UC Berkeley
Other Recent Work

Title
Psychology and Economics

Permalink
https://escholarship.org/uc/item/8jd5z5j2

Author
Rabin, Matthew

Publication Date
1997

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA AT BERKELEY

Department of Economics

Berkeley, California 94720-3880

Working Paper No. 97-251

Psychology and Economics

Matthew Rabin

Department of Economics
University of California, Berkeley

January 1997

Key words: bounded rationality, decision making, fairness, framing effects, heuristics and biases, preferences, psychology, reciprocity, reference-dependent utility

JEL Classification: A12, B49, D11, D60, D81, D83, D91

This is a longer version of an essay under preparation for possible publication in the *Journal of Economic Literature*. I thank John Pencavel and anonymous referees for earlier comments on its structure and content. For comments on this draft, I thank Steven Blatt, Colin Camerer, Peter Diamond, Erik Eyster, Ernst Fehr, Danny Kahneman, George Loewenstein, Ted O'Donoghue, and John Pencavel. For helpful conversations over the past several years on topics covered in this essay, I thank George Akerlof, Gary Charness, Eddie Dekel, Peter Diamond, David Laibson, David I. Levine, George Loewenstein, Rob MacCoun, James Montgomery, Vai-Lam Mui, Drazen Prelec, and especially Colin Camerer, Danny Kahneman, and Richard Thaler. Co-authors on research related to the topics of this essay include David Bowman, Deborah Minehart, Ted O'Donoghue, and Joel Schrag. Helpful research assistance was provided by Gadi Barlevy, Nikki Blasberg, Gail Brennan, Paul Ellickson, April Franco, Marcus Heng, Bruce Hsu, Jin Woo Jung, and especially Steven Blatt, Jimmy Chan, Erik Eyster, and Clara Wang. I am extremely grateful for financial support from the Russell Sage and Alfred P. Sloan Foundations.

Psychology and Economics

Abstract

Because psychology systematically explores human judgment, behavior, and well-being, it can teach us important facts about how humans differ from traditional economic assumptions. In this essay I discuss a selection of psychological findings relevant to economics. Standard economics assumes that each person has stable, well-defined preferences, and that she rationally maximizes those preferences. Section 2 considers what psychological research teaches us about the true form of preferences, allowing us to make economics more realistic within the rational-choice framework. Section 3 reviews research on biases in judgment under uncertainty; because those biases lead people to make systematic errors in their attempts to maximize their preferences, this research poses a more radical challenge to the economics model. The array of psychological findings reviewed in Section 4 points to an even more radical critique of the economics model: Even if we are willing to modify our familiar assumptions about preferences, or allow that people make systematic errors in their attempts to maximize those preferences, it is sometimes misleading to conceptualize people as attempting to maximize well-defined coherent, or stable preferences.

Matthew Rabin
Department of Economics -3880
University of California, Berkeley
Berkeley, California 94720-3880
rabin@econ.berkeley.edu

1. Introduction

Because psychology systematically explores human judgment, behavior, and well-being, it can teach us important facts about how humans differ from traditional economic assumptions. In this essay I discuss a selection of psychological findings relevant to economics.

There are two basic components of the standard economics model of the individual: That she has stable, well-defined preferences, and that she rationally maximizes those preferences. Given some choice set X, a person is assumed to "Max $_{x \in X}$ U(x)." Psychological research can be roughly categorized by how radically it challenges this model, and by the nature of the modifications implied.

Section 2 considers what psychological research can teach us about the true form of the function U(x). By searching for ways to make U(x) more realistic within the rational-choice framework, Section 2 reviews evidence that requires relatively small modifications of the familiar economic framework. I begin by discussing research suggesting that a person's preferences are often determined by changes in outcomes relative to reference levels, not merely by absolute levels of outcomes. In particular, relative to the status quo (or other reference points), people dislike losses significantly more than they like gains. I discuss the implications of this "loss aversion" for choice under uncertainty, and then briefly discuss attitudes towards risk and uncertainty more generally. I then discuss how people depart from pure self interest (as narrowly defined), pursuing "other-regarding" tastes such as fairness, reciprocal altruism, and revenge.

Section 3 presents psychological evidence on systematic errors people make in their attempts to "Max" U(x), and thus poses a more radical challenge to the economics model. Specifically, Section 3 reviews research on biases in judgment under uncertainty. I will discuss in detail a few of these biases (how we under-use base rates, infer too much from too little evidence, and misread evidence as confirming previously held hypotheses), and explain a few more in broad outline. I also discuss some of the psychological evidence on how learning and reasoning do and don't lead people to overcome these biases.

The array of psychological findings reviewed in Section 4 points to an even more radical critique of the economics model. Even if we are willing to modify our standard assumptions about U(x), or allow that people make systematic errors in their attempts to maximize U(x), it is sometimes misleading to conceptualize people as attempting to maximize well-defined, coherent, or stable U(x)'s. I begin by reviewing evidence that people are not fully adept at evaluating their own preferences—we don't always accurately predict our own future preferences, nor even accurately assess our experienced well-being from past choices. I then discuss research on framing effects, preference reversals, and related phenomena in which people prefer some option x to y when the choice is elicited one way, but prefer y to x when the choice is elicited another way. I then turn to evidence that we sometimes make choices because we seek to have justifiable "reasons" for making those choices, rather than because of underlying preferences. I then discuss self-serving biases and "motivated cognition"—how our wishes influence our beliefs. I conclude Section 4 with a discussion of self control and other phenomena that arise because a person's motivations and propensities at different times are inconsistent.

