RATIONAL IRRATIONALITY: A FRAMEWORK FOR THE NEOCLASSICAL-BEHAVIORAL DEBATE

Bryan Caplan
George Mason University

INTRODUCTION

Economists have characterized *beliefs* as "rational," if agents satisfy Bayesian probability axioms; or, more strongly, if they also satisfy the rational expectations assumption[1] [Sheffrin, 1996; Wittman, 1995]. A diverse body of experimental evidence shows that individuals' beliefs deviate from these standards of rationality [Kahneman, Slovic, and Tversky, 1982; Camerer, 1995; Rabin, 1998]. Critics of these findings argue that the anomalies are suspect because financial incentives were absent or inadequate; people would be more rational—perhaps fully rational—if the monetary rewards were large enough [Harrison, 1992; 1990; 1989; Wittman, 1995; Friedman, 1998; Smith and Walker, 1993]. Defenders of the behavioral perspective reply that anomalies are generally robust to this criticism.

The purpose of the current article is not to resolve this controversy, but to provide a shared framework for debate. I present a model of "rational irrationality" and show that the main positions in the neoclassical-behavioral debate about beliefs are special cases of it. In this model, irrationality is a good like any other, and agents optimize by trading wealth for irrationality. Unless otherwise stated, "irrationality" is interpreted as "deviations from rational expectations," of which deviation from Bayesian axioms is a subset. The central assumption of the model of rational irrationality is that agents perceive the price of irrationality *without bias*. On some level, they have rational expectations about the consequences of irrationality, even though they typically hold a positive quantity of irrationality in their consumption bundle.

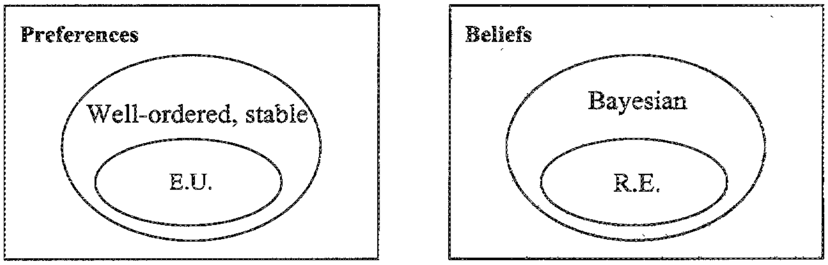
The upshot is that it is not necessary to see the neoclassical and behavioral approaches as two irreconcilable paradigms. The neoclassical-behavioral dispute over beliefs can instead be seen as a disagreement within "normal science" about parameter values; even if researchers cannot agree about their conclusions, at least they are asking the same questions. At the same time, the rational irrationality model does not tautologically define genuine irrationality away: a key falsifiable implication of the model is that (compensated) demand for irrationality must be decreasing in price. Experimental findings that irrationality *increases* as monetary incentives rise thus differ in kind from other anomalies and merit special attention [Hogarth and Reder, 1987].

Bryan Caplan: Department of Economics, Center for Study of Public Choice, George Mason University, 4400 University Drive, Fairfax, Virginia 22030-4444. E-mail: bcaplan@gmu.edu

Eastern Economic Journal, Vol. 26, No. 2, Spring 2000

Copyright 2000. All rights reserved

FIGURE 1
A Taxonomy of Rationality



Shaded region designates forms of "irrationality" considered herein.

The next section provides an elementary taxonomy of rationality and irrationality to clarify the scope of the current paper's investigations. Section three presents the simple model of rational irrationality. Section four analyzes four special kinds of wealth/irrationality indifference curves—neoclassical, near-neoclassical, near-behavioral, and behavioral—and their implications for the neoclassical-behavioral debate. The fifth section shows that several forms of wealth-*enhancing* irrationality can also be understood within the rational irrationality framework. Section six concludes the paper.

A TAXONOMY OF RATIONALITY AND IRRATIONALITY

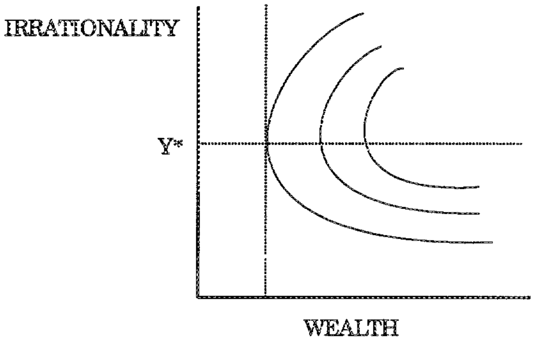
"Rationality," in economic parlance, is equivocal in at least two ways. Not only has it been ascribed to both preferences and beliefs, but in each domain there is a spectrum of rationality standards, from least to most demanding. In the interests of clarity, this section provides an elementary taxonomy of rationality. By no means intended to be exhaustive, its function is only to delimit the scope of the rational irrationality model's application.

At the outset, then, it is necessary to distinguish preferences from beliefs [Montgomery, 1996; Aumann, 1987]. An agent's preferences indicate how he would behave (i.e., what he would choose to do) in all conceivable situations. An agent's beliefs indicate what probability he would assign to any conceivable situation actually being the case. Figure 1 illustrates this contrast by showing preferences and beliefs as two disjoint sets.

Within the set of preferences, one can then draw the sub-set of "well-ordered and stable" preferences. This is typically the weakest hurdle for preferences to qualify as "rational" [Becker, 1962]. Adding on more restrictive conditions—most commonly, the expected-utility axioms of choice under uncertainty—further shrinks the set of rational preferences [Camerer, 1995]. Of course, both more and less demanding rationality criteria can and have been proposed [Harless and Camerer, 1994].

Copyright 2000 All rights reserved

FIGURE 2
Wealth/Irrationality Indifference Curves



Next consider the set of beliefs, where quite different senses of "rationality" apply. Typically, the least restrictive definition requires only that beliefs satisfy the familiar Bayesian probability axioms. These set no limits on agents' prior probabilities; in principle, an agent who is rational in this weak sense could be grossly and systematically mistaken about what the world is actually like. Imposing the more restrictive rational expectations assumption rules out such cases, requiring that agents not only satisfy the Bayesian axioms but also hold unbiased prior probabilities.

The remainder of this paper discusses the rationality of beliefs alone. "Rational" is used interchangeably with "satisfies the rational expectations assumption." The domain of "irrational" beliefs, the shaded region in Figure 1, is the *union* of (1) non-Bayesian beliefs, and (2) Bayesian beliefs that fail to satisfy the rational expectations assumption. This is not meant to suggest that this is the only form of "irrationality" worth considering—or to ignore the experimental evidence on "irrationality" in other senses of the word. Rather, the current paper focuses on violations from rational expectations because this problem is at once important and tractable. Analyzing behavior without well-ordered preferences is quite difficult; but given well-ordered preferences, analyzing irrational beliefs is—it will be argued—a manageable task.[2]

RATIONAL IRRATIONALITY

Suppose an agent has well-defined preferences over both personal wealth and beliefs; he cares about his wealth, but also has a "bliss belief" y^* that (holding wealth constant) he would most like to believe [Akerlof and Dickens, 1982; Akerlof, 1989; Caplan, 1999a, 1999b]. Wealth should be conceived in broad terms to include not just consumption and portfolio value, but also human capital, health, leisure, and so on; in fact, one could just partition all arguments in the utility function into "beliefs" and "everything else," and use wealth as a synonym for the latter. These preferences can then be represented with indifference curves in wealth/irrationality space as demonstrated in Figure 2. Wealth is on the x-axis, and the absolute value of the deviation from rational expectations is on the y-axis; an agent with rational expectations con-

Copyright 2000 All rights reserved

sumes zero y. The only unusual feature of these indifference curves is that they bend backwards at the bliss belief y^*; an agent is assumed to have some *specific* belief that he feels attracted to, rather than a contrarian desire to be as irrational as possible.

