered intractable questions of absolutely first importance for legal scholars. Problems about what people value and how, how and why we overcome collective-action problems, and the social-psychological bases of liberalism and authoritarianism have suffered from division between the two camps, neither of which can do them justice. Positivist approaches beg the most important question: why people act as they do. Without this, *every* "explanation" is just a correlation. Interpretivism, for its part, carries the cardinal sin of being non-falsifiable: as its most candid practitioners concede, the hermeneutic circle really is closed.

In Part I of this Essay, I survey the most provocative findings of neuroeconomics and the closely allied inquiry into the neurology of moral reasoning. I then present the

competing claims that scholars have made for the relevance of this work to law and social inquiry. In Part II, I set the rise of neuroeconomics in the context of the twentieth-century methodological split between positivism and interpretivisim, showing how the invisibility of other minds drove the schism. In Part III, I present several very important problems that the schismatic methods have followed into culs de-sac, and suggest how even a moderately successful neuroeconomics could aid a fruitful rapprochement. I argue that the kind of information that neuroeconomics provides cannot dissolve the methodological opposition I have described, and that we should not wish otherwise, as each methodological approach has value not reducible to the terms of the other. Nonetheless, the modest rapprochement that neuroeconomics promises might be very valuable. Part IV concludes.

I. Neuroeconomics So Far

The tools of neuroeconomics are blunt enough that most findings involve a simple, even inevitably simplistic, distinction between areas of the brain associated with emotions and others associated with cognitive activity such as conscious memory and calculation.[4] The basic technique of MRI brain-scans is to measure the flow of blood to various brain regions as subjects perform exercises in reasoning and choice. Findings are

[4] *See* Jonathan D. Cohen, *The Vulcanization of the Human Brain A Neural Perspective on Interactions Between Cognition and Emotion*, 19 J. ECON. PERSPECTIVES (No. 4) 3-10 (2005) (laying out this methodological account). Because this essay is intended for legal scholars interested in the results of neuroeconomic studies, and a fair amount of training in neurology would be necessary would be necessary to assess researchers' classifications of the various brain regions, I distinguish in the text of the essay between "cognitive" and "emotional" regions. I do, however, note in footnotes which regions are under examination in each study I summarize. Almost all studies identify the prefrontal cortex with reasoning ability and cognitive control, that is, the ability to direct action in keeping with abstract commitments or intentions, especially when this involves overriding intuitions or reflexes. *See id.* at 10. By contrast, a set of subcortical structures, which evolved earlier and are located deeper in the brain, are thought to register immediate emotional or visceral responses to events or circumstances, responses that are transmitted directly to regions of the frontal cortical lobes: the amygdale, the medial and orbital regions of the frontal cortex, and the insular cortex. *See id.* at 8-9.

generally of this form: variations in the reasoning exercise correspond to changes in relative activity among brain regions, suggesting that certain exercises invoke a greater share of cognitive capacity while others invoke emotional responses in greater measure.[5] These results tend to be most interesting in situations of a particular type: where theorists have speculated that several reasoning tasks are indistinguishable as a matter of principle (where the principle may be either self-interest maximization or some stipulated moral rule), yet individuals actually engaged in the exercises reach different results across tasks.[6] The distribution of neural activity in each task tends to suggest something about the character of divergence in actually observed reasoning across tasks.

Interpretation of these results generally relies on a model of brain activity in which conscious processing is regarded as a relatively scarce resource, able to concentrate on only one problem (or at most on a handful of problems) at a time, and thus jealously husbanded.[7] The model regards unconscious processing as relatively abundant, able to receive, integrate, and evaluate a great deal of information simultaneously.[8] Thus, on this model, an efficient brain strives to convert conscious processing to unconscious processing by, in effect, generating unconscious programs to handle information and navigate circumstances that would otherwise require conscious processing.[9] Learning to ride a bicycle or drive a car provides a simple example: at first the activity requires most of the mind's available conscious attention; soon enough, though, it recedes to the "background" of unconscious processing, and one can work out a dinner menu or a math problem, have a phone conversation, or compose a sonnet while riding or driving.

---

[5] *See id.* at 4-6.
[6] I discuss several finds of this sort in I.A-B, below.
[7] *See* Cohen, *supra* n. 5 at 4-6.
[8] *See id.*
[9] *See id.*

Without hazarding complex hypotheses in evolutionary or psychodynamic psychology, neurological researchers often suppose that emotional responses express unconscious processing, developed either genetically or in individual maturation, while cognitive reasoning makes greater demands on conscious processing.[10]

A. Moral Reasoning

The "trolley problem," much discussed by Judith Jarvis Thomson, is a modern classic in the minor genre of moral philosophy dilemmas.[11] The problem is to generate a moral principle that accounts for widely held intuitions about two hypothetical decisions. Both decisions have the same formal stricture: whether to sacrifice one life to preserve a larger number of lives. Yet depending how the decision is structured, people persistently come to very different conclusions. Hence the difficulty: is moral reasoning principled at all, or is it just dressed-up ad-hockery?

The first version of the decision is the simplest. The subject is asked to imagine herself the switchperson on a railroad. As a train approaches, she realizes that five people stand on the track, and that she has no way to warn them. All she can do to avert their deaths is to throw the switch and divert the train to a side-track – where, unfortunately, a single person is standing. Most subjects agree that it is appropriate – a word eliding the difference between "permissible" and "required" to throw the switch. This suggests utilitarian reasoning: other things equal, the greatest aggregate good is served by preserving five lives, even at the cost of one.

---

[10] *See id.*

[11] *See* JUDITH JARVIS THOMSON, *Killing, Letting Die, and the Trolley Problem*, *in* RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY 78-93 (1986); THOMSON, *The Trolley Problem*, *in id.* at 94-116. The problem was originally formulated by Philippa Foot. *See* Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 5 OXFORD REV. (1967).

Now change the problem. Suppose there is no chance to switch the train to a second track. Instead, as the train barrels toward the five innocents, the subject is to imagine herself standing on a footbridge over the track. (This is sometimes called the "footbridge problem," for the sake of keeping track. Pun inadvertent.) She is standing next to an enormously large person. (Readers will have their own favorites, perhaps ranging on partisan grounds from Arnold Schwarzenegger to Ted Kennedy.) If she pushes her companion onto the track, his weight will stop the train. She is herself too slight to make the difference, even if she were inclined to self-sacrifice.

Formally speaking, the footbridge problem appears identical to the original trolley problem: Do you take an action that will result in one death but save five lives? Yet many subjects who approved of the action in the first problem now find it inappropriate. The minority who stick to their utilitarianism tend to hesitate and waver before overriding a deep reluctance to push, which they did not feel, at least not in the same degree, when choosing the utility-maximizing action meant pulling the switch.

How, then, to reconcile the two sets of responses under a single principle? Does the ambition even make sense? One effort involves suggesting, on loosely Kantian grounds, that the difference is in using another as a means to an end: in the footbridge problem, sacrificing one's large companion is *instrumentally necessary* to saving the five others, while in the trolley problem, the death of the one is incidental to the act of saving the five. Inconveniently, however, a simple change in the facts of the trolley problem spoils the Kantian hypothesis. Suppose the side-track is a loop that rejoins the main track just before the place where the five are standing, but that there is a large person on the side-track whose bulk will be enough to stop the train. Now, throwing the switch to

move the train will still save the five, but the death of the one is instrumentally necessary. Yet most subjects now move back to a utilitarian judgment, and say throwing the switch is appropriate.

The search for a unifying principle seems lost. We are free to say, of course, that people are sometimes right and sometimes wrong; everything said so far is description, not prescription. What, though, can we say about the way moral reasoning proceeds? That we accept utilitarianism up to the point of a switch, but not to the point of a shove? If so, how should we understand the difference?

After decades of discussion that circled these problems, a group of neuroscientists devised an experiment effectively testing the "switch v. shove" account. They presented subjects with a battery of 60 hypothetical decisions, codes into three categories: (1) "up-close and personal" moral decisions involving relatively intimate actions, including a version of the footbridge problem, a case of harvesting one living person's organs involuntarily to save five others, and a case of throwing people off an overcrowded lifeboat; (2) impersonal moral problems, including a version of the trolley problem and a case of voting for a policy that would cause more deaths than the alternative; and (3) non-moral decisions, such as whether to take a train or a bus given certain time constraints.[12] The moral-personal problems were associated with increases in MRI signal of between 20% and 40% in brain areas associated with emotion; by contrast, none of these areas experienced as much as a 20% increase while subjects addressed the impersonal moral

[12] *See* Joshua D. Greene, et al., *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 SCIENCE 2105, 2107 (Sept. 14, 2001). The areas of the brain coded for cognitive activity in this experiment were the medial frontal gyrus, the posterior cingulate gyrus, and the left and right anterior gryi. The middle frontal gyrus and the left and right parietal lobes were associated with cognitive activity and control, i.e., the self-conscious governance of moral judgment by abstract principles.

problems, and for two of the four areas the increase was below 10%.[13] The areas associated with cognitive processing, however, showed increases of 15%-25% during the impersonal moral problems, and much lower, or even declining activity for the personal moral problems.[14] Non-moral problems were associated with small increases or even decreases in activity in areas associated with emotion, and increases rivaling and sometimes surpassing those of impersonal moral problems in areas associated with cognitive processing.[15] The results suggested that problems involving direct injury to another person invoke more or less automatic, emotional responses of aversion, which are usually powerful enough to pre-empt consciously held contrary principles.[16] Problems that did not invoke these responses appeared to engage a different system, a body of utilitarian principles associated with cognitive processing.[17] Thus "moral reasoning" appeared to be divided among two broad systems of cognition, which responded to different features of a problem according to different criteria and thus produced incompatible results.

The researchers advanced the experiment in a later version, in which they sought to understand relations among the competing systems of moral response.[18] This time, their key distinction was between "easy personal" judgments and "hard personal" judgments. Both categories proposed tradeoffs between one life and benefits to others and asked the subject to imagine taking the life at close range. The exemplary "easy

---

[13] *See id.* at 2106.
[14] *See id.*
[15] *See id.*
[16] *See id.* at 2107.
[17] *See id.*
[18] *See* Joshua D. Greene, et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 NEURON 389-400 (Oct. 14, 2004). The region of the brain associated with cognitive activity and control was the anterior cingulated cortex, while emotional responses were associated with the dorsolateral prefrontal cortex region. *See id.* at 390.

personal" case, which they called "infanticide," asked the subject to consider killing an unwanted newborn. The exemplary "hard personal" case presented a decision whether to smother a crying infant to prevent enemy soldiers from discovering and murdering an entire family – including the infant.[19] The ambition of the experiment was to observe brain activity as two types of response – emotional aversion and utilitarian cognition – came into contest. They found first that difficult judgments – measured by the time subjects took to make their decisions, which corresponded fairly closely to researchers' conception of the questions as easy or difficult – corresponded to increased activity in brain regions associated with mediating conflict among responses and enforcing cognitive control over emotional impulses.[20] Second, they found that subjects who reached a utilitarian judgment about the crying baby case – who, in other words, overrode strong emotional aversion with principle – showed higher levels of activity in the cognitive-control regions than those who finally declared it inappropriate to smother the child.[21] In other words, the more conflict between emotion and principle, the more cognitive activity; and cognitive activity rises to an even higher level when abstract principle overrides emotional response.[22]

Even taken in a charitable spirit, this research raises many more questions than it answers. The categories of "emotion" and "cognition" are diffuse and have uncertain boundaries, not least vis-à-vis each other. Utilitarian reasoning, for instance, presupposes a judgment about who counts as a member of the relevant moral community – a question

[19] *See id.* at 394-96.
[20] *See id.* at 397.
[21] *See id.*
[22] Interestingly, the judgment in favor of smothering the crying baby is also associated with increased activity in one region associated with emotion. Greene and his colleagues suggest, following David Hume, that this region may be associated with an emotional motivation to cognitive activity, perhaps a commitment to principle or a benevolent attitude toward humanity.

fought in the nineteenth century over race and class, and in the twenty first over non-human species. In one of the classic American comments on race and moral blindness, Mark Twain had Huckleberry Finn declare that a steam-engine explosion had hurt nobody – only "killed a nigger." [23] Is the judgment underlying that statement cognitive or emotional? What about the change in ideas and sentiments that produces my reaction as I look at the phrase I have just typed: intense emotional discomfort at a word that carries the ugliest associations, combined with a principled belief that a modestly unsettling example is the best way to communicate the force of the question? What systems would be engaged by, for instance, a "principle" of emotional spontaneity – perhaps adopted by someone who believed his ethical and interpersonal shortcoming resulted from an excess of abstraction and lack of connection with his own feelings? The examples are easy to multiply, and are not trivial: they lie near the heart of actual moral experience.