Because specific psychological research is so often interpreted (by both psychologists and economists) under the shadow of general debates about whether alternatives to familiar assumptions can or should be engaged by economists, I briefly assess some of these arguments in Section 5. I then conclude the essay by discussing some of the ways economists can incorporate psychological research.

The customary disclaimer of review essays—that they do not pretend to be exhaustive—applies here: The topics covered will necessarily be only a small fraction of economically-relevant psychology.[1] Other restrictions on scope abound. My primary emphasis is on reviewing what psychology tells us about modifying our general model of individuals, rather than on any of its specific economic applications. For instance, while fairness and reference-level effects (reviewed in Section 2) and framing effects (reviewed in Section 4) are likely to contribute to downward stickiness in wages, I shall leave it for other forums to explore these implications. Also, I emphasize *what* psychologists have learned about people, rather than *how* they have learned it. Consequently, though many experiments are described, the focus of this essay is not at all on experimental methods *per se*.[2]

Mainstream economics employs a powerful combination of methods: methodological individualism, mathematical formalization of assumptions, logical analysis of what conclusions follow from those assumptions, and sophisticated empirical field testing. I believe these methods are tremendously useful, and should not be abandoned. But these methods often raise a barrier to incorporating psychological insights into economics: While psychology investigates humans in all their richness, economics requires models that are not so rich as to retard the process of drawing out their economic implications.[3] For a discipline such as economics that places a high

---

[1] Some topics of obvious economic relevance omitted are status, envy, and social comparisons; the social psychology of conformity, group solidarity, and herd behavior; the tendency of "extrinsic motivation" (e.g., organizational incentive schemes) to drive out "intrinsic motivation" (e.g., the internal drive to excel at your vocation); self-perception theory (that we are often uncertain about our own motivations and other facts about ourselves, and infer some of these facts from our behavior); and a mass of research on learning from cognitive and developmental psychology. Another topic I have omitted is "non-psychological" models of bounded rationality. Researchers have formulated models of bounded rationality (based on intuition, computer science, or artificial intelligence) that are meant to capture cognitive limits of economic actors, but which do not invoke research on the specific patterns of errors that human beings make. In some arenas (e.g., "unforeseen contingencies" in complex contracts) it makes sense to focus on abstract conceptions of bounded rationality. In other contexts, however, I believe the orientation of the research presented in Section 3, to identify patterns in the types of errors humans make, is more likely to be enlightening. For an excellent review of the role of bounded rationality in economics, broader in scope than the topics I cover here, see Conlisk (1996).

[2] Descriptions of experimental methods employed by psychologists can be found in virtually any psychology text. For an excellent set of reviews of the results of experimental economics, see Kagel and Roth (1995). While most of the findings I present in this essay are by psychologists, I do not confine myself to reviewing research by people with PhDs in psychology; I will often draw on evidence from experimental economics as well. My focus, however, is on experiments that explore the psychological nature of individuals, not on research seeking to replicate economic institutions in the laboratory.

[3] Any mainstream economist can appreciate Robert Louis Stevenson's (1884, p.100) advice to aspiring young novelists:

> Let him not mind if he miss a thousand qualities, so that he keeps unflaggingly in pursuit of the one he has chosen. ... let him bear in mind that his novel is not a transcript of life, to be judged by its exactitude; but a simplification of some side or point of life, to stand or fall by its significant simplicity.

premium on the logic and precision of arguments and the quantification of evidence, incorporating all facets of human nature is neither attainable nor desirable.

As messy as complicating our familiar model of humans will be, however, it is not legitimate for economics to continue to ignore psychological research.[4] Many psychological findings are robust enough, tractable enough, and of enough potential economic importance that we ought begin to integrate them into economics. Other findings raise more fundamental challenges, and incorporating them into economic research will take longer. But even in these cases, economists ought to become aware of the shortcomings of our models, regret these shortcomings, and keep our eyes open for ways to remedy them. Most importantly, we must abandon meta-arguments about whether it is "possible" that psychologists have identified economically relevant departures from rationality, self interest, and other familiar assumptions. Of course it is possible, and in fact it is true. In this essay, I review what research shows some of those departures to be.

Substituting "model" for "novel," the quote corresponds to our advice to an aspiring young economist developing her first formal model—that she should not try to create a model that is a "transcript" of some economic phenomenon, but rather one that is a "simplification of some side" of the phenomenon.