As with utility theory in general, one should not read too much psychological content into this choice model. A hungry person might state that he buys food not because he "wants" it but because he "needs" it. A worker could balk at the suggestion that he "reveals his preference for leisure" when he takes a break after a double shift. An agent who holds systematically biased beliefs that "just seem obvious" to him can be seen in the same light.

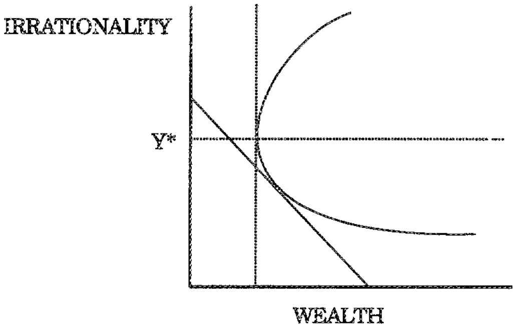
In psychological terms, cognitive and motivational biases are different: the motivational depend on the emotions, but the cognitive do not [Nisbett and Ross, 1980]. But the rational irrationality model treats them symmetrically for analytical purposes.[3] Still, the introspective experience of the effect of incentives could depend on the type of the bias. For a motivational bias like over-confidence, an agent might describe his response to incentives as: trying to be reasonable, suppressing his emotions, or making an effort to give opposing arguments a fair hearing. For cognitive biases like the availability bias,[4] the same agent might say that incentives make him more likely to doubt his initial intuition, look at aggregate rather than anecdotal evidence, ask for expert advice, or spend extra time researching the question.[5] Modeling cognitive and motivational biases symmetrically does not imply that they are introspectively equivalent.

In most situations of practical interest, systematically biased beliefs have a negative impact on personal wealth. Intuitively, an agent who consistently responds optimally to the way the world *isn't* almost certainly fails to respond optimally to the way the world *is*. As Nisbett and Ross put it, "The *costs* of willy-nilly distortions in perception are simply too high to make them a cure-all for the disappointed or threatened perceiver. In general, misperceptions make us less able to remedy the situations that threaten us or give us pain than do accurate perceptions" [1980, 234]. Suppose, for example, that a doctor genuinely thinks that 95 percent of the people who test positive for a disease have it, when in fact merely 2 percent do, as in the familiar base rate experiment [Casscells, Schoenberger, and Graboys, 1978]. Earnestly acting on this biased perception endangers the doctor's career prospects, risking malpractice suits, dissatisfied customers, and loss of professional reputation. Similarly, an agent who is systematically over-confident, mistaking 80 percent probability for perfect certainty, is likely to make a wide variety of losing bets; a poorly calibrated weather forecaster would be likely to lose his credibility and his audience [Lichtenstein, Fischhoff, and Phillips, 1982]. People who overestimate their own abilities, and accordingly bargain for more than they can reasonably expect to get, may reduce their average earnings [Babcock and Loewenstein, 1997].

Agents' wealth/irrationality "budget lines" therefore normally have the familiar negative slope as seen in Figure 3. A wealth/irrationality budget line shows the combinations of wealth and irrationality that are feasible. For the sake of convenience these expected losses are drawn as linear, i.e., proportional to the degree of bias, though this need not be the case. The budget line's intersection with the x-axis shows an individual's wealth assuming he strictly conforms to the rational expectations assumption. The budget line's intersection with the y-axis, in contrast, indicates the

Copyright 2000 All rights reserved

FIGURE 3
The Wealth/Irrationality Budget Line



level of rationality necessary to actually drive an actor's wealth down to zero. If wealth is defined broadly enough, the budget line's intersection with the *y*-axis could be interpreted as the point beyond which more extreme irrationality would be fatal.

Rationally irrational agents choose their utility-maximizing combination of wealth and irrationality based on an unbiased judgment about the tradeoffs. The higher the private cost of irrationality becomes, the flatter the budget line gets, and the smaller the optimal purchase of irrationality becomes as Figure 3 shows. The crucial assumption is that agents on some level have rational expectations about the slope of their wealth/irrationality budget line; they perceive the impact of their irrationality on their wealth *without bias*. This is what differentiates rational irrationality from the competing hypothesis of unqualified irrationality. The point here is not to define unqualified irrationality away, but to lay the groundwork for empirical comparison by spelling out the rational irrationality model's details.

If the private impact of irrationality on wealth were zero, the wealth/irrationality budget line would be vertical, and an optimizing agent would always choose his bliss belief y^*. This polar case plausibly arises under many circumstances. For example, it is pleasant to believe that your job is socially beneficial, but it unclear how biased beliefs on this point would be costly in terms of material wealth [Klein, 1994]. The same applies to many religious and political beliefs: for most people, there are no practical repercussions of doubting the theory of evolution or believing that one's nation is the "best in the world" [Caplan, 1999a].

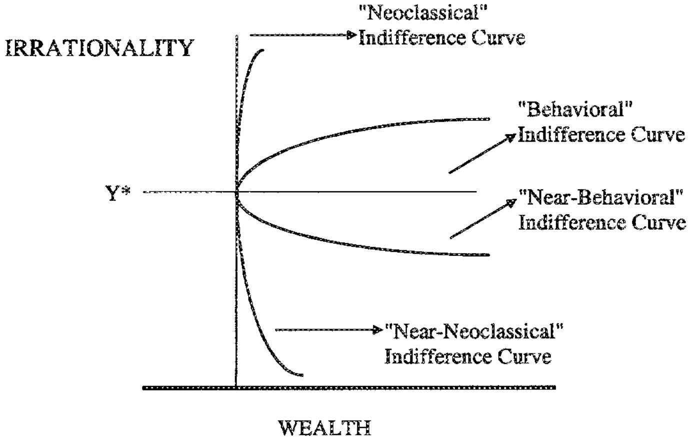
THE NEOCLASSICAL-BEHAVIORAL CONTINUUM AND ITS PRACTICAL SIGNIFICANCE

"Neoclassical" vs. "Behavioral" Indifference Curves

In standard neoclassical models, agents have no preferences over beliefs. Beliefs are a tool for getting more desirable commodities, not an end in themselves. "Neoclassical" wealth/irrationality preferences can therefore be represented as vertical indifference curves: "neoclassical" agents care only about their wealth, not their opinions.

Copyright 2000. All rights reserved

FIGURE 4
"Behavioral" vs. "Neoclassical" Indifference Curves


"Neoclassical" Indifference Curve
IRRATIONALITY
"Behavioral" Indifference Curve
Y*
"Near-Behavioral" Indifference Curve
"Near-Neoclassical" Indifference Curve
WEALTH

They worry about their actual job safety, not how safe they *believe* their job is [Akerlof and Dickens, 1982]. Similarly, they do not care about self-image or how certain they are [Dickens, 1985; Lichtenstein, Fischhoff, and Phillips, 1982]. The implication is that they have rational expectations if irrationality has *any* negative impact on wealth, however trivial.