Nonetheless, even these relatively primitive findings are enormously provocative. They call into question two fairly widespread ideas in moral psychology. One is the cognitivist idea that moral reasoning is basically a matter of applying principles, and that moral development occurs through movement from one principle to another.[24] The other is the anti-cognitivist idea that moral judgments are basically elements of perception – that sensory information hits consciousness already interpreted as "good," "bad," and so forth.[25] The results surveyed here strongly suggest interaction between two sets of

---

[23] MARK TWAIN, THE ADVENTURES OF HUCKLEBERRY FINN XXX (XXXX).

[24] This view is canonically associated with Lawrence Kohlberg's account of moral reasoning as culminating in the individual's capacity to apply abstract principles of general reach to specific cases. *See* LAWRENCE KOHLBERG, ESSAYS ON MORAL DEVELOPMENT (1981).

[25] A version of this view is sometimes associated with Martha Nussbaum's early work, notably MARTHA C. NUSSBAUM, THE FRAGILITY OF GOODNESS: LUCK AND ETHICS IN GREEK TRAGEDY AND PHILOSOPHY 13-17, 40-43, 316-17 (1986) (describing this view). Nussbaum has continued to take her inquiry into the relationship between reflective and unreflective, embedded and self-conscious, or emotional and cognitive moral evaluation in interesting and productive directions. *See* NUSSBAUM,

responses. One is emotionally powerful and largely pre-conscious in its evaluations, in the manner envisaged in the anti-cognitivist position. The other, which comes into play when the first system comes into conflict with explicit principle, seems to involve conscious application of principle both to decision problems and to the emotions that those problems evoke. The scenarios of the trolley problem, thus, do not display a single system of reasoning, but rather engage two systems, often complementary, which come into conflict as a series of hypothetical trains chug toward disaster.

B. Rationality and Reciprocity: Ultimatum Games and Other Decisions

The ultimatum game has been among the most productive experiments in behavioral economics. The game involves two players and a pot of money. One player proposes a division of the money, which the second player accepts or rejects. If the second player accepts the division, then the players claim their respective shares of the money. If the second player rejects the division, however, the pot of money "dissolves" and neither player takes anything.

The game provides a clean test of a prediction of purely self-interested behavior. According to that prediction, the first player should propose that the second take the smallest divisible unit of money – a penny, let us say, or, if a penny is intuitively no longer money at all these days, a nickel – because he will seek to maximize his gain from the transaction. And the second player should accept the division, because any offer above zero improves his position.

HIDING FROM HUMANITY: DISGUST, SHAME, AND THE LAW 21-70 (2004) (discussing the interaction of emotional response and moral judgment).

Offered the opportunity to demonstrate self-interested rationality, test subjects demurred. Student players' modal offer is almost always 50% of the pot, and mean offers are between 40% and 45%.[26] Respondents are not much more conventionally rational than offerors: when the offer falls below 20%, they reject it about half the time.[27] Skeptics suggested that the stakes of the game were too low to engage fully rational behavior, and that players misunderstood the one-off character of the transaction and imagined they were establishing reputations for selfishness or generosity; but experiments with higher stakes (up to a $400 pot) and with opportunities to learn and appreciate the rules of the game yielded much the same results.[28]

The ultimatum game has benefited from one of the first major cross-cultural experiments in behavioral economics. When students around the world turned up pretty much the same lab results, researchers took the ultimatum game to 15 herding, hunter-gatherer, and slash-and-burn agricultural societies in Asia, Africa, and South America.[29] Mean offers ran as high as 57% among the Lamalera, a Malaysian island people, and as low as 25% among the Quichua, semi-nomadic horticulturalists in South America. The Tsimane of Bolivia rejected not one of seventy offers, including five offers of below 20%. The Machiguenga of Peru, who live semi-nomadically in the tropical forests of the Amazon, most closely approximated the neoclassical model of self-interested behavior: almost half their offers were below 20%, and of those only one was rejected.[30]

---

[26] *See* Joseph Henrich, et al. *"Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies* at secs. 1-2 (forthcoming BEHAVIORAL AND BRAIN SCIENCES).
[27] *See id.*
[28] *See id.*
[29] *See id.*
[30] *See* Table 2 *in* Henrich, et al.

The persistent results of student ultimatum games inspired a functional explanation: that societies do better when their members are motivated to behave reciprocally and to punish non-reciprocal behavior.[31] Results from the intercultural ultimatum game experiments enriched the interpretive effort. To begin with, offers were generally too high to be consistent with a context-adjusted version of self-interest, in which players would be predicted to make offers that maximized expected income given relevant rejection thresholds.[32] Second, higher offers were significantly correlated with what the researchers called "payoff to cooperation" in economic life, that is, how much economic well-being depends on actors outside the household.[33] Higher offers were also positively correlated with the social groups' degree of market integration, that is, how frequently members engaged in market transactions.[34] The interpretation these results suggest, then, is that ideas of fairness and reciprocity matter everywhere, and modify the neoclassical supposition that individuals will act solely to maximize their own self-interest. But reciprocity and fairness appear to be variable, not fixed, both in their hold on individuals and, particularly, in their content across cultures. Moreover, day-to-day experience of reciprocity as integral to social life, whether through cooperation or through market transactions, appears, as first look, to train people in reciprocal behavior.[35] The results are compatible with the general interpretation of reciprocity as

---

[31] *See* Ernst Fehr & Urs Fischbacher, *The Nature of Human Altruism*, 425 NATURE 785-91 (2003).
[32] *See* Henrich, et al., at sec. 4.2.
[33] *See id.* at sec. 5.
[34] *See id.*
[35] This argument about market psychology is classically captured in one of Adam Smith's lesser-known arguments:

> The offering of a shilling, which to us appears to have so plain and simple a meaning, is in reality offering an argument *to persuade one to do so and so as it is in his interest*. Men always endeavour to persuade others to be of their opinion even when the matter is of no consequence to them. … And in this manner, every one is practising oratory on others thro the whole of his life.—You are uneasy whenever one differs from you, and you endeavour to persuade [him] to be

evolutionarily adaptive, but maintain considerable space for cultural plasticity and, consequently, individual habituation into or against reciprocity.

How, then, to conceive of the relationship between self-interest and reciprocity as elements of players' reasoning in ultimatum games? Is reciprocity an abstract normative principle akin to utilitarian reasoning, a reflective override of simple selfishness that takes into account the overall functional benefit of rewarding cooperative behavior and punishing defecting behavior? Does rational self-interest operate like a meticulous calculator of benefits and burdens, or like a hungry child reaching spontaneously for what most pleases it? Is the relationship between the two "principles" an integrated one, as philosophers hoped that the "principle" behind the trolley problem would be, or is, as neuroimaging studies suggest, a contest between distinct aspects of motivation and judgment?

A group of Princeton researchers addressed a part of this question by measuring the neural activity of respondents in an ultimatum game.[36] Researchers told respondents that some of their offers would come from living offerors, whom respondents briefly met before the game began, while other offers were generated by a computer.[37] In fact, the researchers manipulated the offers to ensure consistent distribution of "fair" (50/50) and

---

> of your mind … . In this manner [people] acquire a certain dexterity and adress [sic] in managing their affairs, or in other words in managing of men … . That is bartering, by which they adress [sic] themselves to the self interest of the person and seldom fail immediately to gain their end. The brutes have no notion of this …

ADAM SMITH, LECTURES ON JURISPRUDENCE 352 (R.L. Meek et al. ed.) (Oxford University Press 1978) (1762-63). I have elaborated on the consequences of this argument in Jedediah Purdy, *A Freedom-Promoting Approach to Property: A Renewed Tradition for New Debates*, 72 U. CHI. L. REV. 1237, 1251-58 (2005).

[36] *See* Alan G. Sanfey, et al., *The Neural Basis of Economic Decision-Making in the Ultimatum Game*, 300 SCIENCE 1755-1758 (June 13, 2003). Researchers associated the bilateral anterior insula with negative emotional states such as resentment and disgust, the dorsolateral prefrontal cortex with cognitive activity such as calculation, and the anterior cingulated cortex with cognitive conflict, i.e., a struggle to maintain cognitive control in the face of a strong emotional response.

[37] *See id.* ar 1756.

"unfair" (90/10 and 80/20) offers.[38] Respondents accepted all 50/50 offers and nearly all 70/30 offers. Rejection rates, however, ranged between 15% and 60% for lower offers; moreover, acceptance rates for computer-generated "unfair" offers were at least 20 points higher than for "unfair" offers linked to a human offeror.[39] These results suggest that the motive to refuse offers perceived as unfair is linked to an idea about *interpersonal* reciprocity, that is, that fairness is a matter of reciprocal relations between human beings, and unfairness a willful withholding of that reciprocity.

The neural results of the study were particularly interesting. Unfair offers from human partners were associated with elevated levels of activity in two areas of the brain. One, the dorsolateral prefrontal cortex (DLPFC) is associated with "cognitive processes such as goal maintenance and executive control," and was also associated with utilitarian judgments in the moral reasoning studies.[40] The other, the bilateral anterior insula (BAI), is associated with "negative emotional states" such as anger and disgust, as well as with sensations of hunger and thirst.[41] Not only were both regions engaged by unfair offers "from" human offerors; activity in the BAI was higher while considering offers that respondents rejected than while considering offers they accepted.[42] Moreover – although the researchers caution against quick conclusions from comparisons of concurrent activity in different areas of the brain – it appears that respondents reject offers when BAI activity is greater than DLPFC activity, and vice-versa.[43]

---

[38] *See id.*
[39] *See id.*
[40] *See id.* at 1757; Jonathan D. Cohen, *The Vulcanization of the Human Brain A Neural Perspective on Interactions Between Cognition and Emotion*, 19 J. ECON. PERSPECTIVES (No. 4) 3-24, 14 (2005).
[41] *See* Sanfey et al., at 1756.
[42] *See id.* at 1757.
[43] *See id.*

An unskeptical view of findings that are unavoidably crude suggests several tentative judgments. First, the motive of reciprocity in rejecting unfair offers appears linked with very basic experiences of affront and deprivation: hunger, thirst, repugnance, all of which register as physical and emotional facts, not problems for deliberate evaluation. Second, when confronted with unfairness, this motive appears to come into direct conflict with a cognitive effort at maximizing self-interest even at the cost of absorbing an insult. Under conditions of fairness, the two might easily work in harmony; but as with the trolley problem, under circumstances that pull apart their cooperation, two quite distinct modes of reasoning enter a struggle for dominance. Third – taking into account cross-cultural findings in ultimatum-game experiments – the threshold of "insult" that engages the BAI emotional response appears to vary with the principles of reciprocity or non-reciprocity carried out in the everyday activity of the respondent's society. Moreover, it is plausible to suppose that economic maximizing is also socially learned behavior, at least in some degree: the content of the DLPFC's cognitive override does not seem likely to be genetically programmed in its particulars, any more than a gene is likely to turn up for classical utilitarianism, strict Kantianism, or Kaldor-Hicks efficiency. The overall impression, then, is of an uneasy relation between cognition and emotion, in which the two are sometimes complementary, sometimes at odds, but distinctly and essentially expressions of different types of neural activity.

Other neuroimaging results are also intriguing. In another finding connected with reciprocity, researchers have found that areas of the brain associated with emotion and social engagement are activated in people who achieve high levels of cooperation in

repeated prisoners' dilemma type reciprocity games.[44] These studies also show that participants with high levels of cooperation show greater activity in the associated brain areas when playing with other people than when playing with computers, while less cooperative subjects register the same (relatively low) level of brain activity when engaged in a cooperation game with a computer as when playing the same game with a human being.[45] The studies also show brain activity associated with receiving rewards or pleasurable feedback activated in connection with human reciprocity (and suppressed when the other player defects), and not reaching the same level when a computer's "reciprocity" produces the same monetary award that human reciprocity generated.[46] Taken together, these results suggest that special satisfactions accompany reciprocity, but that not all individuals experience these satisfactions equally, and that our social functionality may vary our emotional response to reciprocity.