[4] The belligerent discuriosity with which mainstream economics has at times dismissed behavioral evidence parallels the words used by Isaiah Berlin (1993, p. 114) in sympathizing with the frustrations of a critic of mainstream enlightenment philosophy:

> Kant has rightly won the day, but Hamann and his followers express a continual revolt against taking so much so blandly for granted, against leaving out so much, perhaps necessarily, but with too little regret, with no qualms, as if what the theory cannot embrace is mere expendable rubbish: psychological idiosyncrasy, oddities and quirks, which theory cannot notice and which in a rational universe will themselves be ironed out, so that the facts will be only such as the final infallible theory fits.

Despite remnants of this attitude, however, it is clear that many of the psychological results presented below and elsewhere will in coming years become both uncontroversial and well-integrated into economics, and that already economists are becoming more receptive to such behavioral research. The number of behaviorally-inspired papers appearing in prominent economics journals has increased dramatically in recent years, and many mainstream economists clearly feel that this type of research is a healthy development. Most skepticism these days comes in the form of demands for clearer statements of what specifically we learn from behavioral research, rather than *a priori* insistence that we will surely learn nothing.

2. Preferences

The premise of this section is that many psychological findings can be used to improve the realism of economics simply by modifying the utility functions we employ. Hence, this section of the essay involves the least radical departure from mainstream economics.[5]

Reference Levels, Adaptation, and Losses

Overwhelming evidence shows that humans are often more sensitive to how an outcome differs from some reference level than to the absolute level of the outcome itself. While applying this principle to consumption and income, Kahneman and Tversky (1979a, p. 277) stress that the salience of changes from reference points is a basic aspect of human nature:

> An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point (Helson (1964)). Thus, an object at a given temperature may be experienced as hot or cold to the touch depending on the temperature to which one has adapted. The same principle applies to non-sensory attributes such as health, prestige, and wealth. The same level of wealth, for example, may imply abject poverty for one person and great riches for another—depending on their current assets.

Understanding that people often feel changes more intensely than absolute levels suggests that we ought incorporate into utility analysis such factors as habitual levels of consumption. For instance, instead of utility at time t, u_t, depending solely on present consumption, c_t, it may also depend on a "reference level," r_t, determined by factors like past consumption or expectations of future consumption. Hence, instead of positing a utility function of the form $u_t(c_t)$, utility should be written in a more general form, $u_t(r_t, c_t)$.

The relative lack of economic models with reference-dependent preferences does not appear to be a case of fundamental methodological resistance; rather, it seems merely to be a bad habit.[6] Reference dependence deserves

---

[5] For positive analysis, the usefulness of the utility-maximization framework depends on whether choice data can be usefully organized by positing that people maximize stable utility functions. A more controversial and rarer question about this framework is whether the preferences which people seem to maximize correspond to the well-being they actually experience. Many economists consider such a question off limits, feeling that "by definition" the actions of informed people reflect what makes them happy. But there is a coherent sense in which even outcomes intentionally chosen may not maximize a decision maker's experienced well-being. I shall return to this issue in Sections 4 and 5.

to be, and is gradually becoming, an important part of economic modeling. But reference points influence behavior in domains, and to degrees, that economists will find surprising.

One ubiquitous pattern stands out: Losses resonate more than gains. In a wide variety of domains, people are more averse to losses than they are attracted to same-sized gains.[7] One of the main realms where loss aversion plays out is in preferences over wealth levels. That the displeasure from a monetary loss is greater than the pleasure from a same-sized gain is also implied by a concave utility function, which economists typically use as the explanation for risk aversion. What distinguishes loss aversion from conventional risk aversion is that the value function abruptly changes slope at the reference level, so that people are significantly "risk averse" for even small amounts of money. People dislike losing $10 more than they like gaining $11, and hence prefer their status quo to a 50/50 bet of losing $10 or gaining $11. (Tversky and Kahneman (1991) suggest that in most domains where sizes of losses and gains can be measured, people value moderate losses roughly twice as much as equal-sized gains.) While such "first-order" risk aversion is widely observed, the standard concave-utility function implies that people are close to risk-neutral for small stakes.[8] A "kink" in the utility function (illustrated later in Figure 2) is required to get such an attitude towards risk.[9]

Loss aversion is related to the striking *endowment effect* identified by Thaler (1980, 1985). Once a person comes to possess a good, she immediately values it more than before she possessed it. Tversky and Kahneman (1991, pp. 1041-2) describe an experiment conducted by Kahneman, Knetsch, and Thaler (1990) that nicely illustrates this phenomenon:

> Kahneman, Knetsch, and Thaler [1990] tested the endowment effect in a series of experiments, conducted in a classroom setting. In one of these experiments a decorated mug (retail value of about $5) was placed in front of one third of the seats after students had chosen their places. All participants received a questionnaire. The form given to the recipients of a mug (the "sellers") indicated that "You now own the object in your possession. You have the option of selling it if a price, which will be determined later, is acceptable to you. For each of the possible prices below

---

[6] Early examples of economists' considering such issues include Duesenberry (1952), Easterlin (1974), Markowitz (1952), McGlothlin (1956), Pigou (1951), and Ryder and Heal (1973).