In contrast, on behavioral accounts, irrationality and incentives are essentially unrelated; people have an irreducible propensity to make and cling to systematic errors. Cognitive anomalies are usually seen as inherently unresponsive to incentives. Motivational anomalies too—though admittedly payoff-dependent in theory—are often seen as payoff-independent in practice [Rabin, 1998; Dickens, 1986, 1985; Lichtenstein, Fischhoff, and Phillips, 1982]. As Piattelli-Palmarini puts it, "Between our rationality and our cognitive pride, we will choose the latter, and are willing to pay the price for so doing" [1994, 3]. The natural way to diagram the behavioral position on irrationality is with indifference curves that are *horizontal* at the bliss belief y^*. Just as the polar "neoclassical" agent is rational regardless of how weak the incentives are, the polar "behavioral" agent is *ir*rational regardless of how strong the incentives are as can be seen in Figure 4.[6]

Critics of behavioral anomalies often maintain that subjects have insufficient financial motivation to make them think rationally [Harrison, 1992; 1990; 1989; Wittman, 1995; Smith, 1991]. Wittman for example remarks: "Mistakes are quite likely to occur when no one suffers from them (and are especially likely when the researcher is searching for ways to trick the unsuspecting subjects)... [G]iven the insignificant incentives in experimental work, one should be surprised when experimental results do confirm economic theory" [1995, 41-2]. Examining polar cases of neoclassical and behavioral indifference curves suggests that this criticism of experimental anomalies is slightly off the mark. With rational respondents, lack of incentives merely explains an increase in the *variance* of beliefs; it would not account for the occurrence of the *systematic* mistakes emphasized in the experimental literature[7] [Harrison, 1990, 27-30; Smith and Walker, 1993].

Copyright 2000. All rights reserved

Within the rational irrationality framework, there is a straightforward way to reformulate the standard neoclassical critique. Suppose that instead of perfectly vertical wealth/ irrationality indifference curves, actors have "*near*-neoclassical" indifference curves: almost vertical, but curving outwards around$y^* > 0$ as Figure 4 shows.[8] Under most conditions, "near-neoclassical" actors cannot be distinguished from neoclassical actors: Neither is irrational if the price is non-trivial. The latent difference reveals itself only as the price of irrationality nears zero—as it would in an experiment without monetary incentives. In these choice environments, near-neoclassical actors consume appreciable amounts of irrationality, whereas perfectly neoclassical actors are as rational as ever. If indifference curves were really vertical, then regardless of the absence of incentives, critics of behavioral economics would lack an explanation of systematic experimental choice anomalies. But with "near-neoclassical" preferences, the main critique of behavioral economics is intelligible: people exhibit obvious irrational bias in the absence of material incentives, but at a small positive price this disappears.

The present paper does not try to resolve the behavioral-neoclassical controversy. What the rational irrationality model provides is a common framework for debate. There are not two irreconcilable approaches, but two endpoints on a continuum of possibilities. If the polar neoclassical position is empirically untenable, the polar behavioral position is not the only alternative. Economists with neoclassical priors (i.e., the prior judgment that people's wealth/irrationality indifference curves are vertical) might instead make the marginal move to the near-neoclassical view as experimental anomalies multiply. Conversely, economists with behavioral priors (i.e., that people's wealth/irrationality indifference curves are horizontal) need not leap to the neoclassical view if some evidence of price-sensitive irrationality accumulates. They could instead concede only that indifference curves are "near-behavioral": the consumption of irrationality is price sensitive but positive even at high prices. See Figure 4.

Adopting rational irrationality as a framework can also illuminate Harrison's [1989] claim that it is more informative to check for deviations from rationality in payoffs rather than beliefs.[9] Harrison notes that in terms of both statistical and economic significance, experimental divergence from full rationalisty is often large, but due to the flatness of most payoff functions, divergence from income maximization is usually small. How can this argument be related to the neoclassical-behavioral continuum? Suppose that an agent deviates from rationality in neither beliefs nor payoffs. (See Table 1.) This is consistent with *both* neoclassical and near-neoclassical preferences: the agent might be rational because he is congenitally rational, or because irrationality is costly. Similarly, if an agent deviates from rationality in *both* beliefs and payoffs, there is evidence for behavioral or near-behavioral preferences: the agent might be irrational because he is congenitally irrational, or because irrationality isn't costly enough. Observed deviations from rationality in beliefs alone, however, are consistent with everything except for polar neoclassical preferences; only neoclassical preferences are inconsistent with costless irrationality.[10] Finally, note that the fourth box in Table 1 is empty. Deviations from rational beliefs are what drag payoffs below their maximum expected value. It is thus impossible to simultaneously have rational beliefs and diverge from income maximization.

TABLE 1
Rational Irrationality, Beliefs, and Payoffs

Deviation from rationality in... is consistent with...	**Beliefs**	
Payoff	**Yes**	**No**
Yes	near-behavioral, behavioral	—
No	near-neoclassical, near-behavioral, behavioral	neoclassical, near-neoclassical

Practical Significance

Even if behavioral anomalies disappear in the face of financial incentives, findings of biased mistakes matter. As Smith and Walker observe, "[T]here are both low-stake and high-stake economic decisions in life, and all are of interest" [1993, 249]. The former can draw near-neoclassical agents to deviate from rational expectations; the latter may provoke near-behavioral agents to set aside their biases [Caplan, 1999a; Camerer, 1987; Frey and Eichenberger, 1994]. In addition, sometimes the *marginal* cost of irrationality is low even though the *total* stakes involved are high; as Russell and Thaler note, "the more efficient the market, the *less* discipline the market provides. In a fully arbitraged market, all goods (assets) yield the same characteristics per dollar (returns), thus individuals can choose in any manner without penalty" [1985, 1081].

The practical significance of *any* deviation from neoclassical preferences cannot be confined to lab experiments because this is only one form of low stake decision-making [Kirchgässner and Pommerehne, 1993]. Consider some others: the private cost of systematic misestimates of inflation is small over some ranges [Akerlof and Yellen, 1985]. Contingent valuation surveys overestimate the willingness to pay for environmental amenities [Harrison and Kriström, 1995]. More generally, surveys about issues with expressive value merit suspicion due to the negligible private cost of biased responses [Brennan and Lomasky, 1993; Boulier and Goldfarb, 1998]. Above all else, if rational expectations fail in experiments without adequate incentives, one would expect irrationality to play an important role in democratic elections due to the trivial private consequences of a vote [Caplan, 1999b; Akerlof, 1989].

Wittman specifically tries to *contrast* hypothetical surveys and elections: "When errors involve little cost (for example, answering a survey question incorrectly in a cognitive-psychology experiment), then little cognition will be employed. When errors involve great cost (for example, purchasing unreliable equipment for the military), then greater cognition will be involved..." [1995, 59]. Considering the enormous unlikelihood that a vote will change an electoral outcome, it is rather the *similarity* of surveys and elections that stands out. Wittman's disanalogy conflates private and social costs: in an election, errors may have large *social* costs, even though they rarely

Copyright 2000. All rights reserved.

will in laboratory surveys. But the *private* incentive structure of surveys and elections is identical: in both cases, the private cost of errors is effectively zero [Brennan and Lomasky, 1993, 38-41].

Similarly, even if behavioral anomalies are only slightly sensitive to material incentives, it makes a practical difference. Near-behavioral agents will make a wide range of systematic mistakes, but if they encounter the same situation repeatedly the price of irrationality may at last induce them to correct their errors. There are also self-selection effects to consider: if some people are closer to the behavioral pole than others, then people unwilling to abandon their biases even at high prices may switch to activities where their biases make less difference. This makes the long-run supply of rationality more elastic than evidence from experiments with randomly selected subjects might lead one to expect [Camerer, 1987].