A small study with results published in 2002 suggested another intriguing conclusion: that people making choices with a range of specified possible outcomes are more averse to ambiguity (which characterizes a situation in which it is impossible to calculate the respective odds of various outcomes) than to risk (in which the odds of various outcomes are known).[47] Experimental subjects confronted two urns containing balls of three colors. After selecting one of the two urns, they received a cash reward

---

[44] *See* James K. Rilling, et al. *A Neural Basis for Social Cooperation*, 35 NEURON (Issue 2) 395-405 (2002) (associating activation of the orbitofrontal and ventromedial frontal cortex and anterior cingulate cortex with emotional gratification from arranging or experiencing mutually cooperative social interaction); Kevin McCabe et al., *A functional imaging study of cooperation in two-person reciprocal exchange*, 98 PROC. NAT'L ACAD. SCI. (No. 20) 11832-35 (Sept. 25, 2001) (associating parts of the prefrontal cortex with emotional gratification from cooperation).

[45] *See* McCabe et al., supra n. __ at 1833-34.

[46] *See* Rilling, et al., *supra* n. [two footnotes back]

[47] *See* Kip Smith, *et al*., *Neuronal Substrates for Choice Under Ambiguity, Risk, Gains, and Losses*, 48 MANAGEMENT SCI. (No. 6) 711-18 (2002) (associating calculating behavor under known risk with a neocortical dorsomedial system and emotionally driven aversion to uncertainty with a ventromedial system).

depending on the color of a ball drawn at random from that urn. (In some rounds, they instead selected an urn knowing that they would a cash withdrawal from an initial endowment, also in an amount depending on the color of the ball drawn.) The balls in one urn represented a wider variation in possible rewards or subtractions, while in the other the balls were designated with a narrower range of cash consequences.

Subjects played two versions of the same game. In one, they knew the number of balls of each color in each urn, and so were able to calculate the odds of each possible outcome. This was thus a situation of risk. The odds were so designed that the expected payoff from each urn was identical: therefore, the deciding factor in subjects' choice was their taste for the degree of risk, that is, for a lower chance of a higher payoff or the reverse.[48] In the other version, although the spread of possible results was known and was the same as in the first, the numbers of balls of two of three colors were unknown for the urn with the smaller spread. For the urn with the larger spread, all numbers remained known: this urn was identical in every respect with the larger-spread urn in the first version of the game. Thus, although participants knew the spread of possible results for each urn, they could not know the odds of each result only for the higher-spread urn. This was thus a situation of choice between risk and ambiguity. The experiment thus measured the difference between preference for high-spread versus low

The experiment showed that participants were much more likely to choose the lower-spread urn under conditions when facing ambiguity than under conditions of risk. That is, participants were averse to ambiguity in itself, seeking to avoid exposure to unknowable odds of either harm or benefit.[49] Moreover, in subjects facing choices that

[48] Subjects were given the alternative of expressing no preference.
[49] *See id.* at 714.

presented ambiguity, brain activity was high in regions associated with emotional response, while those making judgments wholly about risk showed activity concentrated in regions associated with calculation and "executive processes."[50] The implication is that risk engages a portion of the brain adapted to reasoning much like that modeled in economic accounts of rationality, while ambiguity engages more basic responses to threat or hazard – what the researchers define as "settings in which the competence of the decision maker is challenged."[51] That is, what we do not feel we can know and control may elicit decision-shaping responses that move us away from instrumental rationality in the direction of a kind of self-protective crouch. One system, metaphorically speaking, advances the interests of *home economicus*; the other protects the frightened animal. Depending which system a situation engages, it is a very different actor that will respond.

C. Implications for Law?

Reception of neuroeconomics by economics-oriented legal scholars has been instrumental, that is, interested in how neuroscience can help law to facilitate welfare-maximizing behavior.[52] It has also tended to suffer from the problem of mere correlation. The background argument is the widely recognized functional account of modern law as facilitating transactions among strangers at great distances, separating economic life from social and geographic proximity, which both facilitated and constrained it for many

[50] *See id.* at 716-17.
[51] *See id.* at 7111, n. 1.
[52] *See* Terrence Chorvat & Kevin McCabe, *The brain and the law*, 359 PHIL. TRANS. R. SOC. LOND. 1727-36, 1734-35 (2004); Terrence Chorvat er al., *Law and Neuroeconomics* (unpublished manuscript, available at http://ssrn.com/abstract=501063.

years.[53] A major part of this institutional achievement is to generate impersonal institutions that can substitute effectively for interpersonal trust.[54] On this account, the contribution of neuroeconomics is that economic self-interest and the reciprocity characteristic of trust have different correlates in the brain, and may well express distinct and potentially competing motivational systems. If this is true, then those who design legal systems should be alert to both the benefits of consonance between the systems and the hazards of bringing them into conflict, causing what one might think of as drag or friction, an emotionally founded reluctance to participate in economically rationally, impersonal transactions.[55]

This is not necessarily an empty contribution, but it flirts with the emptiness of mere correlation. The problem is nicely, and inadvertently, captured in a passage from Terrence Chorvat and Kevin McCabe's argument to this effect: "If [government] can frame violations of its rules *as defections from social norms*, they are more likely to be punished and often private punishment is enough to enforce this obligation. To the extent that society is able to frame defection as cheating and invoke social sanctions, enforcement will become easier."[56] The difficulty is that, if *social norms* is synonymous with whatever new information neuroeconomics has provided us, then there is no new information at all, except a new layer of correlation. It is not news, after all – although it raised some eyebrows among law-and-economics scholars in the previous decade – that social norms make a difference in people's decisions, that perceptions of "good" and

[53] For this argument, *see* HERNANDO DE SOTO, THE MYSTERY OF CAPITAL: WHY CAPITALISM SUCCEEDS IN THE WEST AND FAILS EVERYWHERE ELSE (2001); Douglas North, *Institutions*, 5 J. ECON. PERSP. 97-112 (1991).
[54] *See id.* Chorvak and McCabe rehearse this point at 1734-35.
[55] *See* Chorvak & McCabe at 1734-35.
[56] *Id.* at 1734 (emphasis added).

"bad," "fair," and "unfair" matter to legal compliance.[57] There has been a fair amount of effort to specify the appropriate understanding of the relationship between law and norms.[58] There may be a time when neuroimaging reveals something non-obvious about this relationship; but for now, to identify the contribution of neuroeconomics with the contribution of norms theory is to invite the charge of correlation, and to have no answer.

Nonetheless, one sees what the efforts at application are getting at, and it is potentially profound. We know that human beings act persistently against our interests – individual and collective. Sometimes we are sub-optimal; at other times, we are downright destructive, or worse, murderous. We know, moreover, that the worst things we do often express motives that are not evil, not essentially aimed at waste or destruction, but misplaced or misshapen. Fear is a valuable, even a necessary emotion; but we fear things we should not, whether ambiguity or nuclear waste or Tutsis, and so we make mistakes, or worse. A visceral reluctance to harm another human being may be an essential check on brutality, but it may that when harming another will save hundreds, our deepest moral intuition may be too strong, and lead us astray. In contrast, if we are considering a political decision to launch cruise missiles at a city we have never seen and whose residents we have never met, the same intuition may be too weak, and give too much play to expedience and fear. Neuroeconomics promises a glimpse of the neural systems where these intuitions, feelings, emotional and cognitive programs, operate. And this, in turn, suggests that we could better understand why we do what we do, and how we might improve. Is this right, or is it correlates all the way down?

[57] Some of the material from norms and reciprocity literatures. Note that Bob Scott, at least, has complained that norms may be observationally indistinguishable from economic rationality.
[58] More of same. Meares, Kahan, etc.

In the next Part, I propose that the feeling of promise attached to neuroeconomic observations expresses frustration with a methodological divide characteristic of twentieth-century social science, and that understanding the culs-de-sac where this divide has led can help us to discern the most important questions where neuroeconomics can make a contribution.

II. Other Minds and the Great Divide

A. The Positivist Claim

It may seem odd, in an essay on the latest methodological advances in the human sciences, to pause over the quaintly named philosophical "problem of other minds."[59] The problem with other minds, Anglo-American philosophers of the early twentieth century concluded, is that they are inaccessible to us.[60] We can make no direct observation of the subjective experience of others, nor can we share in it ourselves. This distinguishes others' subjective experience from two domains that we do observe directly: our own subjective experience, and external phenomena – that is, the world outside us, except the bits of it contained with the minds of others.[61] About our observations of our own experience, we have a fair amount of confidence: I am happy, I am in pain, I am fearful, I am elated – we observe these facts as directly as we observe

[59] For an introduction to the history of philosophical treatments of this question, *see* ANITA AVRAMIDES, OTHER MINDS (2001) (surveying the history of thought with attention to this and cognate problems); *cf.* ALEC HYSLOP, OTHER MINDS 5-6 (1995). Hyslop denies that the problem is one of observability, and distinguished between observing and experiencing inner states, arguing that the real difficulty is "an asymmetry in respect of knowledge" generated by the impossibility of experiencing what others experience. This strikes me as a helpful distinction. As will emerge later, it imposes some limits on what one might hope to gain by the observations of neuroeconomics.

[60] *See* AVRAMIDES, OTHER MINDS at 3-5 (introducing and considering the problem).

[61] *See id.*

anything.[62] The same holds of our observations of external objects: she is dancing, he is falling, they are on fire – unless our eyes are tricking us, we observe these things directly. And such direct observation has implications for knowledge.[63] Although it is possible to stake out an epistemologically skeptical position and stick to it on principle, direct observation is far and away the best candidate for knowledge: we know the things we observe, or we know very little at all.[64]

Other minds are different. Here trouble begins. How do we know that others have subjective experience like our own? When someone says, "I am in pain," how do I know that he means what I mean when I use that phrase? Or, perhaps more persuasively, when he says, "I love her," why should I believe that the experience behind the phrase is same that I observe in myself when I reply, "No, *I* love her"? What do if we then insist, simultaneously, "Well, *she loves me*"? All the evidence we have is words and actions. We can suppose that they express the same inner states that we experience in ourselves; but we cannot know.

It would entirely forgivable to feel that one had not caught the force of the "problem" so far. Is this in fact a problem, except for severe neurotics?[65] A little more

---

[62] *See id.*
[63] *See id.*

[64] Avramides suggests that skepticism about knowledge of the external world tends to be refuted in practice by everyday experience of the world, while everyday experience in fact raises and reinforces the problem of other minds. *See id.* at 2 ("The person who is asked to justify her belief that [her] cat has a mind will very soon find herself asking how she knows that other *human beings* have minds. And, similarly, the person denies minds to computers may soon find herself wondering why she is so sure there are *any* other minds").

[65] There is a view, with impressive philosophical pedigree, that this is just the right characterization of an obsession with the problem of other minds. *See* STANLEY CAVELL: THE CLAIM OF REASON:WITTGENSTEIN, SKEPTICISM, MORALITY AND TRAGEDY at 463 (suggesting the clinical concept of narcissism usefully diagnoses the perception of oneself as unknowable to others in some essential way, and also reveals an underlying and frustrated wish to be known) 468-69 (describing a productive history of the problem of other minds as requiring " an account of the particular insanity required, or caused, or threatened, in the very conceiving of the problem") (1979).

historical texture may be of some help. The problem of other minds took wind to its sails from Logical Positivism, a philosophical movement that flourished, particularly in Austria and England, in the early decades of the twentieth century.[66] The ambition of Logical Positivism was to distinguish between propositions whose truth or falsehood could be determined and others that lacked truth-value and were in fact expressed only conceptual confusion or emotional assertion masquerading as something more.[67] One category of statements that notoriously lacked truth-value in the Logical Positivist view was moral claims: good and bad, right and wrong, were dressed up assertions of preference, and best treated as such.[68] Another such category was statements involving alleged facts about which, in principle, no reliable evidence could be obtained – such as statements about the nature of God, or about other minds.[69] Claims about the contents of other people's minds were exemplary of the kinds of confusion of prejudice or wishful thinking with knowledge-claims that the Logical Positivists sought to banish from respectable inquiry.[70]

As David Grewal has recently argued, the advent of the now-dominant economic model of behavior was directly linked with the intellectual culture of Logical Positivism in general, and with the problem of other minds in particular. Lionel Robbins, the English economist who contributed much to bringing about the so-called ordinalist revolution in Anglo-American economics, founded an important part of his case on the

---

[66] For a very helpful introduction to the ambitions and method of logical positivism, *see* A.J. Ayer, *Editor's Introduction in* LOGICAL POSITIVISM 3-28 (A.J. Ayer, ed.) (1959). Like nearly any school or body of thought, logical positivism encompassed a variety of inflections and emphases around its core commitments.