[7] See Kahneman, Knetsch, and Thaler (1991) for an excellent review. Papers on this theme also include Kahneman (1992), Kahneman, Knetsch, and Thaler (1986a, 1986b, 1990, 1991), Kahneman and Thaler (1991), Kahneman and Tversky (1979a, 1982c, 1984), Tversky and Griffin (1991), and Tversky and Kahneman (1981, 1986, 1991, 1992). Economic applications include Benartzi and Thaler (1995), Bowman, Minehart, and Rabin (1996), Camerer, Babcock, Loewenstein, and Thaler (1995), Coursey, Hovis, and Schulze (1987), Fershtman (1993), Frank and Hutchens (1993), Hartman, Donae, and Woo (1991), Knetsch (1992), Knetsch and Sniden (1984), Loewenstein and Sicherman (1991), Samuelson and Zeckhauser (1988), Shalev (1994, 1995a, 1995b), and Thaler (1980, 1985).

[8] See Arrow (1974), Pratt (1964), and Samuelson (1961). In many cases where concavity is invoked to explain observed behavior, calibration would lead to rejection of the reference-independent, expected-utility model. Suppose, for instance, we observe that people typically turn down 50/50 gambles of losing $10 and winning $11 over a wide range if initial wealth levels. The degree of concavity of the utility function needed to explain this within the standard framework would imply that the person would, for instance, turn down 50/50 gambles in which she loses $100 or gains $10,000.

[9] The above statements refer to choice behavior when people are expected-utility maximizers. Segal and Spivak (1990), for instance, develop a model of first-order risk aversion outside the expected-utility framework.

> indicate whether you wish to (x) Sell your object and receive this price; (y) Keep your object and take it home with you...." The subjects indicated their decision for prices ranging from \$0.50 to \$9.50 in steps of 50 cents. Some of the students who had not received a mug (the "choosers") were given a similar questionnaire, informing them that they would have the option of receiving either a mug or a sum of money to be determined later. They indicated their preference between a mug and sums of money ranging from \$0.50 to \$9.50. The choosers and the sellers face precisely the same decision problem, but their reference states differ. ... [the choosers] face a positive choice between two options that dominate [their reference state] ... The sellers ... must choose between retaining the status quo (the mug) or giving up the mug in exchange for money. Thus, the mug is evaluated as a gain by the choosers, and as a loss by the sellers. Loss aversion entails that the rate of exchange of the mug against money will be different in the two cases. Indeed, the median value of the mug was \$7.12 for the sellers and \$3.12 for the choosers in one experiment, \$7.00 and \$3.50 in another. The difference between these values reflects an endowment effect which is produced, apparently instantaneously, by giving an individual property rights over a consumption good.

The behavior described here is usefully conceptualized as a case of loss aversion comparable to that identified in choice among lotteries. Individuals who are randomly given mugs treat the mugs as part of their reference levels or endowments, and consider not having a mug to be a loss, whereas individuals without mugs consider not having a mug as remaining at their reference point.

As established by Knetsch and Sinden (1984), Samuelson and Zeckhauser (1988), and Knetsch (1989), a comparable phenomenon—the *status quo bias*—holds in multiple-good choice problems. Here, loss aversion implies that an individual's willingness to trade one object for another depends on which object she begins with: Individuals tend to prefer the status quo to changes that involve losses in some dimensions, even when these losses are coupled with gains in other dimensions. Knetsch and Sinden (1984) and Knetsch (1989), for instance, demonstrated the status quo bias by randomly giving one set of students candy bars, and the remaining students decorated mugs. Later, each student was offered the opportunity to exchange her gift for the other one—a mug for a candy bar or vice versa. 90% of *both* mug-owners and candy-owners chose not to trade. Because the goods were allocated randomly and transaction costs were minimal, the different behavior for the two groups of subjects must have reflected preferences that were induced by the allocation.

Knetsch (1989) experimentally demonstrates that such preferences can be usefully captured by utility functions defined over reference levels as well as consumption levels.[10] Consider the following rather stylized graph. The loci of points, I_A and I_B, represent indifference curves for bundles of Goods 1 and 2 for a consumer at, respectively, reference point A and reference point B.

[10] See also Tversky and Kahneman (1991) for a formalization.

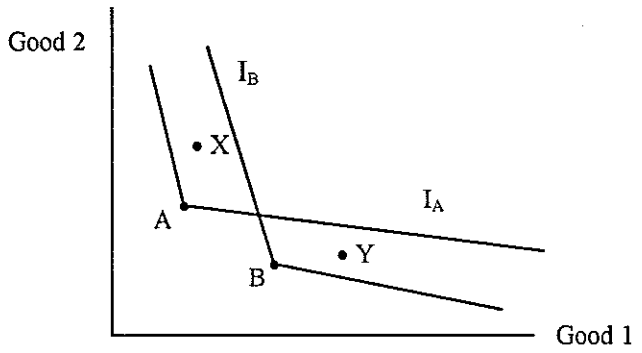

Figure 1

Such indifference curves capture the status quo bias, because they imply that the consumer strictly prefers A over B if she is at A, and B over A if she is at B. More generally, she will tend to prefer bundles that avoid losses of any goods; thus, she will prefer X over Y if she is at A, but Y over X if she is at B.