Empirical Implementation

How exactly would empirical workers use this analysis? Two approaches suggest themselves. The first is to test whether (and when) rational irrationality can be rejected in favor of a model of unqualified irrationality. The second is to empirically calibrate the elasticity of irrationality with respect to incentives under various conditions.

Can the rational irrationality model be rejected? The rational irrationality model is consistent with a much broader range of observations than standard neoclassical models, but it cannot be reconciled with a special class of anomalies (income effects aside). Rational irrationality leaves open the possibility that biases *fail to decrease* as incentives for rationality intensify. But it rules out the possibility that stronger incentives for rationality actually make subjects *less* rational. This is noteworthy because several extant sources report findings along these lines [Einhorn and Hogarth, 1987; Camerer, 1995]. Einhorn and Hogarth rationalize these results as follows: "Performance, however, depends on both cognition and motivation. Thus, if incentive size can be thought of as analogous to the speed with which one travels in a given direction, cognition determines the direction. Therefore, if incentives are high but cognition is faulty, one gets to the wrong place faster."[11] [1987, 63] More straightforwardly, one could say that individuals simply perceive the wrong wealth/irrationality budget constraint.[12]

What the rational irrationality framework highlights is that such findings are not one more anomaly among many, but a challenge to rationality on the deepest level. As such, they deserve more scrutiny: Harrison [1992, 1990] finds that the marginal incentives in many well-known experiments have been a few pennies or less, and Thaler acknowledges that these findings have not been replicated "at very large stakes" [1987, 96] .

One recent and noteworthy study that sheds light on this question is Daniel Friedman's [1998] study of Monty Hall's 3-doors paradox. Friedman begins by experimentally demonstrating the paradox's strength: subjects switch about 30 percent of the time, even though they would switch 100 percent given unbiased probability esti-

Copyright 2000. All rights reserved.

mates. He then estimates linear probability models with "switch" (=1 if the subject switched, 0 otherwise) as the dependent variable, and tests the sensitivity of the basic results to modifications of the experiment. One variant is to move from regular to "intense" incentives, which increases the ex ante marginal cost of biased estimation five-fold—from $.10/turn to $.50/turn.[13] Friedman uses two different variables to capture the impact: Intense, a dummy variable equal to 1 if a subject had "intense" incentives, and Switchbonus, a continuous variable equal to (cumulative earnings for switching – cumulative earnings for not switching).[14]

Friedman reports that in a multiple regression, the coefficient on Intense is actually negative. Its discrete effect is to reduce the "switch" percentage by 9 percentage-points, a decline significant at the 5 percent level. [Friedman, 1998, 940] However, the coefficient on the continuous measure Switchbonus is positive and significant. In combination, the coefficients on Intense and Switchbonus indicate that at least after five turns, intense incentives yield better performance than regular incentives. As Friedman explains, "Since *Switchbonus* already captures the most relevant positive aspect of intense incentives, the significantly negative *Intense* coefficient should be thought of as measuring a residual impact" [*ibid.*, 944].

From the standpoint of the rational irrationality model, then, Friedman's findings are not anomalous, even though he presents strong evidence that considerable irrationality exists. The rational irrationality model moves the analytical spotlight from the "anomaly" in the usual sense of the term (violation of full rationality) to the anomalous response of rationality to stronger material incentives. Now suppose instead that Friedman had found that the coefficients on *both* Intense and Switchbonus were negative. This would definitely count as empirical evidence against the unlimited applicability of the rational irrationality model, since it would imply that consumption of irrationality rises when the price of irrationality increases.

One potentially important caveat for empirical researchers to take into account is the short-run versus the long-run elasticity of rationality. Since Friedman's discrete and continuous measures of incentives have different signs, strictly speaking, his results predict that *at first* subjects become less rational when the stakes increase. The positive continuous effect of incentives overtakes the negative discrete effect by turn six. In effect, then, the short-run response of rationality to incentives in the 3-doors game may be negative even though the long-run response is positive.

Divergence between the short- and long-run elasticity of rationality would presumably be far more pronounced if subjects had additional time during which to access a library or the internet. For example, in an alternative experimental design, subjects might play for an hour, receive an hour break of "free time," then return to play for one additional hour. Would higher stakes still have a temporarily perverse effect? Or as incentives rose would players be increasingly inclined to skip lunch and search out a correct explanation [Frey and Eichenberger, 1994]?[15]

Calibrating the elasticity of irrationality. A second way for empirical researchers to use the rational irrationality framework is to calibrate the elasticity of irrationality with respect to incentives. Assuming that irrationality declines as the price of irrationality rises, it is still vital to know "how much, under what conditions?" Re-

Copyright 2000 All rights reserved

turning to Friedman, one can roughly calculate the elasticity of irrationality in different conditions.

Extrapolating from the probit version of his results, Friedman infers that with intense incentives, subjects would be rational 90 percent of the time after 57 turns.[16] Following the same procedure [Friedman, 1998, 940] I calculate that with standard incentives, the corresponding rate after 57 turns would be only 35 percent . With non-intense incentives, subjects would reach the same level of rationality only after an estimated 257 turns. In other words, after 57 turns of play, increasing incentives by a factor of 5 increases the percentage of rational responses by a factor of 2.6, and reduces the number of turns required for 90 percent rationality by a factor of 4.5. Focusing on the more immediate effect of incentives, Friedman's results also imply that after 10 turns, the switch rate given intense and non-intense incentives will be 25 and 22 percent respectively. The implied elasticity still has the "right" sign. Yet 10 turns out, magnifying the incentives five times increases the percentage of rational responses by only 14 percent .

Overall, then, Friedman's results suggest that the short-run elasticity of irrationality with respect to incentives is fairly low, but the long-run elasticity is quite high. The reason for discussing the study, though, is not its specific conclusions, but rather to illustrate how the rational irrationality model might redirect empirical research. Within the rational irrationality framework, searching for deviations from full rationality remains useful. In fact, placing irrationality within a consistent theoretical framework would tend to defuse purely methodological objections against anomalous experimental findings. The rational irrationality framework only demands that evidence of anomalies be accompanied by additional work on their incentive-elasticity under different conditions.

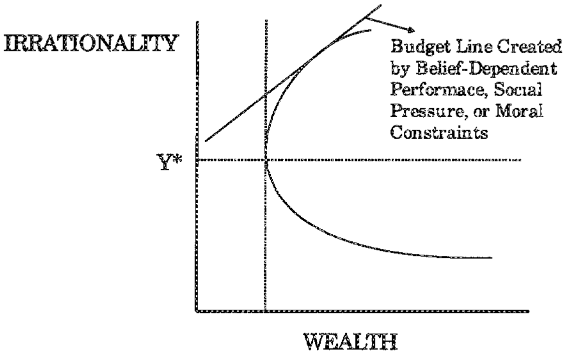
INCENTIVES FOR IRRATIONALITY: POSITIVE ILLUSIONS, MORAL CONSTRAINTS, AND SOCIAL PRESSURE

A simple intuition lies behind the negative slope of the wealth/irrationality budget line: if people respond optimally to the way the world is *not*, then except by pure chance they fail to respond optimally to the way the world *is*. But there are three notable exceptions to this principle. The first case is belief-dependent performance, where an agent's beliefs are actually an argument in his production function; the second, social pressure to deviate from rationality; the third, "self-serving bias" to circumvent fixed moral constraints. In each of these cases, agents actually face an *upward*-sloping wealth/irrationality budget line shown in Figure 5. As in Figure 3, the budget line is drawn as a straight line for illustrative purposes, but it could easily be non-linear. The steeper the line, the more wealth an agent forgoes if he refuses to delude himself. In effect, agents face a trade-off between a good (wealth), and a bad (irrationality in excess of y^*). To get more of one, they must endure more of the other, much like one might tolerate more pollution in order to produce additional steel.