[67] *See id.* at 17-23 (emphasizing the logical positivist mistrust of claiming truth, verifiability, or falsifiability for statements about value or about the mental states of others).

[68] *See id.* at 21-23.

[69] *See id.* at 17-21.

[70] *See id.* at 17 ("There could be no question of our literally sharing one another's sense-data, any more than we can literally share one another's thoughts or images or feelings").

impossibility of knowing the utility levels of others.[71] Robbins's motivation was to purify economics as a positive science by purging it of knowledge-claims that could not pass the Logical Positivist bar.[72] Although an ordinalist approach to economics had found earlier adherent in England, and on similar philosophical ground, and likewise on the Continent, Robbins's arguments were decisive in the ordinalist triumph in Anglo-American economics.[73] This was particularly pressing because much of English economics into the 1930s was classically utilitarian; that is, it sought to advise policymakers on the best means of maximizing the utility of the nation, summed across persons.[74] Such an effort necessarily involves hypotheses about the relative utility of

---

[71] *See* Robert Cooter & Peter Rappoport, *Were the Ordinalists Wrong about Welfare Economics?* 22 J. ECON. LIT. 507, 520-24 (1982) (recounting Robbins's critique, on these grounds, of the ambitions to utility measurement of earlier modes of English economics). Robbins pioneered his critique of previous economics theory in LIONEL ROBBINS, AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC THEORY 136-58 (summarizing his attack on the concept of cardinal measures of utility and insisting, *inter alia*, on the importance of value-relativism in setting the limits of economic inquiry). This also comports with the capsule history of ordinalism's rise given by Amartya Sen. *See* AMARTYA K. SEN, *The Possibility of Social Choice*, *in* RATIONALITY AND FREEDOM 65-118, 71 (2002) ("[E]conomists came to be persuaded by arguments presented by Lionel Robbins and others (deeply influenced by 'logical positivist' philosophy) that interpersonal comparisons of utility had no scientific basis: 'Every mind is inscrutable to every other mind and no common denominator of feelings is possible.' Thus, the epistemic foundations of utilitarian welfare economics were seen as incurably defective") (quoting Robbins).

[72] *See* Cooter & Rappoport, *Ordinalists* at 522 (recounting the influence and positivism in Robbins's circle and concluding that "Robbins went a long way in the positivist direction of excluding ethical and mental concepts from science").

[73] Most significant among these was William Stanley Jevons, a nineteenth-century theorist, relatively marginal in his day, to whom Robbins looked as an indispensable predecessor. Philosophically, Jevons was a utilitarian skeptic: he believed that subjective pleasure, or "psychological hedonism," was the proper measure of well-being, but that there was no way to unify the measurement of pleasure across persons because satisfactions were incorrigibly idiosyncratic. *See* LIONEL ROBBINS, *The Place of Jevons in the History of Economic Thought*, *in* THE EVOLUTION OF MODERN ECONOMIC THEORY AND OTHER PAPERS ON THE HISTORY OF ECONOMIC THOUGHT 169-88 (1970) (recounting Jevons's thought). Robbins argued that this dimension of Jevons's thought was unnecessary, but that its skeptical impulse, purified of its other commitments, laid a proper foundation for scientific economic inquiry. *See* ROBBINS, NATURE AND SIGNIFICANCE at 85 ("the hedonistic trimmings of the works of Jevons and his followers were incidental to the main structure of a theory which – as the parallel development in Vienna showed – is capable of being set out and defended in absolutely non-hedonistic terms"). Cooter and Rapport agree with this characterization of Robbins's contribution, concluding, "Jevons used the subjective definition and remarked that there is no compelling way to compare the pleasures of different people. Robbins merely embedded this familiar claim in positivist philosophy." Cooter & Rappoport, *Ordinalists* at 522.

[74] For a hostile characterization of this approach, *see* ROBBINS, NATURE AND SIGNIFICANCE at 4-11 (describing the "materialist" conception of economics, which attempted empirical measurements of well-being). For a friendlier description, *see* Cooter & Rappoport, *Ordinalists* at 512-20 (describing the method

different persons, which English utilitarianism accomplished chiefly through the assumption of declining marginal utility, that is, that with increasing wealth, the gain in utility from each new sum is less than at lower levels of wealth.[75] But the claim to know that one person is experiencing a greater level of satisfaction than another concerns the subjective experience of another – a cardinal instance of wishful thinking or idle speculation masquerading as a claim to truth. Mixing such speculation with economic science was like mixing morality with mathematics.

Robbins made precisely this argument in his famous *Essay on the Nature and Significance of Economic Science*,[76] where he launched his argument against the then-prevalent English fusion of utilitarian ethics and economic analysis:

---

and ambitions of the "material welfare school" of English economics). It is abundantly clear that the ambitions of the material welfare school reappear in the contemporary work of Amartya Sen, who has proposed a version of welfare economics attentive to the measurement of "capabilities," that is, the range of human potentials that individuals are able to realize in the activity. This is a more nuanced and open-ended conception than the emphasis on physical functioning that informed the classical materialist approach; nonetheless, it proposes that by understanding how resources matter to human beings by enabling them to function in their lives, it is possible to say something useful about the relationship between distribution of resources and aggregate well-being – that is, it is possible to make meaningful, if not entirely precise, comparisons of welfare across persons. Sen has developed this position in many essays. *See Goods and People*, at 509 in *Resources, Value, and Development*, *supra* n. 66; *Markets and Freedoms*, *in* RATIONALITY AND FREEDOM 501 (2003); *Opportunities and Freedoms*, *in* RATIONALITY AND FREEDOM 583; *Freedom and the Evaluation of Opportunity* , *in* RATIONALITY AND FREEDOM 659; and *passim* in AMARTYA SEN, DEVELOPMENT AS FREEDOM (1999).

[75] Robbins accordingly made an attack on the "scientific" status of the hypothesis of declining marginal utility central to his arguments. *See* ROBBINS, NATURE AND SIGNIFICANCE at 137-38 ("The Law of Diminishing Marginal Utility … is derived from the conception of a scarcity of goods in relation to the ends which they serve. It assumes that, for each individual, goods can be ranged in order of their significance for conduct; and that, in the sense that it will be preferred, we can say that one use of a good is more important than another. Proceeding on this basis, we can compare the order in which one individual may be supposed to prefer certain alternatives with the order in which they are preferred by another individual. … But it is one thing to assume that scales can be drawn up showing the *order* in which an individual will prefer a series of alternatives. It is quite a different thing to assume that behind such arrangements lie magnitudes which themselves can be compared. This is not an assumption which needs anywhere be made in modern economic analysis, and it is an assumption which is of an entirely different kind from the assumption of individual scales of relative valuation"). As this somewhat long excerpt shows, rejection of the assumption of diminishing marginal utility was lodged at the heart of Robbins's argument about the non-comparability of personal utility levels.

[76] LIONEL ROBBINS, AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC SCIENCE (1935).

> Now, of course, in daily life we do continually assume that the [interpersonal] comparisons can be made. But the very diversity of the assumptions actually made at different times and in different places is evidence of their conventional nature. In Western democracies we assume for certain purposes that men in similar circumstances are capable of equal satisfactions. Just as for purposes of justice we assume equality of responsibility in similar situations as between legal subjects, so for purposes of public finance we agree to assume equality of capacity for experiencing satisfaction from equal incomes in similar circumstances as between economic subjects. But although it may be convenient to assume this, there is no way of proving that the assumption rests on ascertainable fact. And, indeed, if the representative of some other civilization were to assure us that we were wrong, that members of his caste (or his race) were capable of experiencing ten times as much satisfaction from given incomes as members of an inferior caste (or an "inferior" race), we could not refute him. We might poke fun at him. We might flare up with indignation, and say that his valuation was hateful, that it led to civil strife, unhappiness, unjust privilege, and so on and so forth. But we could not show that he wrong in any objective sense, more than we could show that we were right.[77]

In an essay a few years later, Robbins again used the example of cultural variation in assumptions about utility, evoking a Brahmin who purports to be capable of ten times as much happiness as an untouchable. He concluded,

> I had no sympathy with the Brahmin. But I could not escape the conviction that, if I chose to regard men as equally capable of satisfaction and he to regard them as differing according to a hierarchical schedule, the difference between us was not one that could be resolved by the same methods as were available in other fields of social judgment.[78]

In Robbins's telling, the problem of other minds not only scuttles the specific claim that people experience equal utility under like circumstances: it makes untenable any interpersonal comparison of utilities. This is so because claims about the utility of others are neither falsifiable nor verifiable: in the face of disagreement, arguers can have

---

[77] *Id.* at 140. I am grateful to David Grewal for drawing my attention to this passage, which he uses in making the same argument about Robbins: that his role in the ordinalist revolution was inspired by the problem of other minds. *See* Grewal, *supra* n. __.

[78] Lionel Robbins, *Interpersonal Comparisons of Utility: A Comment*, 48 ECON. J. 635-41, 636 (1938). Again, David Grewal showed me this remark of Robbins's and has given substantially the same analysis of it.

no recourse to facts; they can only reiterate their judgments more emphatically. As Robbins wrote, they can "flare up," call the opposing view "hateful," or claim that holding it will have undesirable results for the social world; but they cannot show that it is false, any more than the opponent can show that it is true. Such a claim has no place in a scientific method of social inquiry.[79]

The essential connection between Robbins's conclusion and the method of contemporary economics is that the ordinalist account of rationality and utility can provide a model of decision-making that avoids any claim to compare the utility of various persons. Instead, the microeconomic model that Robbins helped to inaugurate avoided any claims to observe or make substantive inferences about the mental states of others. Instead it assumes that (1) individuals have introspective knowledge of their own preferences, so they are able to evaluate choices according to the levels of satisfaction they will bring about *for them*; (2) the decisions they make, which are objectively observable, will be consistent with, and thus reveal, their preferences; and (3) the observed pattern of decisions will consequently be utility-maximizing for each person. Note that these claims allow no inference about whether one person's utility in a given state of affairs is greater or less than another person's in the same (or, for that matter, a different) state of affairs. The only kinds of knowledge this approach requires – or permits – are introspective and externally observable, which it bridges by means of the

[79] Naturally it is entirely possible to state a normative principle that people shall be regarded as experiencing like utility under like circumstances, and to make judgments about public policy on that basis; but this is not social inquiry or explanation. Treating interpersonal comparisons of utility in this manner gives them exactly the same status that the logical positivist C.L. Stevenson gave to ethical judgments. He argued that they made sense only as conventional agreements to name certain objective states of the world "bad" or "good" and to draw imperatives of action from those conventional descriptions. He denied utterly that the stipulation of "bad" or "good" was itself subject to reason evaluation. *See* C.L. Stevenson, *The Emotive Meaning of Ethical Terms*, *in* LOGICAL POSITIVISM 264-81 (A.J. Ayer, ed.) (1959). Disagreements as to this issue admitted of only the kinds of tantrums that Robbins suggested might follow from quarrels over interpersonal utility.

assumption that introspective knowledge of one's own utility function translates into objectively observable decisions.

It should be evident that this version of economic method has at least an elective affinity with non-utilitarian modes of welfare economics, particularly Pareto's conception of efficiency, which requires only non-interference with individual utility levels, and includes no notion of aggregating utility, which would require an interpersonal metric of utility.[80] My interest here, though, is in a different point: the connection between an economic method that avoids the problem of other minds and the general conception of scientific social inquiry that arose with logical positivism and took on board its conception of what could constitute knowledge. The microeconomic model of decision-making that I have just sketched built into its premises the sort of epistemic chastity that this model of social inquiry celebrated, and thus provided a tool-kit well adapted to the formulation of decorous questions and chaste answers.