In addition to loss aversion, another important feature of how people assess departures from their reference levels is that they have *diminishing sensitivity*—the marginal change in perceived well-being is greater for changes that are close to one's reference level than for changes that are further away. As with loss aversion, Kahneman and Tversky (1979a, p. 278) argue that diminishing sensitivity reflects a more fundamental feature of human cognition and motivation:

> Many sensory and perceptual dimensions share the property that the psychological response is a concave function of the magnitude of physical change. For example, it is easier to discriminate between a change of 3 and a change of 6 in room temperature, than it is to discriminate between a change of 13 and a change of 16 . We propose that this principle applies in particular to the evaluation of monetary changes. Thus, the difference between a gain of 100 and a gain of 200 appears to be greater than the difference between a gain of 1,100 and a gain of 1,200. Similarly, the difference between a loss of 100 and a loss of 200 appears greater than the difference between a loss of 1,100 and a loss of 1,200, unless the larger loss is intolerable. Thus, we hypothesize that the value function for changes of wealth is normally concave above the reference point ... and often convex below it... That is, the marginal value of both gains and losses generally decreases with their magnitude.

In the context of preferences over uncertain monetary outcomes, diminishing sensitivity implies that the slope of a person's utility function over wealth becomes less steep as her wealth gets further away from her reference level. Because for losses relative to the reference level, "further away" implies lower wealth levels, diminishing sensitivity has a provocative implication: While people are likely to be risk averse over gains, they are often risk-*loving* over losses. As evidence, Kahneman and Tversky (1979a) found that 70% of subjects report that they would prefer a 3/4 probability of losing nothing and 1/4 probability of losing $6,000 to a 2/4 probability of losing nothing and 1/4 probability each of losing $4,000 or $2,000.[11] The preferred lottery here is a mean-preserving spread of

[11] See also Fishburn and Kochenberger (1979) for evidence of diminishing sensitivity.

the less-preferred lottery; hence, the responses of 70% of the subjects are inconsistent with the standard concavity assumption. More generally, there is evidence that people often have risk-seeking preferences over negative outcomes.[12] Consider again the utility function U(r,c), where c is the consumption level and r is the reference level. Kahneman and Tversky (1979a) illustrate the value function for gains and losses in the following graph, which incorporates both loss aversion and diminishing sensitivity into the utility function defined over gains and losses in wealth:

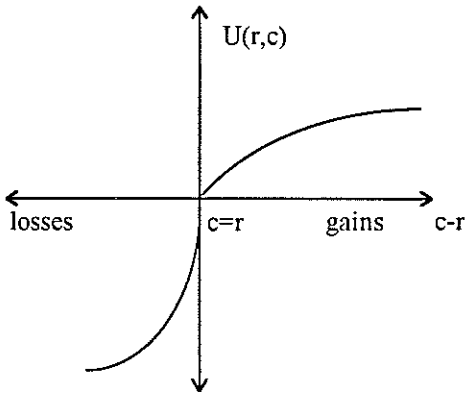

Figure 2

In order to study the effects of reference points in dynamic contexts within a rational-choice, utility-maximization framework, we need to take account of how people feel about the effects of their current choices on their future reference points.[13] To maximize their long-run utilities when reference points matter, people must determine two things: how current behavior affects future reference points, *and* how they feel about changes in their reference points. One approach to the first question, pioneered by the economists Ryder and Heal (1973), attempts to model explicitly the process by which reference points change. In particular, they assume that $r_t = \alpha c_{t-1} + (1-\alpha)\, r_{t-1}$, where $\alpha \in (0,1)$ is a parameter measuring how quickly people adjust their reference points. In a "rational-expectations" model, people will take this formula into account when maximizing their long-run well-being. Such an account of how reference levels are determined seems intuitive, but not very well grounded in behavioral evidence.[14]

---

[12] The "risk-loving" tendency does battle with the diminishing marginal utility of income, which people also display. Thus, for large losses that might push a person to extremely low consumption levels, risk aversion may reappear. Empirical evidence yields mixed results regarding risk attitudes over losses.

[13] As discussed below in both the first and last subsections of Section 4, evidence is strong that people *don't* rationally foresee consequences for their future preferences of their current behavior, and often underweight those consequences even if they do foresee them. Consequently, attempts to understand dynamic choice with reference-dependence entirely within the utility-maximization framework may stretch that framework too far.