Note that the positively-sloped budget line reaches a tangency only at some point *more* irrational than y^*. Even if a person likes being rational ($y^*=0$), the positive slope of the budget line provides an incentive for systematic bias in equilibrium. As before,

Copyright 2000. All rights reserved.

FIGURE 5
Belief-Dependent Performance, Social Pressure, and Moral Constraints



the responsiveness of beliefs to incentives varies along the neoclassical-behavioral continuum. The closer your indifference curves are to the neoclassical extreme, the *more* inclined you are to manipulate your views to your material advantage. More neoclassical agents readily "psyche themselves out" to enhance their performance, rationalize to avoid binding moral constraints, and change their opinions as social fashions change. The delusions of those closer to the behavioral pole, in contrast, are relatively stable; their views may be biased, but they are not conveniently interpreting reality to enhance their prospects.

Belief-Dependent Performance

Suppose that some of an agent's beliefs are arguments in that agent's production function. The clearest example is the placebo effect, where individuals' health really improves because they falsely believe they are receiving effective medical treatment. Taylor interestingly notes that "Placebo effects are so powerful that no new treatment or drug can be approved for general use in the practice of medicine unless its effectiveness has been evaluated against that of a placebo" [1989, 118]. But there are many non-medical examples too. Psychological studies have also found that frequently individuals with realistic—rather than over-optimistic—probability assessments are more likely to be depressed. With their mood less positive, they are often objectively less successful as a result [Camerer, 1995; Taylor, 1989]. Similarly, a student who believes he is likely to do well on a test feels calm and confident, and in consequence probably actually does better.

There is also evidence that the "illusion of control" is on net wealth-enhancing.[17] Stress impairs performance, and events believed to be uncontrollable are normally more stressful than equally unpleasant controllable ones. The production function of a person with the "illusion of control" about an unpleasant event is therefore less impaired by stress than the production function of someone immune to the illusion. As Taylor explains:

Copyright 2000. All rights reserved.

> When an event—even a painful or upsetting event—is perceived to be under personal control, the event does not produce as much stress as one perceived to be uncontrollable. When people experience uncontrollable stressful events, they react more negatively. Their physiological systems respond dramatically, leading to an increase in adrenalin secretion, which in turn has accompanying side effects such as the pounding of the heart, nervousness, and sweatiness. Psychological distress is greater when a stressful event cannot be controlled. People who are under stress that they cannot control also perform more poorly on other tasks. Concentration is limited, so it may be difficult to attend properly to what they need to do. [1989, 75-6]

In terms of the rational irrationality model, agents with any of these forms of belief-dependent performance actually face an upward-sloping budget line—at least over some range as in Figure 5. The flatter the slope of the price line, the greater the impact of irrational beliefs on performance. A person with rational expectations about the probability of success can rationally expect to be less successful and therefore less wealthy than the systematically overconfident. A person with rational expectations about the controllability of stressful events can rationally expect a higher level of stress and a lower level of wealth than a person subject to the illusion of control. As a result, even agents with no intrinsic taste for irrationality ($y^{*}=0$) on these topics will decide to be irrational in equilibrium. Perhaps this even explains why overconfidence, the illusion of control, and some other biases predominate empirically [Lichtenstein, Fischhoff, and Phillips, 1982; Langer, 1982; Taylor, 1989]: If over-rating your prospects actually improves your prospects, evolutionary forces tend to weed out tastes for rationality, not irrationality [Cosmides and Tooby, 1996, 64-68; Waldman, 1994].

Social Pressure

Belief-dependent performance can matter even for a solitary game against nature, but the scope for wealth-enhancing irrationality expands in multi-player games. The power of positive thinking aside, nature provides no incentive to be irrational, but other humans often do. Suppose people are more likely to hire, befriend, praise, and give political power to those who share their beliefs, and more likely to boycott, shun, denounce, and persecute those who disagree with them [Kuran, 1995; Klein, 1994; Becker, 1974]. In this environment, less rationality could easily make you more wealthy, giving rise to another instance of the upward-sloping wealth/irrationality budget line.[18] The more intense the social pressure (positive or negative), the flatter the slope of the line [Frey and Eichenberger, 1989].

What is remarkable is that using social pressure, a nucleus of true believers who support some irrational view (that is, with $y^{*}>0$) can generate positively-sloped budget lines for everyone. Facing social pressure to share in popular illusions, people with no intrinsic inclination towards irrationality (that is, with $y^{*}=0$) may nevertheless opt to be irrational. A test subject may have no strong feelings about the relative length of two sticks. When surrounded by confederates who all claim that the shorter

Copyright 2000. All rights reserved

stick is longer, though, test subjects frequently say—and perhaps convince themselves—that they agree with the popular position. On a grander scale, a small clique of committed Communists or believers in the caste system may make dissent so costly that their doctrine wins widespread acceptance [Kuran, 1995]. The "true believers" put pressure on their close associates to share their beliefs, these associates in turn pressure others, and irrationality consequently "trickles down" from activists to the general public.

Moral Constraints and Self-Serving Bias

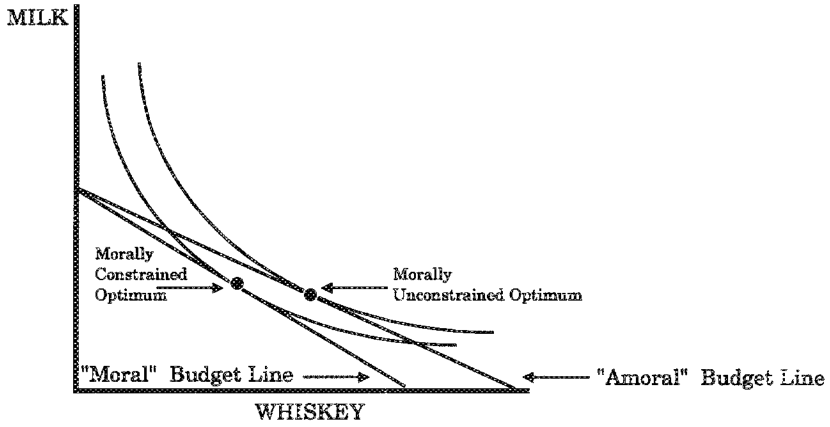
Rabin [1995] distinguishes between "moral preferences" and "moral constraints." Moral preferences are *in* the utility function; the more intensely an individual cares about a moral goal, the more he wants at a given price. Moral constraints, on the other hand, shrink one's permissible budget set. People pursue moral goals because they want to, whereas they obey moral constraints because they feel they must. Rabin explains that, "For *given* beliefs, there isn't a big behavioral distinction between these two models of morality," but matters change "when people can manipulate their beliefs" [1995, 4]. If people have moral *preferences*, mistaken beliefs misdirect their sincere efforts to do good, making them subjectively worse off. But if they instead face moral *constraints*, mistaken positive beliefs can expand their budget set and make them subjectively better off.

For example, suppose an agent chooses between two consumptive goods, milk and whiskey, shown in Figure 6. The "amoral" budget line shows that combinations the agent can financially afford; the tangency shows the most-preferred bundle, morality aside. If avoiding alcohol were an internalized moral preference, this would be modeled by shifting the agent's indifference curve. The contrary possibility, however, is that an agent avoids alcohol because of a perceived *external* moral constraint that forbids or discourages its consumption. Unlike the internalized moral preference, such a moral constraint would shift the effective budget line in along the x-axis, leaving the agent on a lower indifference curve. *Assuming the agent realizes the true alcohol content of whiskey*, he has to choose within the "moral" budget set instead of the "amoral" budget set. The knowledge assumption is critical: The more the agent underestimates its alcohol content, the less the moral constraint matters, and the higher his utility.