In a fine polemical statement of this methodological ambition, the logical positivist Otto Neurath declared that social inquiry must be subject to the same standard as any other inquiry: creating a unified set of generalizations "equal to the task of serving, as often as possible, to *predict* individual events or certain groups of events."[81] A method of social inquiry geared to these criteria must contain no meaningless statements, which would at best obscure and, far more often, outright confuse the inquiry. Neurath characterized any claim to knowledge of the self-understanding of others as meaningless in this sense, that is, not subject to objective assessment and thus adding other confusion

[80] This is the main argument of Grewal, *supra*. *See also* SEN, *Social Choice* at 71-72 (recounting the development in which the rejection of interpersonal comparisons of utility, by reducing "the informational base on which any social choice could draw" cleared the way for "a so-called 'new welfare economics,' which used only one basic criterion of social improvement, viz, the 'Pareto Comparison").
[81] Otoo Neurath, *Sociology and Physicalism*, *in* AYER, *supra* n. __ at 282-320, 293 (italics original).

or nothing at all. Methods that claimed a place for "empathy" or "understanding" of the motives or ideas on which people act came in for special scorn.[82] Like claims about the utility of others, or ethical statements, claims resting on "empathy" or "understanding" could be neither verified nor falsified, because they did not refer to a realm of phenomena to which inquirers had access. They might chance to be true, as claims about the utility of others or about the nature of God might chance to be true; but there was no building a science out of such shots in the dark. All social inquiry must be behaviorist, that is, concerned only with the observable activity of individuals and groups.

Readers will recognize the connection between Neurath's methodological polemic and the famous later claim of Karl Popper, himself deeply influenced by logical positivism, that social inquiry must produce falsifiable predictions or else be subject to endless and epistemically fruitless manipulation.[83] They will also recognize the connection with the economics-influenced model of social explanation familiar from law and economics, rational-choice theory, and economically influenced sociology: a simple model of individual decision-making, involving no claims about the mental states of others, which produces testable hypotheses about the behavior of individuals and institutions given certain circumstances.

B. The Interpretivist Response

The positivist model of social explanation has not overwhelmed all alternatives. Instead, it has produced a schism in social inquiry between positivists and interpretivists, who maintain what Neurath rejected: that an adequate account of human activity requires

[82] *See id.* at 296-97.
[83] *See* KARL POPPER, THE OPEN SOCIETY AND ITS ENEMIES (1945).

attention to the self-understanding of others. The most rigorous and articulate exponent of this view is the philosopher and law professor Charles Taylor. In a decades-long series of arguments, Taylor has developed a position that may be stated in the following propositions. (1) Human beings are self-conscious animals, aware of making decisions from among alternatives.[84] (2) We experience ourselves as making these decisions not willy-nilly, but for reasons: for instance, we choose an action because we think it is right, or better than the alternative, or more dignified, or more consistent with who we think we are; we avoid or regret an action that we regard as wrong, worse, asinine, or out of character (unless we think our character needs amending!).[85] (3) Consequently, we always act under descriptions or interpretations of our actions and the context in which they take place. Our lives never present themselves to us as raw fact, but always as interpreted activity.[86] (4) These descriptions are not idiosyncratic, but deeply shaped by shared ideas about human beings, the social world, and the natural world. It is only shared understandings of this sort that make certain objective actions into "performing mime theater," "hitting on someone," "coming out of the closet," "negotiating a contract," or "being a Trappist Monk."[87] Try performing mime or coming out of the closet in, say, Homeric Greece, and you will (or would, if the experiment were literally possible) be lucky to get away with merely the kind of mutual incomprehension and irritation that Robbins envisioned with his hypothetical Brahmin. (5) As the last point

---

[84] *See* CHARLES TAYLOR, *What Is Human Agency? in* 1 PHILOSOPHICAL PAPERS: HUMAN AGENCY AND LANGUAGE 15-44 (so arguing); *Self-Interpreting Animals*, *id.* at 45-76 (same) (1985).
[85] *See* CHARLES TAYLOR, SOURCES OF THE SELF: THE MAKING OF THE MODERN IDENTITY 25-52 (1989) (describing the relationship between self-conscious and self-interpreting agents and the questions of value they confront).
[86] *See id.* (so arguing).
[87] *See id.* at 29-42; TAYLOR, *Interpretation and the Sciences of Man*, *in* 2 PHILOSOPHICAL PAPERS: PHILOSOPHY AND THE HUMAN SCIENCES 15-57 (1985) (so arguing, with particular reference to the concept and practice of negotiation).

indicates, shared understandings differ across time and place. Actions that are altogether intelligible in one setting, even regarded as essentially human there (such as striking a contract or publicly acknowledging one's sexuality), will not make sense in another setting. Indeed, for most people in that second setting, even the idea of the action may be inaccessible: can we really know what it meant to be one of the medieval Christians who sent their young to Jerusalem in the Children's Crusade? (6) Consequently, there is a relevant sense in which people, unlike any other object of inquiry, do not obey the same rules across time and space. They will react differently to the same external stimuli, because those stimuli will have different meaning for them, depending who they are.[88] Of course, they will display behavioral regularities within a cultural practice that they belong to, for that is what makes a cultural practice; but to explain what they are doing, you will need reference to the practice and the shared understandings that make it intelligible to its participants.[89] If you try to abstract from those considerations, you will get two results. First is an account that presents correlations between stimuli and behavior without reference to the underpinnings of the activity, which are not simply physical facts, but also self-understandings shaped by participation in shared interpretations.[90] Of course, you may be willing to jettison self-understandings, as long as physical facts suffice to fill out the correlations. But then the problem is the second

[88] *See* TAYLOR, *Human Agency* at 43 ("if we take the view that man is a self-interpreting animal, then we will accept that a study of personality which tries to proceed in terms of general traits alone will have only limited value"); *Interpretation* at 53 ("[T]he … most fundamental reason for the impossibility of hard prediction [of human action] is that man is a self-defining animal. With changes in his self-definition go changes in what man is, such that he has to be understood in different terms. But the conceptual mutations in human history can be and frequently do produce conceptual webs which are incommensurable, that is, where the terms cannot be defined in relation to a common stratum of expression").

[89] *See* TAYLOR, *Interpretation* at 55 ("The entirely different notions of bargaining in our society and in some primitive ones provide an example. Each will be glossed in terms of practices, institutions, ideas in each society which have nothing corresponding to them in the other").

[90] *See id.* at 38 ("We can allow, once we accept a certain set of institutions or practices as our starting point … that we can easily take as brute data that certain acts are judged to take place in certain states").

feature of the result: an account whose validity stops at the border of the practice in which the correlation occurs, whether in time or in space.[91]

Let us say, then, that you want to answer a social question with real stakes: Why did nationalist and fascist ideologies display wide appeal in the twentieth century? Correlatively, how can we tell whether, and if so where, such ideologies present a threat to freedom and tolerance today? Taylor's argument implies that trying to answer such a question by reference solely to objective facts – the incentives the constitutional structure give to politicians, the trajectory of the national economy, the historical proximity of military defeat or colonial domination – will not take us as far as we should want to go. The same logic applies to another urgent contemporary question: Why do certain populations produce suicide bombers while others – also poor, on the losing end of recent history, and beset by intractable political orders – do not? In providing such explanations and predictions, Taylor suggests that we will want to supplement our positivism, at least, with an attempt to understand how the ideological appeal of fascism or martyrdom offers a new collective understanding – of national identity, for instance, and of the nation's history – and recruits followers into it. That is, we will want to make a trip to the heart of shared understandings, of "empathy," in Neurath's dismissive term.

The difficulty is that, even though the positivist program is unsatisfactory, its criticism of the interpretive mode is potent. That the minds of others are not directly available to us is not a niggling point, but a basic impediment to judging whether any particular inquiry is elucidating or obscuring the object of study. In beginning an interpretive inquiry, one must make methodological choices, and those involve

[91] *See id.* at 55 (on the consequences of incommensurability of interpretations and practices across time and space).

presuppositions about the content of other minds. The method may be psychoanalytic, seeking the sources of violent ideologies in suppressed anger and distorted sexual desire.[92] It may be Marxist, involving claims about the relationship of ideology to economic position.[93] It may, in the manner of Taylor himself, simply an attempt to give a rich description of the shared understandings that arise and are articulated around particular political movements.[94] Or, taking a cue from Tocqueville, it may put substantial trust in the acuity and synthetic power of the observer, inviting her to put together a story about ideological motivation out of disparate observations and interpretive intuitions.[95]

How, then, to say which approach is helpful and which misleading? There is no fully satisfactory answer. Each of these approaches can purport to interpret a given episode consistent with its methodological commitments. Moreover, the predictions they issue will tend to incorporate those commitments so as to be non-falsifiable: for instance, "I predict that violent ideologies will have appeal to populations where suppression of anger and shame at sexuality result in sado-masochistic impulses directed at out-groups." That may well be right; but the statement does not come in testable form. The non-

---

[92] *See*, *e.g.*, MICHAEL IGNATIEFF, THE LESSER EVIL: POLITICAL ETHICS FOR AN AGE OF TERROR 126 (associating nihilistic and fanatical political violence with the psychic sources) (2004).

[93] *See*, *e.g.*, ASHOW DHAWALE, THE SHIV SENA: SEMI-FASCISM IN ACTION 64-69 (2000) (applying Marxist class analysis the political success of Hindu nationalist party in Bombay); RAYMOND WILLIAMS, THE COUNTRY AND THE CITY 96-107 and *passim* (1973) (describing transformation in rural property relations and ideas of rural life in broadly Marxist terms).

[94] *See*, *e.g.*, TAYLOR, SOURCES, *supra* n. _ at 393-418 (describing the worldview and idea of personhood that, on his account, gave impetus to both humanitarian political movements and nationalism in the nineteenth century).

[95] For the cardinal instance of the type, *see* ALEXIS DE TOCQUEVILLE: DEMOCRACY IN AMERICA (trans. George Lawrence, ed. J.P. Mayer, Perennial Library 1988) (1850). A relatively recent work of interpretive sociology proceeding frankly in this vein is ROBERT BELLAH, et al., HABITS OF THE HEART: INDIVIDUALISM AND COMMITMENT IN AMERICAN LIFE (1996) (arguing, on the basis of a synthetic interpretation drawn from many individual conversations, that American ideas of the dignified and meaningful life are difficult to reconcile with any rich practice of community and commitment, and thus tend to be self-undermining in practice).

appearance of violent ideologies will simply be evidence of the relative weakness of sado-masochism in politics. To know more, we would have to do what we cannot: have access to the mental states of others, to "see" the descriptions under which they reason and act.

This is the origin of the wish that neuroeconomics addresses: the wish to see other minds, and so to heal the rift in social inquiry, restricting ourselves to observable phenomena while acknowledging the importance of the descriptions under which people act. In the next part of this essay, I treat the promise and the limits of the answer neuroeconomics gives to this wish.

III. Three Problems for Neuroeconomics

A. Collective Action: Reciprocity, Rationality, or rationality?

1. Collective-action Problems

Legal and public-policy analysis have given attention for decades to "collective-action problems." The structure of the problems is that individuals acting on the neoclassical model of rationality will find it in their interest not to contribute to arrangements that benefit the group overall. Canonical examples include the Prisoner's Dilemma, where each player's pursuit of a dominant (always rationally preferred) strategy yields a sub-optimal outcome; the Voter Paradox, where each individual's good reason not to participate in a collective practice causes the practice to break down; and the Tragedy of the Commons, where individual pursuit of maximum gain from a common resource results in exhaustion of the resource.