[14] Bowman, Minehart, and Rabin (1996) combine the Ryder and Heal approach of rational-expectations, reference-dependent utilities with a utility function that incorporates loss aversion and diminishing sensitivity. Duesenberry

Evidence is similarly sparse about people's *preferences* over changes in reference points. If preferences are described by a utility function $U_t(r_t,c_t)$, then the question remains: How does U_t depend on r_t? Without assumptions about this relationship, there will be a relatively small set of circumstances where loss aversion and diminishing sensitivity can be integrated into models of dynamic utility maximization. Yet relatively little behavioral evidence speaks to how U_t depends on r_t.[15]

There have been some initial attempts to study loss aversion, the endowment effect, and the status quo bias in economic and related contexts. Hartman, Doane, and Woo (1991) find empirical evidence for the existence of a status quo bias in consumer demand for electrical utilities; Bowman, Minehart, and Rabin (1996) replicate evidence by Shea (1995a, 1995b) that consumers are more averse to lowering consumption in response to bad news about income than they are to increasing consumption in response to good news, and argue that this behavior is a natural implication of loss aversion.[16]

Uncertainty and Expected-Utility Theory

Traditional economic models of choice under uncertainty assume that people are subjective expected-utility maximizers: They maximize the weighted average of utilities they could get in different uncertain outcomes, where the weights are equal to the perceived probabilities of the outcomes. That is, if S represents a set of mutually exclusive and exhaustive outcomes a person can experience from decision, U(s) is her utility for an outcome $s \in S$, and p(s) is the probability she assesses to that outcome, then her utility can be represented by $\Sigma_{s \in S}\, p(s) \cdot U(s)$.

Over the past several decades, economists and psychologists have amassed a great deal of evidence that the expected-utility model is behaviorally wrong in systematic ways.[17] Whereas the expected-utility model is defined by assuming that preferences are linear in probabilities, some behavior is better explained by utility functions that are non-linear in probabilities. Consider an example that Machina (1987, p. 128) borrows from Allais (1953,

---

(1952) implicitly posited a function closer to $r_t = \text{Max}_{\tau<t}\, c_\tau$—that a person's reference level was her highest past consumption level.

[15] For exceptions, see Loewenstein and Sicherman (1991), Frank (1989, 1992 Ch. 15), and Frank and Hutchens (1993), who have identified a tendency for people to prefer income profiles that are steady or increasing profiles of income over time to same-sized decreasing profiles, strongly indicating that people prefer not to become accustomed to consumption when they know they cannot maintain that level. For consideration of some other factors involved in the preference for increasing wage profiles, see Loewenstein and Sicherman (1991, pp. 76-82).

[16] See also Cohen and Knetsch (1990), who consider the role of the status quo bias in judicial decision making.

[17] In part because this literature is vast and well-integrated into economic theory, I review it only briefly here. Many examples I cite reflect research initiated largely by economists, not psychologists. While this research has yet to become widely integrated into economic applications (but see Hogarth and Kunreuther (1985, 1989, 1990), Camerer and Kunreuther (1989), Dekel (1989), Weil (1990), Dow and Ribeiro da Costa Werlang 1992), Sarin and Weber (1993), Epstein and Wang (1994), Machina (1994), Fox, Rogers, and Tversky (1995)), it has over the years been the domain with the most extensive interaction between economic theorists and mathematical psychologists.

1979). Each of four lotteries, a_1, a_2, a_3, and a_4, consists of (sometimes zero) probabilities of winning \$0, \$1 million, or \$5 million:

	a_1	a_2	a_3	a_4
Prob \$0	.00	.01	.90	.89
Prob \$1 million	1.00	.89	.00	.11
Prob \$5 Million	.00	.10	.10	.00

In Allais's original and subsequent studies, people tend to choose a_3 over a_4, but a_1 over a_2. This contradicts expected-utility theory, which says that a person's preferences over what happens in some probabilistic "contingency" (here, her preferences between an 11% probability of \$1 million versus a 10% probability of \$5 and a 1% probability of \$0) are independent of what happens in another contingency. The choice of a_3 over a_4 says the person likes the idea of replacing an 11% probability of \$1 million with a 10% probability of \$5 million and a 1% probability of \$0. By contrast, the choice of a_1 over a_2 says that the person dislikes the exact same tradeoff—here she is *un*willing to replace an 11% probability of \$1 million with a 10% probability of \$5 million and a 1% probability of \$0. This type of result, which has been replicated many times and has come to be known as the *Allais paradox*, can be interpreted as saying that when probabilities of a worst outcome are tiny (as in a_1), a person demands a huge increase in the probability of the best outcome to compensate for increasing the probability of the worst outcome by a little (consequently, a person rejects a_2 in favor of a_1); but when the probability of the worst outcome is large (as in a_4), a person is more willing to accept an increase in the likelihood of the worst outcome to increase the probability of the best outcome (consequently, a person accepts a_3 over a_4). Machina (1982), Dekel (1986), and others formalize such a hypothesis.

Subsequent experiments have found other departures from expected utility.[18] Many researchers (see especially Kahneman and Tversky (1979a)), have emphasized how people are more attentive to small changes in probabilities when probabilities are very close to zero and one than they are for other ranges. Others have shown how notions of "regret" and anticipation can be formalized, and helped to explain some anomalous departures from expected utility (see, e.g., Loomes and Sugden (1982, 1987a, 1987b), and Gul (1991)). Some researchers have highlighted the interaction between probabilities of decision weights and the differential attitudes towards gains and losses discussed in the previous subsection. In particular, Tversky and Kahneman (1992, p. 297) conclude that decision weights and the utility function described in the previous section combine to imply "a distinctive fourfold pattern of risk attitudes: risk aversion for gains and risk seeking for losses of high probability; risk seeking for gains and risk aversion for losses of low probability." All in all, this literature has given rise to a host of formal