How can agents make moral constraints less binding? Rabin [1995] focuses on belief manipulation through selective acquisition of information, but biased information processing can in principle be just as effective. In fact, if agents choose their normative as well as positive beliefs, the impact of moral constraints disappears entirely; the "moral budget line" disappears, and they select a point on a higher indifference curve tangent to the "amoral budget line." If beliefs about moral constraints get in the way of self-interest, they could just abandon their beliefs in the relevant binding moral constraints. Suppose, however, that people's *normative* beliefs are fixed, but *positive* beliefs are not. Then an agent who maximizes utility subject to personal moral constraints might be better off with unreasonable positive beliefs, or "self-serving bias"[19] [Dahl and Ransom, 1999; Babcock and Loewenstein, 1997]. Facing what

Copyright 2000. All rights reserved.

FIGURE 6
Moral Constraints



appears to be a binding moral constraint, he rationalizes his way around it, "cheating" on his moral beliefs by manipulating his non-moral beliefs. The more he rationalizes the greater his utility from material goods. For example, the person in Figure 6 might conveniently underestimate the alcohol content of whiskey, moving the "moral budget line" closer to the "amoral budget line."

Note that Figures 5 and 6 illustrate two quite different tradeoffs. Figure 6 shows an ordinary tradeoff between two goods, complicated only by the fact that perceived moral constraints shift the budget line in. Figure 5 endogenizes the budget line for Figure 6; it shows that as positive beliefs connected to moral constraints become more biased, perceived moral constraints relax, making a higher level of wealth feasible. The only thing holding back unlimited rationalization is a distaste for irrationality. In effect, with choice over beliefs there is an additional margin (wealth/irrationality) to optimize along. The individual with fixed normative beliefs and flexible descriptive beliefs thus faces a *positively*-sloped wealth/irrationality budget line.

The more onerous the moral constraint is, the flatter is the budget line's slope; the payoff to rationalization increases with the absolute divergence between your morally-unconstrained utility-maximizing action and your morally-constrained utility-maximizing action. The factors Rabin [1995] identifies as mitigating self-serving bias, such as "salience injection," "moral dogmatism," and "moral priming" make equilibrium beliefs more rational by flattening wealth/irrationality indifference curves.

Empirical Implementation

How could empirical researchers use this analysis of wealth-enhancing irrationality? There are two dimensions worth considering. The first is to empirically distinguish between beneficial irrationality and low-cost irrationality. The second is to estimate the incentive-elasticity of wealth-enhancing irrationality. This section discusses the direction empirical work along these lines might take, using overconfidence anomalies [Lichtenstein, Fischhoff, and Phillips, 1982] to illustrate.

Copyright 2000. All rights reserved.

Beneficial irrationality versus low-cost irrationality. Some discussions of wealth-enhancing irrationality [Taylor, 1989] can be unclear whether those with "positive illusions" are (1) better off *on balance* because they are much happier but only slightly less objectively successful, or (2) better off because they are more objectively successful as well as happier. The rational irrationality model formally distinguishes these two cases: hypothesis (1) is just the standard case shown in Figure 3, while hypothesis (2) is the special case shown in Figure 5. One useful task for empirical researchers would be to determine which of these possibilities is actually the case. For example, does overestimating one's ability to correctly answer general interest questions actually increase the fraction of correct responses by reducing stress?

One way to get at this question, which to my knowledge has not been tried, is to combine standard tests of general knowledge with two sorts of incentives: a reward for a subject's total number of correct responses, and a reward for subjects to accurately guess the number of questions one answered correctly. Designating the number of correct answers as C, the subject's believed number of correct answers $\tilde{C}$, and rewards/prices p_1 and p_2, one such compensation formula would be: $Payoff = p_1C - p_2(\tilde{C} - C)^2$. If hypothesis (1) is correct, then raising p_2 should reduce $\tilde{C}$, but leave C unchanged; if hypothesis (2) is correct, then raising p_2 should reduce both $\tilde{C}$ and C.

Calibrating the elasticity of irrationality. For cases where hypothesis (2) applies, the next task for empirical study is to calibrate the incentive-elasticity of irrationality. The procedure is essentially the same as that in the 3-doors experiment: vary both incentives and other conditions and see how the expected degree of rationality changes. Suppose, for example, that one administers a test of 100 general knowledge questions. One could estimate each subject's "bliss belief" for $\tilde{C}$ by setting both p_1 and p_2 equal to 0. The rest of subjects' indifference curves could then be mapped out by varying p_1 and p_2. For example, one could set $(p_1, p_2) = (\$.20, \$.00)$ for one group of subjects, and set $(p_1, p_2) = (\$.30, \$.00)$ for those remaining. If the average values of V and $\tilde{C}$ for the first group were 55 and 58, and the average values of C and $\tilde{C}$ for the second group were 60 and 65, one could estimate that subjects were willing to increase their bias from 5 percent to 8 percent in exchange for an extra \$1.50 worth of income.

The most unusual facet of this problem is that the indifference curves become backward-bending after y^*. Moreover, while this paper's figures show irrationality as symmetric around y^*, this is not necessarily the way wealth/irrationality indifference curves actually look. Perhaps beliefs are more malleable (indifference curves are steeper) beyond y^* than they are between 0 and y^*, indicating that agents are more stubbornly committed to their wealth-impairing irrationality than they are to their wealth-enhancing irrationality. Thus, calibrating the elasticity of irrationality requires empirical work on both cases.

CONCLUSION

The central message of this paper is that the neoclassical-behavioral debate can be analyzed as an elasticity question. "Pure neoclassical" wealth/irrationality indif-

Copyright 2000. All rights reserved.

ference curves are vertical; such agents care only about wealth, not beliefs. "Pure behaviorial" indifference curves are horizontal at y^*; irrationality is constant and unrelated to material incentives. Both endpoints and all of the intermediate cases between them fit coherently into the rational irrationality model. Moreover, this framework is flexible enough to analyze diverse violations of rational expectations. Games of "man versus nature" where more irrationality has a negative impact on private wealth are the most obvious. But it can also be easily applied to games of "man versus himself" (for example belief-dependent performance and fixed moral constraints) and "man versus society" (for example social pressure) where more irrationality makes you materially better off.

While the central purpose of this paper is simply to provide a common framework for debate, it also has two key substantive findings. The first is that experiments where increasing incentives for rationality makes people *less* rational pose a unique challenge. Such anomalies merit additional attention because they are as inconsistent with rational irrationality as they are with rationality. Perhaps the long-run incentive-elasticity of rationality is positive even if the short-run response is negative. Higher stakes in a rigid setting have been found to make biases worse [Thaler, 1987]. But what about higher stakes combined with sufficient time to do more research, ask experts, or experiment?