As Amartya Sen has phrased the matter helpfully – albeit at a characteristically high level of abstraction – in a way that captures its two-sided implications: "Game-theoretic analyses have contributed to a better understanding of some of the difficulties that the concept of 'rationality' must face and have clarified the nature of some problems that social organization must deal with."[96] Put differently, collective-action problems may present at least two kinds of difficulty for social practices and institutions. First, they may present prudential problems: it might be unwise to introduce or maintain a resource-governance regime with the structure of a commons tragedy. Second, collective-action problems may present a theoretical difficulty: if it is true that actions contributing to social optimality are individually irrational, those who seek to understand social practices and institutions may have to revise their understanding of their topic. Here, however, the other side of collective-action problems enters. It would be a basic mistake to imagine that once a theoretical account of rationality has "proven" the "irrationality" of a practice that is in fact widespread, such as voting, theorists and voters alike must either give up the practice or go in search of "non-rational" explanations. In the natural sciences, observed data is decisive: a result inconsistent with a theory falsifies the theory. In theoretical social inquiry, we owe facts at least the decency to contemplate that their persistence might suggest a problem in a theory that purports to debunk them.

The situation calls for a middle way, which neither fetishizes existing practices nor hypostasizes a particular conception of rationality. Responses have taken many tacks, from Robert Axelrod's argument that reiterated games produce reciprocity among

[96] *See* AMARTYA K. SEN, *Goals, Commitment, and Identity*, *in* RATIONALITY AND FREEDOM 206-224 (207) (2002).

self-interested individuals[97] to Carol Rose's contention that a plurality of motives is necessary to explain the emergence of cooperative practices[98] to James Acheson's studies of actually existing common resources and the management practices that accompany them.[99] Here I concentrate on two, Dan Kahan's proposal that reciprocity is at least as motivationally important as self-interest narrowly understood, and Richard Tuck's argument (developing some themes of Sen's) that a conception of rationality that emphasizes agency over welfare would move us nearer the experience of judgment and choice.[100]

2. Reciprocity as Sympathy: Revising welfare

Kahan's argument nicely summarizes a strong version of the sorts of behavioral economics inquiries sketched in Part I in connection with neuroimaging experiments. He assets that according to "the reciprocity theory" of motivation to collective action, (1) "Most people think of themselves and want to be understood by others as cooperative and trustworthy and are thus willing to contribute their fair share to securing collective goods. By the same token, most individuals loathe being taken advantage of … if they perceive that most other individuals are shirking, they too hold back to avoid feeling (or being) exploited."[101] (2) Either reciprocity or non-reciprocity can characterize a social state: "there is no 'dominant' strategy … interdependencies [of reciprocity-based motivation] tend to generate patterns of collective behavior characterized by multiple equilibria

[97] *See* ROBERT AXELROD, THE EVOLUTION OF COOPERATION 3-35 (1984).
[98] *See* Carol M. Rose, *Property as Storytelling: Perspectives from Game Theory, Narrative Theory, Feminist Theory*, 2 YALE J. L & HUM. 37 (1990).
[99] *See* JAMES M. ACHESON, THE LOBSTER GANGS OF MAINE 48-50, 73-76 (1988).
[100] *See* Dan M. Kahan, *The Logic of Reciprocity: Trust, Collective Action, and Law*, 102 MICH. L. REV. 71 (2003); RICHARD TUCK, FREE RIDING (unpublished manuscript, on file with author) (2005).
[101] *Id.* at 74.

punctuated by tipping points" between reciprocal and non-reciprocal equilibria.[102] (3) In light of (1) and (2), policymakers should seek to promote contribution to collective goods not just by creating positive economic incentives, but also by promoting trust, leading people to believe that others will contribute, and thus to contribute themselves, in the virtuous circle of a high-reciprocity equilibrium.[103] Moreover (4), individual motivation to reciprocate or withhold contributions both varies across individuals at any time and varies with dynamic interactions between individuals' present motives and what they perceive to be the motives and actions of relevant others.[104]

How should we understand Kahan's "reciprocity view"? Just what kind of modification does it propose to the conventional understanding of collective-action problems? The unifying idea of the proposal is that people's motivations are neither as fixed nor as self-regarding as neoclassical accounts would have it. Such a modification, however, may tell in any of several different respects, as Sen outlines a discussion of "several quite distinct components of 'privateness' in the concept of persons used in standard economic theory."[105] Conventional economic analysis on Sen's account takes on board three aspects of "motivation," to use Kahan's term, and interprets each in a "private" way, to use Sen's word. First, on the level of *welfare*, it assumes that a person's "welfare depends only on his or her own consumption (and in particular, it does not involve any sympathy or antipathy toward others).[106] Second, on the level of *goals*, it assumes that "[a] person's only goal is to maximize his or her own welfare … (and in

[102] *Id.* at 75.
[103] *See id.* at 76-77.
[104] *See id.* at 78-79.
[105] *See* SEN, *supra* n. __ at 213.
[106] *Id.*

particular, it does not involve directly attaching importance to the welfare of others)."[107] Third, on the level of *decision* or *action*, "[e]ach act of choice of a person is guided immediately by the pursuit of one's own goal (and, in particular, it is not restrained by the recognition of other people's pursuit of their goals)."[108]

Kahan's proposal operates on the first level, contending that reciprocity plays a major role in constituting welfare, or utility. It would be possible to reverse-engineer the formulation, positing that welfare is function of satisfied preferences or achieved goals, and thus that reciprocity must be one's goal; and that preferences in turn are read from choices ("revealed preferences"); but this would be artificial. The claim is that most people's experience of welfare is interpersonally dependent: we feel good when others treat us reciprocally and regard us as taking the same attitude toward them. This interpretation thus involves no modification of conventional claim (2) or (3), which make goals and choices functions of welfarew: those can proceed undisturbed once Kahan has specified the substantive content of welfare. Kahan's quarrel, then, is not with the formal structure of rationality that undergirds the theory of collective-action failure, but with the assumption that the source of the welfare that individuals seek to maximize excludes the one quality that could make the collective-action problems self solving: reciprocity itself.

3. Reciprocity as Agency: Revising "rationality"

In his discussion of collective-action problems, Sen proposes a different modification, addressed less to the experience of welfare than to the character of agency. Sen aptly identifies Kahan's type of interdependent personal welfare as "sympathy,"

[107] *Id.*
[108] *Id.* at 214.

which "refers to one person's welfare being affected by the position of others," meaning here not their welfare, but their attitude of reciprocity or non-reciprocity.[109] While acknowledging the possibility of this modification, Sen evinces a great interest in what he calls "commitment," which involves conceiving of choice as founded not exclusively in individual welfare (whether imagined as including or excluding sympathy), but also on "identity," or "our view of ourselves … the way we view our welfare, goals, or behavioral obligations."[110] Sen's objection to the model of rationality that generates theoretical collective-action problems is that they are too scanty in their conception of rationality: they undervalue the importance to choice of our capacity for self-scrutiny, reflection, and context spanning commitment.[111] These processes of reasoning, in turn, necessarily reflect the substance of the person's *identity*, the groups or institutions or traditions with which she aligns herself.[112] The weight of Sen's modification is not on the weigh experience registers on the person as a source of welfare, but rather on what satisfactions, commitments, principles, or other aspects of her identity the agent reflectively affirms as reason for action. Choice is thus not a function of welfare, but the independent fulcrum of reflective action, responsive to welfare, but reflectively and not compulsorily so.

---

[109] *Id*. Using the term in this way involves a modification of Sen's meaning, as he intends "position" to refer to others' level of welfare.
[110] *Id.* at 214-15.
[111] As Sen puts it, "A person is not only an entity that can enjoy one's own consumption, experience and appreciate one's welfare, and have one's goals, but also an entity that can examine one's values and objectives *and choose in the light of those values and objectives*." AMARTYA K. SEN, *Rationality and Freedom in* RATIONALITY AND FREEDOM 3-64, 36 (2002).
[112] *See Goals* at 215. As Sen puts it, "A person's concept of his own welfare can be influenced by the position of others in ways that may go well beyond 'sympathizing' with others and may actually involve identifying with them. Similarly, in arriving at goals, a person's sense of identity may well be quite central. And, perhaps most important in the context of the present discussions, the pursuit of private goals may well be compromised by the consideration of the goals of others in the group with whom the person has some sense of identity."

One consequence of Sen's account is that many different rankings of priorities, including many different relative weightings of personal welfare, may be consistent with rationality. Where the neoclassical model produces dominant strategies with suboptimal consequences, Sen contends, there is simply nothing authoritative about the account of rationality that generates those strategies. It is just as cogent to say that one rationally chooses not to defect from a collective enterprise because of the value one places on one's own future well-being, the aggregate long-term well-being of those who share in the enterprise (such as other citizens), or the principles the enterprise seeks to bring about (for example, general prosperity or political liberty).[113] It may be a stretch to insist that such considerations are all constituents of "welfare," unless the move is purely axiomatic, i.e., whatever prefer we reveal by choice must be welfare-maximizing. It is not, however, a stretch to say that these are the considerations one has reflectively adopted and pursued.

This way of putting the argument may seem miss the force of collective-action problems. It is all very well to say that people prefer states of the world in which joint enterprises can flourish and the right principles (that is, the principles these people reflectively endorse) govern. But why not, at any moment, let other people do the work, or take a little extra for oneself? And if any one person should be susceptible to that temptation, why not everyone? And so forth down the familiar path until we are all non-voters imprisoned for medium-length terms amidst pastures exhausted by overgrazing.

Sen does not fully develop an answer, but his response contains the cornerstones of two lines of argument. One is that identification with a relevant group of others *means* accepting certain rules of conduct toward others, and *not* engaging in relentless

[113] *See* SEN, *Goals* at 212. As Sen puts it here, "The point is not that rationality must take us to the communal principle, rejecting the individualistic one, but that there is a genuine ambiguity here about what rationality might require[.]"

calculation toward self-interested maximization.[114] Self-interest bounded by such rules is, of course, entirely ordinary; but self-interested calculation that persistently disregards such rules includes a failure or refusal to identify with others.[115] To say that rationality includes the substantive commitments generated by interpersonal and group identification, then, is to say that it includes a valuation of shared rules as things to be obeyed.

A second line of argument is more thinly developed in Sen, but provocative nonetheless. This has to do with the relationship between agency and identity. As Sen stresses, "'We' demand things; 'our' actions reflect 'our' concerns; 'we' protest at injustice done to 'us.''[116] That is, in speech and writing people habitually ascribe agency to collectives with which they identify. Although that form of speech might be mere rhetorical license, it also might indicate something about the experience of agency. If I vote the winning candidate in a presidential campaign where I care strongly about the outcome, and the candidate wins by more than one vote, do I experience myself as having committed a superfluous act of no significance, or as having a share in bringing about a victory? Introspection suggests that the answer is the latter, and that having no such experience – feeling the act of voting was entirely inefficacious – would indicate that one had already disidentified from the partisan or civic identity that made sense of the act of voting. Moreover, reflection suggests that the experience of agency in such collective accomplishments is not the fraction of effort that one's contribution made up, e.g., one

[114] *See id.* at 216-17.
[115] Sen's formulation connects his thought with that of Alexis de Tocqueville, who regarded "self-interest properly understood" as the motivation that would checked individualism and permitted widespread social cooperation in the United States and, potentially, in democratic, individualistic societies generally. As Sen points out, he draws his emphasis on the importance of customary rules in part from Adam Smith, who stressed their "great use in correcting misrepresentations of self-love concerning what is fit and proper to be done in our particular situation." *See* SEN, *Goals* at 217 (quoting Smith).
[116] *Id.* at 215.

forty-millionth part of a victory. Consider a winning baseball team. Although individual contributions are relatively easy to tally in that sport, a player whose contribution was "non-necessary" to the win in the sense that it did not provide the margin of victory would be confused, maybe even narcissistic, if he decided the win therefore had nothing to do with his participation. Moreover, each player's experience of achievement would not one-ninth of some sum of achievement, ascribed to a collective agent called "the team." There is no such sum. The experience of agency occurs only in individuals, and each individual has the experience expressed in "we won," an experience that is possible only because of each individual's identification with the team.

On this account, then, an act of voting would be an act of agency dependent on identification with the party, movement, or principles of the candidate. A victory means that the voter's experience is "we won," an undivided experience of achievement that by its nature belongs to each voter who contributed. The intensity of the experience will vary with the magnitude of the achievement (whether the race was hard-fought, how deep and consequential the difference of principle was) and the level of the voter's identification with the party, movement, or principle. It will not vary with whether the voter's contribution provided the margin of victory, nor will it be divided by the number of contributing voters. Such restrictions are in the nature of purely self-regarding agency, but not in the nature of agency premised on identification with a collective.