---

[18] The experimental literature on "non-expected utility" or "generalized expected utility" includes Hodges and Lehman (1952), Allais (1953, 1979), Coombs and Huang (1976), Loomes and Sugden (1987), Camerer (1989), Conlisk (1989), Battalio, Kagel, and Jiranyakul (1990), Kagel, Macdonald, and Battalio (1990), Tversky and Kahneman (1992), Camerer and Ho (1994), Harless and Camerer (1994), and Fox, Rogers, and Tversky (1995). For some reviews of this research, see Machina (1987) and especially Camerer (1992b), and Camerer (1995, Sections III.D. and III.E.)

models providing alternatives to the expected utility model.[19] While none are quite as parsimonious as the expected-utility model, many explain behavior significantly better.[20]

The Allais paradox and other behavioral anomalies have given rise to modifications of the expected-utility model whereby probabilities enter preferences non-linearly. Even these models, however, assume that preferences over uncertain outcomes can be represented as a function of a person's probabilistic assessments—independent of the "type" of uncertainty to which those probabilistic assessments pertain. Research beginning largely with Ellsberg (1961) has shown that people either don't *form* coherent subjective probabilistic assessments, or that the way probabilities enter their preferences depends on the nature of the uncertainty from which those probabilities are derived. Consider a choice problem where a person is told of the following two urns:[21]

Urn 1 100 balls, Red or Black (i.e. not told proportions).
Urn 2 100 balls, 50 Red, 50 Black.

Now suppose people are assigned to each of two choice situations, where they first choose one of the two urns, and then possibly receive $100, depending on the color of a ball randomly drawn from that urn:

Choice Situation A Must choose one of two gambles, a_1 or a_2:
a_1: Ball from Urn 1, $100 if Red, $0 if Black.
a_2: Ball from Urn 2, $100 if Red, $0 if Black.

Choice Situation B Must choose one of two gambles, b_1 or b_2:
b_1: Ball from Urn 1, $100 if Black, $0 if Red.
b_2: Ball from Urn 2, $100 if Black, $0 if Red.

---

[19] For formal models along these lines, see Loomes and Sugden (1982, 1987), Machina (1982, 1994), Luce and Narens (1985), Dekel (1986), Gilboa (1987), Weber and Camerer (1987), Fishburn (1988, 1989), Chew and Epstein (1989), Cox and Epstein (1989), Schmeidler (1989), Wakker (1989), Gul (1991), Tversky and Kahneman (1992), Epstein and Le Breton (1993), Shafir, Osherson and Smith (1993), Border and Segal (1994), and Tversky and Wakker (1995).

[20] Several recent surveys (see especially Camerer (1992b) and Harless and Camerer (1994)) have assessed how these different models fare. Harless and Camerer specifically address the relative gain in predictive power of different generalizations of expected utility. They provide a menu of theories that fare well in the sense (putting it roughly) that no alternative theory explains the data better without adding extra degrees of freedom. (A theory may be more consistent with observed data simply because it makes less precise predictions.) Their menu of four models which do well contains two generalizations of expected utility, expected utility itself, and *expected value*—expected utility theory where utilities are assumed to be linear (i.e., risk neutrality). (Expected utility is itself a mathematical and historical generalization of expected value theory.) Harless and Camerer (1994, p. 1286) conclude that there is a parsimony argument in favor of the expected-utility model when modeling choice between lotteries with the same "support" (i.e., the lotteries put positive probability on exactly the same set of outcomes). But they also conclude that, "When lotteries have different support, there is *never* a price-of-precision which justifies using expected utility; anyone who values parsimony enough to use expected utility over all of the generalizations should use expected value instead of expected utility." That is, for this class of choice situations, *no* preference economists might have for precision-of-prediction rather than behavioral accuracy can justify attachment to expected utility theory. (Hey and Orme (1994) present less clear-cut arguments, but conclude that "for many subjects, the superiority of several of the generalizations [of expected utility] is not established.")

[21] This is a variant of one of Ellsberg's (1961) experiments, and similar to the version presented by Einhorn and Hogarth (1987, pp. 43-4).

Experimental evidence indicates that many people prefer both a_2 over a_1 and b_2 to b_1.[22] Such revealed preferences (if assumed not to reflect pure indifference, which experimental variants have ruled out) are inconsistent with preferences that are a function of outcomes and probabilities alone. That is, the choice data are inconsistent with subjective expected utility theory, even if we generalize expected utility theory to allow for non-linear preferences. This is because the choice of a_2 over a_1 means that a person is choosing as if she believed that the percentage of red balls in Urn 1 is less than .5. But the choice of b_2 over b_1 is as if the person believes that the percentage of black balls in Urn 1 is less than .5.

Ellsberg and other researchers have concluded from this and other evidence that people are often *uncertainty-averse* or *ambiguity-averse*: People's dislike of known or objective uncertainties ("risk") is less severe than their dislike of unknown, ambiguous, or subjective uncertainties ("uncertainty"). In this example, people preferred a ball-color gamble based on Urn 2, with its known 50/50 distribution of red and black balls, to a gamble based on Urn 1, with its unknown distribution of ball colors.