The second conclusion is that even the staunchest critics of the behavioral findings must concede that many people's preferences are not neoclassical, but only near-neoclassical. This seemingly small concession has strong implications. Evolutionary claims about the possible scope of irrationality must distinguish between cases where irrationality is privately costly, private costless, or even privately beneficial; only in the first case are evolutionary arguments for perfectly neoclassical preferences at all compelling. If near-neoclassical preferences are widespread, then behavioral economics provides a distorted picture of market behavior[20], but *ipso facto* offers a compelling account of much non-market behavior. Most notably, *real-world* agents usually lack financial incentives to be rational in politics [Akerlof, 1989; Caplan, 1999a, 1999b]. The implication for experimental design: to simulate market conditions, you must use private material rewards, but to simulate political conditions, you must *not*.[21]

NOTES

For discussion and useful suggestions I would like to thank Don Boudreaux, Tyler Cowen, Pete Boettke, Jim Schneider, Geoffrey Brennan, Bill Dougan, Bill Dickens, Mitch Mitchell, Ed Lopez, J.C. Bradbury, Todd Zywicki, David Bernstein, Robin Hanson, Dan Klein, Alex Tabarrok, Nicky Tynan, Timur Kuran, Ron Heiner, Roger Congleton, Fab Rojas, David Levy, Kenneth Koford, Daniel Friedman, two anonymous referees, seminar participants at George Mason, participants at the Public Choice Outreach seminar and the Public Choice Society meetings, and members of my Armchair Economists' listserv. Gisele Silva and Eric Crampton provided excellent research assistance. The standard disclaimer applies.

[1] In contrast, they have characterized *preferences* as "rational" if they are well-ordered and stable; or, more strongly, if they also satisfy the expected-utility axioms of choice under uncertainty [Becker, 1962; Camerer, 1995]. The current paper focuses solely on the rationality of beliefs. The second section of this paper presents a basic taxonomy of irrationality.

Copyright 2000. All rights reserved.

2. As an anonymous referee puts it, "Indeed, if one thinks of rational expectations as having a 'correct' understanding of the 'mechanisms' that actually determine outcomes in the real world, then someone could have rational expectations—that is, understand and 'forecast' the world very well—yet still *behave* 'irrationally,' meaning in a way that did not involve maximizing his or her utility."
3. This paper takes the sensitivity of cognitive biases to incentives as an empirical question, as for example, Tversky and Kahneman seem to: "This article has been concerned with cognitive biases that stem from the reliance on judgmental heuristics. These biases are not attributable to motivational effects such as wishful thinking or the distortion of judgments by payoffs and penalties. Indeed, several of the severe errors of judgment reported earlier occurred despite the fact that the subjects were encouraged to be accurate and were rewarded for the correct answers" [1982, 18]. It should be noted however that other researchers appear to take responsiveness to incentives as definitional; if a form of irrationality decreases as payoffs rise, it couldn't have been cognitive in the first place. As Piattelli-Palmarini puts it, "It is in fact important to distinguish carefully between cognitive illusions and simple errors of judgment or blunders due to inattention, distraction, lack of interest, poor preparation, genuine stupidity, timidity, braggadocio, emotional imbalance, and so on... It should be emphasized that cognitive science consists of probing the ordinary mental structures of the human species, free from such spurious effects as are due to motivation, emotions, or aggressivity" [1994, 141].
4 The availability bias, as Tversky and Kahneman explain, arises because "people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences may be brought to mind" [1982, 11] even though factors other than frequency and probability—such as familiarity and salience—predictably affect the ease of recalling examples.
5. "Incentives do not operate by magic: they work by focusing attention and by prolonging deliberation. Consequently, they are more likely to prevent errors that arise from insufficient attention and effort than errors that arise from misperception or faulty intuition" [Tversky and Kahneman, 1987, 90]. But why couldn't incentives prompt people to be more skeptical about their intuition or perception, turn to aggregate evidence, double-check with an expert, or expend more research time?
6. Consider the special case of behavioral indifference curves horizontal at $y^*=0$. A person with such indifference curves has rational expectations not because it is in his self-interest, but because he has an intrinsic taste for truth. He therefore stays rational even when irrationality has a positive effect on wealth (see section 4). "Truth-loving" indifference curves are interesting at least for normative purposes, since they capture the ideals of scientific and philosophic objectivity.
7. Harrison is aware of this difficulty, but after considering three possible explanations he concludes that economists "have no useful business fussing around in an attempt to make sense of unmotivated behavior" [1992, 1441]. Smith and Walker also recognize the need to explain "why subject decisions are not just random responses in the absence of salient rewards" [1993, 248] . Their decision cost model explains biased errors as the product of "bounded rationality," plus truncation. It is not clear, however, that their model can explain price-sensitive systematic errors when truncation problems are not an issue.
8. Near-neoclassical preferences have no connection to Russell and Thaler's notion of quasi rationality: "[D]epending on how the problem is framed, it can be predicted whether the agent will choose x or y. We propose calling any such regular yet nonrational behavior *quasi rational*" [1985, 1072].
9. Harrison [1989] actually uses slightly different terminology: "message space" instead of "beliefs," "expected payoff space" instead of just "payoffs."
10. This point is nearly identical to what Harrison calls his "payoff dominance" critique of experimental findings. In terms of my framework, however strong the experimental case against polar neoclassical preferences, everywhere but that pole is consistent with the evidence.
11. The experimental findings Einhorn and Hogarth discussed here concern utility theory rather than rational expectations; elsewhere in the same volume Hogarth and Reder [1987, 12] apply this point more generally.
12. I owe this formulation to an anonymous referee.
13. With regular incentives, a correct pick earns \$.40, while a wrong pick earns \$.10; with intense incentives, a correct pick earns \$1.00, while a wrong pick earns −\$.50. [Friedman, 1998] For regular incentives, $E(\text{payoff}\mid\text{switch})=\$(2/3\times.40+1/3\times.10)=\$.30$, $E(\text{payoff}\mid\text{don't switch})=\$(2/3\times.10+1/3\times.40)=\$.20$, and thus the expected marginal cost of biased versus unbiased probability estimates is $E(\text{payoff}\mid\text{switch})-E(\text{payoff}\mid\text{don't switch})=\$.30-\$.20=\$.10$. For intense incentives, $E(\text{payoff}\mid\text{switch})=\$.50$, $E(\text{payoff}\mid\text{don't switch})=\$.00$, and thus $E(\text{payoff}\mid\text{switch})-E(\text{payoff}\mid\text{don't switch})=\$.50$.

Copyright 2000. All rights reserved.

14. Switchbonus should not be interpreted as a simple practice effect, because the set of independent variables also includes a trend variable.
15. On a basic Altavista search of "3 doors puzzle," the first three hits explain the correct answer. Even in a search of "switch stay 3," one of the first ten hits offers a correct explanation.
16. For the part of the experiment with both intense and standard incentives, subjects played either 12 or 15 turns.
17. The "illusion of control is defined as an expectancy of a personal success probability inappropriately higher than the objective probability would warrant" [Langer 1982, 231]. More generally, it consists in treating chance events as controllable.
18. A second possible response to social pressure is to engage in "preference falsification," i.e., pretend to share advantageous views without actually adopting them [Kuran, 1995].
19. This is only one form of self-serving bias discussed in the literature. Many forms of self-serving bias have been seen in normative, rather than positive, beliefs, where the concept of systematic bias is more difficult to apply. Self-interest may be empirically shown to influence perceived fairness [Dahl and Ransom, 1999], but without a measure of "true fairness" there is no way to test for rational expectations in the usual sense. In terms of wealth, moreover, self-serving biases have often been seen as *counter*-productive: A person who overestimates his productivity will probably earn less, not more, as a result. Babcock and Loewenstein raise an important caveat when they ask "whether it benefits a party to be less biased, holding constant the beliefs of the other party" [1997, 116]. If the answer to their question is yes for a given form of self-serving bias, then it can be analyzed with a standard negatively-sloped budget line. If the answer is no—less bias is harmful given the biases of others—then it should be analyzed with a positively-sloped budget line.
20. And not even that in every case, as Akerlof and Yellen [1985] show, or as Russell and Thaler's [1985] findings on consumers' detergent choices suggest.
21. Rationality in politics is in effect a public good, which, as numerous experiments show, will tend to be under-supplied even if it is not entirely absent [Ledyard, 1995]

REFERENCES

Akerlof, G. The Economics of Illusion. *Economics and Politics*, Spring 1989, 1-15.