4. The Question for Neuroeconomics

This discussion leaves us with several alternative accounts of why neoclassical models of rationality frequently fail to predict the outcomes of collective-action

problems. One, Kahan's is based on the satisfaction people take in viewing themselves and others as reciprocators and in activity that confirms that view. Another, the first of Sen's alternatives, is that group identification produces loyalty to common rules of conduct, and that loyalty to these is part of the nature of group membership. A third is that group identification changes the experience of agency, producing an experience of undivided achievement based in accomplishments conventionally ascribed to the group.

Of course, these accounts need not be mutually exclusive. It is possible that one, two, all, or none of them is right, and likely that they account for different degrees of motivation across persons, contexts, and time. Neuroimaging in prisoners' dilemma-type games, as discussed above, hints that when the decisionmaker experiences reciprocity or non-reciprocity with another concrete person (not necessarily visible, but believed to be the source of decisions), some kind of spontaneous experience of welfare follows. Ultimatum game results suggest a corresponding experience of bad feeling and impulse to resist or punish upon the experience of non-reciprocal behavior linked to another concrete person. These data, however, do not take us very far, even if we accept that reciprocity with concrete others is motivated in the manner just sketched. We would still know nothing about which account best explains voting: an experience of agency mediated through identification with a collective; an extension of the welfare-enhancing experience of reciprocity to an impersonal scale based on the impression that others are also contributing (voting); or respect for settled rules of conduct that are implied by membership in a collective. Moreover, as the lure of voting clearly exercises different relative force on different people, there is no reason not to believe that each of these motives comes into play sometimes, in some contexts, in some persons.

To the extent that some of these motives are the social-emotional ones whose neural correlates have been observed in reciprocity games, their presence in other contexts might be observable. To the extent that the successful exercise of agency has neural correlates, the role of this experience in contributions to collective enterprises would also be open to investigation. As for rules of conduct implied by group identification, it would be possible to investigate whether these display the same correlates as utilitarian principles, when unproblematic and when under pressure from emotionally laden motives. If so, it might be possible to discern when such principles are in play in reasoning about collective action, and how powerful they are.

B. Commodification: Is there a there there?

1. The Anti-commodification case

A persistent problem for legal theory is the claim that commodification drains experience of moral richness and complexity. *Commodification* refers to governing a resource under the principles of market property, that is, allowing ownership of it with the incident of alienation, so that it becomes an object of market logic. Above all, as a market commodity the resource is fungible with other commodities, meaning that at any time its owner can convert it to cash, and consequently to its cash equivalent in a bundle of other goods. This metric of value is the main object of the anti-commodification complaint. The charge goes that valuing something in terms of its cash equivalent is inconsistent with valuing it for other qualities, such as beauty, personal attachment, or association with spiritual or cultural traditions. Moreover, according to critics of commodification, market valuation tends to overwhelm or crowd out other forms of

valuation, so that once something is commodified, those who once valued it for more elusive qualities will learn in time to value it for its cash equivalent. A fully commodified world, the argument implies, would be on all points subject to the curmudgeon's complaint about the cynic: that he knows the price of everything, but the value of nothing. Critics have directed variants of the argument at the use of markets for environmental regulation, the sale of organs, and proposals for markets in adopted children. Although there are still stylized debates over commodification's good or bad qualities in general, Margaret Jane Radin, who substantially introduced commodification concerns to the American legal academy, has developed a much more nuanced view, arguing that many domains are properly governed by market-style rights and regimes, but that goods connected with important or subtle forms of qualitative value, such as sexuality, should not be.[117]

There is quite a straightforward response to the complaint against commodification: *Not so!* One can simply deny that there is any inconsistency between valuing things for their qualitative properties and putting a price on them. After all, does the price not reflect our qualitative appreciation of the thing? If parents paid a market rate to adopt a child, it would presumably be because they were moved by all the usual parental motivations. And doesn't the price of beautiful homes or beautiful clothes reflect (in addition, perhaps, to their status value) potential buyers' appreciation of their beauty?

Even if one accepted that the two metrics of value are in some meaningful sense incompatible, rather than believe commodity value is derivative of qualitative value, that

[117] *See* MARGARET JANE RADIN, CONTESTED COMMODITIES: THE TROUBLE WITH TRADE IN SEX, CHILDREN BODY PARTS, AND OTHER THINGS 1-15, 79-101 (1996) (laying out the argument that commodification can have bad consequences for moral, emotional, and social life).

would not imply that commodity value ineluctably overwhelms more qualitative forms of value. It would still be perfectly possible in principle that the two could co-exist. This is a weaker version of *not so*, but nonetheless a variant of the same answer.

There is also an affirmative response to the anti-commodification charge. On this account, commodifying things is not just neutral in its effect on their valuation; rather, it is a way of designating them as mattering and thus of increasing or accentuating their "qualitative" value. This argument has two versions, varying usually with the resource at issue. For non-human resources, such as environmental goods, the suggestion is that, whatever pious noises people may make, in practice when things have no price, we tend to treat them not as priceless but as worthless. To say that one must pay for emitting air pollution, therefore, does not imply that the atmosphere is a "mere commodity"; rather, it carries the message that the atmosphere is *worth something*.

The second version of the affirmative response concerns commodification of aspects of human beings, particularly their time and labor power, but also organs, reproductive capacity, sexuality, and any other quality that another might seek to purchase or rent. The argument, which arose in the debate over commodification of labor, is that by (1) enabling people to exchange for market value what they previously could not, commodification increases their options and thus their effective freedom and (2) by creating a legally enforceable right to refuse to exchange these aspects of one's self, commodification defines a boundary of legal identity that reinforces autonomy and dignity.[118]

[118] I have discussed this argument in more detail *in* Jedediah Purdy, *A Freedom-Promoting Approach to Property: A Renewed Tradition for New Debates*, 72 U. CHI. L. REV. 1237, 1251-58 (2005) (discussing the arguments of Adam Smith and other early-modern market advocates that commodification of labor

2. The failure of inquiry and the question for neuroeconomics

Serious inquiry into the issue of commodification has all but ended. A good part of the reason, I suggest, is that commodification presents a classic other-minds problem. Take it as true that Margaret Radin experiences commodification as draining its objects of their other dimensions of value. That tells us nothing about Richard Posner, or anybody else, would respond to the same act of commodification. To make the point more abstractly and precisely: the prediction of the anti-commodification position is observationally indistinguishable from its contrary, so positive inquiry can tell us nothing. Describe a labor market: at the end of the description, you will have added nothing to our knowledge of whether participation in such a market degrades, ennobles, or merely allocates efficiently the act of labor. Describe a market in environmental pollution credits: the same discouraging proviso applies. Positive inquiry is in this respect restricted to regarding people as black boxes. When the question concerns the experience of value – not the choice that follows, but the qualities at play in the judgment that generates the choice – epistemic chastity toward other minds turns out to be a chastity belt.

Here neuroimaging has some promise of suggesting certain initial cuts, in a phrase, whether there is a there in commodification theory. The image of reasoning suggested by commodification theorists is very much like that evoked by neuroimaging results in the ultimatum game and moral reasoning exercises: struggle between a cognitively operational principle – here, maximizing commodity value – and a less

---

produced increasingly reciprocal relationships and reduced the indignity and degradation of status-based labor obligations).

cognitive response such as aesthetic appreciation, reverence, or love, which arrives in consciousness as a direct perception, rather than being mediated by conscious application of a principle. It is possible, of course, that this description misapprehends the anti-commodification charge, and that the claim is instead that viewing things in terms of their maximum commodity value becomes as "pre-cognitive" as beauty or pain; but suppose otherwise for the moment. If there is anything to this, then we might hypothesize that judgments involving trading off, say, money against beautiful landscapes or time with intimates would correlated with pattern of brain activity like those of the ultimatum game and moral-reasoning dilemmas. Such a finding would lend credence to the idea that these tradeoffs involve essentially different kinds of valuation. Tracking parallel tradeoffs in ways designed to capture degrees of commodification would then suggest something about the substance – or the emptiness – of commodification theory.

C. Authoritarianism: The highest stakes and the otherest minds

1. The positivist problem

Among the most consequential questions one might want social inquiry to address is why authoritarian politics succeeds or fails.[119] Authoritarian movements and

[119] In the formal social-science literature, this question got seriously underway with the publication of *The Authoritarian Personality* in 1950. *See* T.W. ADORNO, et al., THE AUTHORITARIAN PERSONALITY (1950). Sponsored and copyrighted by the American Jewish Committee, this massive study was a direct response to the arresting fact that Nazism had some to power in an educated, democratic, and, it would have seemed, relatively liberal society. The motive to make sense of the fascist appeal through social psychology, however, went back at least to Erich Fromm's *Escape from Freedom*, published early in the Second World War. *See* ERICH FROMM, ESCAPE FROM FREEDOM (1941). Subsequently, Robert Jay Lifton pursued Fromm's and Adorno's psychoanalytic approach to the problem, asking how the apocalyptic political impulse arises, persuading adherents that a climactic act of violence can make the world whole and pure. *See* ROBERT JAY LIFTON, DESTROYING THE WORLD TO SAVE IT:AUM SHINRIKYO, APOCALYTPIC VIOLENCE, AND THE NEW GLOBAL TERRORISM (1999). Departing sharply from the psychodynamic approach, Canadian political psychologist Bob Altemeyer has described authoritarian attitudes as a product of "social learning," a package of attitudes acquired and reinforced in one's social setting rather than related to any predisposition firmly established in early life. *See* BOB ALTEMEYER, ENEMIES OF FREEDOM:

governments have been the bane of democracies throughout the twentieth century, from Nazism to today's radical Hindu and Islamist politics: wherever they come to power through elections or attract popular support, they suggest that democracy may be self-immolating, at least under certain circumstances. Since Theodor Adorno and his collaborators began trying to understand varying individual attraction to fascism in Nazi Germany, students of authoritarian disposition have broadly agreed on the political attitudes whose origins they seek to understand: "suppression of [moral, cultural, and/or racial] difference and insistence upon uniformity," expressed in politics through "autocratic social arrangements in which individual autonomy yields to group authority."[120] These attitudes characterized fascist and nationalist movements in the early and middle decades of the twentieth century, and they characterize today's forms of extremism as well. Understanding why such politics appeals to some and repels others would be a momentous achievement.

Yet the problem has not inspired methodological lucidity. Instead, several persistent difficulties led to the eclipse of the study of authoritarian disposition. First was a confusion of independent and dependent variables. Researchers confounded the hypothetical object of the study – a *disposition* to authoritarian appeals – with its expression, adherence to authoritarian political beliefs.[121] Identifying bearers of

---

UNDERSTANDING RIGHT-WING AUTHORITARIANISM (1988). Others have stayed with the idea of strong predispositions, but looked to biological rather than social bases for these. *See* C.S. Bergeman, et al., *Genetic and Environmental Effects on Openness to Experience, Agreeableness, and Conscientiousness: An Adoption/Twin Study*," 61 J. OF PERSONALITY 159 (1993).

[120] *See id.* at 15. Stenner elsewhere elaborates on her findings in connection with this model: "authoritarians proved to be greatly alarmed by departures from moral and cultural absolutism, by any deviation from unquestioning conformity to external authority. And most characteristic of all, they invariably looked first to leaders and institutions to reinstate and reinforce the normative order, seeking to marshal the authority of the state to 'institute' the maintenance of 'civility' and 'hold up our moral values[.]' *Id.* at 267.

[121] *See* STENNER, *supra* n. __ at 20-23 (outlining this problem).

authoritarian disposition by their expression of authoritarian beliefs collapsed the distinction and left a circular definition of the authoritarian disposition: that disposition evidenced by expression of the same authoritarian beliefs the disposition was meant to explain.[122] Second was a related conceptual failure: the conflation of "authoritarian" attitudes with the political program of "right-wing" parties in the home countries of the investigators, so that "authoritarianism" was run together with other, local elements of right-wing politics, such as the conservative distaste for change in general and the libertarian preference for free markets over redistribution and economic regulation.[123] Third, theorists have mostly lacked any account of when the hypothetical disposition produces actual adherence to authoritarian politics and when it lies dormant.[124]

Stenner's response is to isolate a definition of "authoritarian" disposition that is distinct from conservative and libertarian outlooks and does not depend on the same political attitudes that it is supposed to predict. Moreover, she offers a testable theory of why the authoritarian disposition sometimes finds expression in intolerant political attitudes and sometimes lies dormant. Stenner defines the authoritarian disposition as "some general desire … to transfer sovereignty to, and commit self and others to conformity with *some* collective order," connected to a deep-set belief that the security and trustworthiness of the social world depend on "collective authority and conformity …

---

[122] *See* R. Nevitt Sanford, et al, *The Measurement of Implicit Antidemocratic Trends*, *in* ADORNO, et al., *supra* n. __ at 222 279 (explaining the method of measuring the authoritarian personality, the now-notorious "F Scale"). Altemeyer's "Right-Wing Authoritarianism" suffers from the same difficulty. *See* ALTEMEYER, *supra* n. __.

[123] *See* STENNER, *supra* n. __ at 138-98 (comparing authoritarian with conservative dispositions), 236-68 (comparing authoritarian with libertarian dispositions). This confusion has the methodologically unsettling consequence that, for instance, Russian communists who abhor the chaos of quasi-democratic capitalism score high on a "right-wing authoritarianism" measure that includes a preference for libertarian economic policies as part of its definition of authoritarianism. *See id.* at 149-50 (reporting and commenting on this "finding" by Altemeyer).

[124] *See id.* at 17-20 (outlining this difficulty and a response).

oneness and sameness."[125] She hypothesizes that this disposition finds expression under circumstances of "normative threat," when the "collective order" is threatened by conflicting values, "moral decay," or evidence that political leaders are unreliable.[126] Thus under conditions of apparent stability and moral consensus, authoritarians' views may not differ much from those of others. When events present "normative threat," however, authoritarians will respond with more conformist and punitive positions than non-authoritarians, and will make it a high political priority to rebuild stable and trustworthy authority.[127] By contrast, libertarians are much less likely to express increasingly authoritarian views even under conditions of normative threat.[128] Conservatives, while resistant to change and often hostile to dissensus, tend to mistrust the programs that distinguish authoritarian politics: efforts to reconstitute authority in the face of perceived breakdown, a strategy which many forms of nationalism and fascism exemplify.[129]

Methodologically, Stenner measures authoritarian disposition through questions designed to probe childrearing values. She asks respondents to rank as more important one of a pair of qualities that parents might to inculcate in their children.[130] The pairings present alternatives such as "[the child] follows the rules" or "he follows his own conscience," and "he has respect for his elders" or "he thinks for himself."[131] Stenner

---

[125] *Id.* at 141.
[126] *See id.* at 11-12, 26-29 (describing features of her theory of normative threat).
[127] *See id.* at 26-29, 85-98 (outlining authoritarian responses to normative threat and contrasting these with conservative responses).
[128] *See id.* at 261-68 (describing libertarian subjects as resistant to the idea that social phenomena that constitute normative threat for authoritarians are in fact evidence of "moral decay").
[129] *See id.* at 85-98. Stenner gives the telling example of Britain, where intolerance gets much support from authoritarianism, but little from conservatism, because the country's tolerant traditions make dispositional conservatism an ally of toleration.
[130] *See id.* at 23-25.
[131] *Id.* at 24.

proposes that these questions elicit basic orientations toward the relative value and importance of authority and uniformity versus autonomy and diversity. Because childrearing is a personal and high-stakes activity, she believes the answers should capture genuine, deep-set attitudes.[132] Stenner has found that high scores on her childrearing-based authoritarianism index correspond to heightened levels of authoritarian attitudes under conditions of normative threat, suggesting that the index captures the disposition she hypothesizes.[133]

2. The persistence of interpretation

It is critical to Stenner's methodological contribution that she treats her index of childrearing values as an independent variable. Although this is a promising move as a matter of social-science methodology, it does not dispel the problems that dog authoritarian studies. The aim in specifying the independent variable is to pick out some feature of personality that precedes politics and thus can explain why people with that personality adopt one political attitude rather than another. The difficulty is that, as the interpretivist position proposes, expressions of personality are always inflected through culturally specific worldviews. Thus, normative conceptions of the family and the adult personality that childrearing should produce are hardly more fixed or pre-political than more traditionally political attitudes. Rather, they are the products of constant interaction between ideological and material contests: on the one hand, ideas about the good or just

[132] *See id.* at 24-25. It strikes me as a fair concern that there is indeed a politics of parenting, and people's ideological and partisan affiliations may well influence their announced priorities in raising children. A glance at the cultural politics of the Christian Right in the United States will confirm this point. That said, however, precisely because childrearing is intimate and has substantial and immediate consequences, Stenner's judgment that it provides a relatively independent clue to basic orientations is plausible. Moreover, her findings upon testing the theory are impressive.

[133] *See id.* at 199-268 (detailing findings of Stenner's research).

family and the good or dignified person;[134] on the other hand, the control of resources that structures the "cooperative competition" within families and thus contributes to each succeeding generation's idea of what families actually are and normatively should be.[135] Moreover, ideas of the family are frequently part of explicitly political disputes and the partisan identity of citizens. One need reach no further for evidence than the choice of "family values" as a rallying point for cultural conservatives in the United States; the response that "hate is not a family value"; and the political charge of disputes over same-sex marriage.[136]

Moreover, family structure seems to work in a dynamic relationship with other values. For instance, research in India has found that the survival rates of female children go up considerably relative to male children when mothers become literate and enter the workforce.[137] Amartya Sen has argued that families become more gender-egalitarian as women increase both their bargaining power vis-à-vis their husbands and their ideas of what the world might hold for them.[138] Such changes have at least three, interdependent elements: (1) political decisions to encourage women's education and access to labor markets; (2) women's assertion of enhanced agency in the family after taking advantage of these changes; and (3) changing normative ideas of the family, specifically an increasingly egalitarian view of sons and daughters. Intuitively, these changes should be

---

[134] *See* TAYLOR, SOURCES, *supra* n. __ at 211-33 (describing rise of "the affirmation of ordinary life" of production, reproduction, and emotional attachment as a critical episode in the development of contemporary self-understanding).

[135] For a discussion of the idea, *see* AMARTYA SEN, DEVELOPMENT AS FREEDOM 196-203 (1999). The concept of "cooperative competition proposes that, while families act as a unit vis-à-vis the outside world (at least in many cases), within each family a set of priorities and decision procedures emerges from internal contests, which partly reflect each member's control over resources and choice of alternatives to going along with the family's present form and commitments).

[137] *See id.*

[138] *See id.*

mutually reinforcing, as each would tend to support the others. While this example does not speak to authoritarianism as such, it does instance the interaction of political and childrearing attitudes beyond cases, such as "family values," in which family structure is explicitly politicized.

None of this is to deny that Stenner proposes an enriched set of correlations among attitudes that may reveal much about the interaction of ideas about authority, order, and the threat of disorder at different levels of social structure, politics and the family. Moreover, homologies between attitudes toward family and toward politics add credence to the belief in some invariant need for order and authority and mistrust of plurality and uncertainty, which might indeed be called a "personality." Nonetheless, the concept of personality or disposition is no more than a metaphor for a constellation of attitudes that all occur on the same level of expression: participation in already interpreted world of cultural and political values. Moreover, on this evidence it may or may not be a misleading metaphor, suggesting an essential core of social-psychological potential that moves with a person throughout her life, exhibiting itself subtly in childrearing attitudes and now and again erupting into politics when "normative threat" awakes it.

3. The Question for Neuroeconomics

What would be genuinely new in this area would be to generate neuroimaging correlates for the key experiences in Stenner's model: encountering unwelcome evidence of cultural and ideological diversity, such as gay or interracial couples; facing instances of individuality-oriented childrearing; perceiving political leaders and other emblems of

authority besmirched or shamed, as during Richard Nixon's last years in office or Bill Clinton's period of sexual scandal. Stenner hypothesizes that these experiences trigger the same, or a closely related, sense of threat. We know a little about the neurological correlates of threat, and particularly the primitive sense of threat presented by a confusing and alarming problem that defies our mastery – such as uncertainty about the odds of a gamble, to take a particularly bloodless example. Neuroimagining could falsify or let stand the hypothesis that a common experience of threat unites these events.

Moreover, neuroimaging might reveal something about the character of susceptibility to authoritarian political appeals. If the hypothesis of a common threat structure found some support in imaging, the next step would be to explore the correlates of responsiveness to, let us say, videos of speeches by authoritarian leaders. (Although there will inevitably be partisan carping, I would nominate Patrick J. Buchanan's 1992 address to the Republican National Convention for American subjects.) What are the neural correlates of response to such political appeals? Are they more intense in those whose sense of threat has recently been engaged? And with what other attitudes and responses are they correlated? Is the responsive brain region associated with relief, aggression, higher reasoning? Do the correlates suggest a contest among regions of the brain, as with difficult ethical problems, or an unchecked emotional reaction? How do they differ from the correlates of, for instance, a participant in a structured exercise in reasoned political deliberation?

Hypotheses about authoritarian personality or disposition rely, like Stenner's, on claims that a set of observable attitudes and responses expresses some underlying orientation to the world, which might be identified independently of at least some of its

expressions. Neuroimaging has the potential (1) to test the hypothesis that the pattern reflects a unified underlying response and (2) to characterize the other emotions and cognitive functions associated with that response, and to contrast it with other kinds of political judgment.

As I stressed in discussing the interdependence of family and political attitudes, one must put aside any idea that brain activity is somehow the foundation and one-way cause of emotions and judgments, so that identifying the true underlying facts would enable us to make confident generalization and predictions about the epiphenomena of mind, culture, and politics. Neural correlates are just that, correlates, and the mental experiences to which they correspond are the activity of self-interpreting human beings. Nonetheless, patterns of correlation are facts that bridge some of the distance between observer and observed, diminishing the importance of the problem of other minds.

IV. Conclusion

How do people overcome collective-action problems? What do we value, and does our valuation change along with legal and social institutions? Why do some democratic citizens adopt authoritarian political attitudes while others resist them? It would be hard to identify a more basic set of questions for social inquiry. The first concerns the origins and persistence of social, political, and legal order. The second asks into the basic purposes and principles that guide that order. The third addresses the most elemental intrinsic threat to our liberal version of such order.

Each of these questions is at something of an impasse – particularly the second two, which are foundering on the suspicion that no there is there. The impasse has its

origins in the limitations of the main methodological alternatives in social inquiry. Positivism takes seriously the inaccessibility of other minds and thus stops chastely short of essential questions about the motivations of social phenomena. Interpretivism takes seriously the importance of other minds, but produces non-falsifiable fabrics of interpretation that depend, in the end, on the axioms one has selected to describe the inner lives of human beings.

The tantalizing promise of neuroeconomics is to bridge the gap between the two methods by rendering literally visible the activity of other minds. It will let us see reason, fear, and principle at work, let us watch utility accumulate or dissipate.

The promise is not a mirage, but neither is it the whole world. Reflecting on the three problems where I have suggested neuroeconomics can help overcome methodological limitations reveals a point too basic to overcome. Sometimes interpretation is an imperfect substitute for positive knowledge: when neuroimaging reveals the place of primitive fear in aversion to uncertainty, we have learned something, and can set aside certain interpretive alternatives. Sometimes, however, interpretation is indispensable because the facts of human activity are constituted by interpretation: individual and shared self-understandings, concerning both the nature of the world we confront and the nature of the value it contains, that vary across persons and across cultures. Knowledge of neural correlates can help to distinguish interpretations of interpretations, ideas about what, exactly, is going on "in there"; but it cannot fruitfully get interpretation out of the story by replacing it with a series of observable facts. Too often, those facts comprise and are the correlates of shifting and contested interpretation.

All of this strikes me as hopeful, even wonderful. Human experience is open-ended, in individual lives and in history, because we change as our self-understandings change. Because we are self-aware and self-interpreting creatures, no objective knowledge of our nature could ever reduce us to functions of neural events; those events are reciprocally born of our cultural and mental lives. The new knowledge of neuroeconomics does, however, promise to help us set aside some interpretive mistakes, choose among interpretive alternatives, and add to our stock of positive facts. It may make us less obscure to ourselves and to one another.