Akerlof, G. and Dickens. W. The Economic Consequences of Cognitive Dissonance. *American Economic Review*, June 1982, 307-19.

Akerlof, G. and Yellen, J. Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria? *American Economic Review*, September 1985, 708-20.

Aumann, R. Correlated Equilibrium as a Expression of Bayesian Rationality. *Econometrica*, January 1987, 1-18.

Babcock, L. and Loewenstein. G. Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives*, Winter 1997, 109-126.

Becker, G. Irrational Behavior and Economic Theory. *Journal of Political Economy*, February 1962, 1-13.

____________. A Theory of Social Interactions. *Journal of Political Economy*, June 1974, 1063-91.

Boulier, B. and Goldfarb, R. On the Use and Nonuse of Surveys in Economics. *Journal of Economic Methodology*, 1998, 1-21.

Brennan, G. and Lomasky, L. *Democracy and Decision: The Pure Theory of Electoral Preference*. Cambridge: Cambridge University Press, 1993.

Camerer, C. Do Biases in Probability Judgments Matter in Markets? Experimental Evidence. *American Economic Review*, December 1987, 981-97.

____________. Individual Decision Making, in *The Handbook of Experimental Economics*, edited by J. Kagel and A. Roth. Princeton, NJ: Princeton University Press, 1995.

Caplan, B. Rational Ignorance vs. Rational Irrationality. Unpublished Manuscript, George Mason University, 1999a.

____________. The Logic of Collective Belief. Unpublished Manuscript, George Mason University, 1999b.

Casscells, W., Schoenberger, A., and Graboys, T. Interpretation by Physicians of Clinical Laboratory Results. *New England Journal of Medicine*, November 1978, 999-1001.

Cosmides, L. and Tooby, J. Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty. *Cognition*, January 1996, 1-73.

Copyright 2000 All rights reserved

Dahl, G. and Ransom, M. Does Where You Stand Depend on Where You Sit? *American Economic Review*, September 1999, 703-27.

Dickens, W. Occupational Safety and Health and 'Irrational' Behavior: A Preliminary Analysis, in *Workers' Compensation Benefits: Adequacy, Equity, and Efficiency*, edited by J. Worroll and D. Appell. Ithaca: ILR Press, 1985, 19-40.

___________. Safety Regulation and 'Irrational' Behavior, in *Handbook of Behavioral Economics*, vol. A, edited by B. Gilad and S. Kaish. Greenwich, CT: JAI Press, 1986, 325-48.

Einhorn, H. and Hogarth, R. Decision Making Under Ambiguity, in *Rational Choice: The Contrast Between Economics and Psychology*, edited by R. Hogarth and M. Reder. Chicago: University of Chicago Press, 1987, 41-66.

Frey, B. and Eichenberger, R. Anomalies and Institutions. *Journal of Institutional and Theoretical Economics* 145, September 1989, 423-437.

___________. Economic Incentives Transform Psychological Anomalies. *Journal of Economic Behavior and Organization*, March 1994, 215-34.

Friedman, D. Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly. *American Economic Review*, September 1998, 933-946.

Harless, D. and Camerer, C. The Predictive Utility of Generalized Expected Utility Theories. *Econometrica*, November 1994, 1251-89.

Harrison, G. Theory and Misbehavior of First-Price Auctions. *American Economic Review*, September 1989, 749-762.

___________. The Payoff Dominance Critique of Experimental Economics. Unpublished Manuscript, URL http://theweb.badm.sc.edu/glenn/dominance.pdf, 1990.

___________. Theory and Misbehavior of First-Price Auctions: Reply. *American Economic Review*, December 1992, 1426-1443.

Harrison, G. and Kriström, B. On the Interpretation of Responses in Contingent Valuation Surveys, in *Current Issues in Environmental Economics*, edited by P. Johansson, B. Kriström, and K. Mäler. Manchester: Manchester University Press, 1995, 35-57.

Hogarth, R. and Reder, M. Introduction: Perspectives from Economics and Psychology, in *Rational Choice: The Contrast Between Economics and Psychology*, edited by R. Hogarth and M. Reder. Chicago: University of Chicago Press, 1987, 1-23.

Kahneman, D., Slovic, P. and Tversky, A. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press, 1982.

Kirchgässner, G. and Pommerehne, W. Low-Cost Decisions as a Challenge to Public Choice. *Public Choice*, September 1993, 107-15.

Klein, D. If Government is So Villainous, How Come Government Officials Don't Seem Like Villains? *Economics and Philosophy*, April 1994, 91-106.

Kuran, T. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Cambridge: Harvard University Press, 1995.

Langer, E. The Illusion of Control, in *Judgment Under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic and A. Tversky. Cambridge: Cambridge University Press, 1982, 230-8.

Ledyard, J. Public Goods: A Survey of Experimental Research, *The Handbook of Experimental Economics*, in J. Kagel, and A. Roth. Princeton, NJ: Princeton University Press, 1995.

Lichtenstein, S., Fischhoff, B. and Phillips, L. Calibration of Probabilities: The State of the Art to 1980, in *Judgment Under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic and A. Tversky. Cambridge: Cambridge University Press, 1982, 306-34.

Montgomery, J. Contemplations on the Economic Approach to Religious Behavior. *American Economic Review Papers and Proceedings*, May 1996, 443-7.

Nisbett, R. and Ross, L. *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs: Prentice-Hall, 1980.

Piattelli-Palmarini, M. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds.* New York: John Wiley and Sons, 1994.

Rabin, M. Moral Preferences, Moral Constraints, and Self-Serving Biases. Unpublished Manuscript, University of California Berkeley, 1995.

___________. Psychology and Economics. *Journal of Economic Literature*, March 1998, 11-46.

Russell, T. and Thaler, R. The Relevance of Quasi Rationality in Competitive Markets. *American Economic Review*, December 1985, 1071-82.

Copyright 2000 All rights reserved

Sheffrin, S. *Rational Expectations*. Cambridge: Cambridge University Press, 1996.

Smith, V. Rational Choice: The Contrast Between Economics and Psychology. *Journal of Political Economy*, August 1991, 877-897.

Smith, V. and Walker, J. Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry*, April 1993, 245-61.

Taylor, S. *Positive Illusions: Creative Self-Deception and the Healthy Mind.* NY: Basic Books, 1989.

Thaler, R. The Psychology and Economics Conference Handbook: Comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman, in *Rational Choice: The Contrast Between Economics and Psychology*, edited by R. Hogarth, and M. Reder. Chicago: University of Chicago Press, 1987, 95-100.

Tversky, A. and Kahneman, D. Judgment Under Uncertainty: Heuristics and Biases, in *Judgment Under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press, 1982, 3-20.

__________. and __________. Rational Choice and the Framing of Decisions, in *Rational Choice: The Contrast Between Economics and Psychology*, edited by R. Hogarth, and M. Reder. Chicago: University of Chicago Press, 1987, 1-23.

Waldman, M. Systematic Errors and the Theory of Natural Selection. *American Economic Review*, June 1994, 482-97.

Wittman, D. *The Myth of Democratic Failure: Why Political Institutions are Efficient.* Chicago: University of Chicago Press, 1995.

Copyright 2000. All rights reserved.