Formal models have now been developed where people treat risk and uncertainty differently.[23] Some of these models capture uncertainty aversion by assuming a type of pessimism—people act as if, whatever they choose, uncertainty (but not risk) is likely to be resolved in a way that makes them worse off. In the previous example, it is as if people thought that if they picked Urn 1 when betting on a red ball they would likely draw a black ball, whereas if they bet on a black ball, they would more likely draw a red ball.[24]

---

[22] Research demonstrating behavior along the lines outlined here, more generally including research that explores the differential attitude people have towards objective risk versus uncertainty/ambiguity includes Hodges and Lehman (1952), Ellsberg (1961), Fellner (1961), Roberts (1963), Becker and Brownson (1964), Smith (1969), Sherman (1974), Yates and Zukowski (1976), Macrimmon and Larsson (1979), Kahneman and Tversky (1982b), Curley and Yates (1985, 1989), Hogarth and Kunreuther (1985, 1989, 1990), Curley, Yates and Abrams (1986), Frisch and Baron (1988), Gardenfors and Sahlin (1988), Camerer and Kunreuther (1989), Heath and Tversky (1991), Bernasconi and Loomes (1992), Sarin and Weber (1993), Camerer and Karjalainen (1994), Fox and Tversky (1995), and Tversky and Fox (1995).

[23] See Einhorn and Hogarth (1985, 1987), Bewley (1986), Segal (1987), Gilboa and Schmeidler (1989), Hogarth and Einhorn (1990), Dow and Ribeiro da Costa Werlang (1992, 1994), Sarin and Winkler (1992), Klibanoff (1993a, 1993b), and Epstein and Wang (1994). This notion that people feel differently towards "risk" than "uncertainty" is not new in economics. Knight (1921), for instance, famously differentiated risk from uncertainty, and part of his distinction may reflect the issues discussed here. See Bewley (1986) for such an interpretation, and an early formal model of uncertainty aversion.

[24] Further research suggests that preferences are more complex than the simple form of uncertainty aversion demonstrated by Ellsberg's experiment and variants. Although people do exhibit uncertainty aversion in domains where they feel "incompetent" in assessing probabilities, Heath and Tversky (1990) show that people *prefer* betting on uncertain events whose likelihood they feel competent assessing, rather than betting on known risk. (See also Fox and Tversky (1995).) These comparisons apply to equal-probability assessments, so results cannot be explained simply by the fact that people who feel competent believe they can predict the outcome better.

Social Preferences and Fair Allocations

It is common for undergraduates to encounter the following quote from *The Wealth of Nations* (Smith (1776, p. 26-7)) at the beginning of Economics 1:

> It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard for their own interest. We address ourselves not to their humanity, but to their self-love, and never talk to them of our necessities, but of their advantage.

There is not much to disagree with in Smith's poetic analysis of the motivations driving most market behavior, and no other two-word description of human motives comes close to "self interest" in helping us understand market or non-market economic behavior. Yet pure self interest is clearly not a fully adequate description of human motivation.[25] Realism suggests that economists should move away from the presumption that people are motivated *solely* by self interest. Dawes and Thaler (1988, p. 195) eloquently set parameters for this endeavor:

> In the rural areas around Ithaca it is common for farmers to put some fresh produce on the table by the road. There is a cash box on the table, and customers are expected to put money in the box in return for the vegetables they take. The box has just a small slit, so money can only be put in, not taken out. Also, the box is attached to the table, so no one can (easily) make off with the money. We think that the farmers have just about the right model of human nature. They feel that enough people will volunteer to pay for the fresh corn to make it worthwhile to put it out there. The farmers also know that if it were easy enough to take the money, someone would do so.

Virtually every researcher to examine behavior in experimental settings has concluded that preferences depart from pure self interest in non-trivial ways. It has been experimentally verified that people contribute to public goods more than can be explained by pure self interest, that those free to allocate money as they choose do not universally grab all the money, and that people sacrifice money to retaliate against unfair treatment. Debates remain: Some researchers have produced laboratory environments (e.g., well-organized, private-information spot markets) that induce behavior that is more self-interested than is behavior in other settings. But, disentangling debates over the nature of preferences from strong ancillary assumptions about what institutions and environments matter in the real world, there is no debate among behavioral researchers about whether underlying preferences depart non-trivially from pure self-interest.[26]

---

[25] Few Economics 1 students are exposed to Smith's other great book, *The Theory of Moral Sentiments* (1759), which very much emphasizes that self interest does *not* explain all human behavior.

[26] For experimental research demonstrating that people are not purely self-interested, see Goranson and Berkowitz (1966), Leventhal and Bergman (1969), Greenberg (1982), Kahneman, Knetsch, Thaler, Kunreuther, Luce, and Schweder (1986), Kahneman, Knetsch, and Thaler (1986a), Dawes and Thaler (1988), Thaler (1988), Caporael, Dawes, Orbell, and van de Kragt (1989), Ochs and Roth (1989), Boles and Messick (1990), Camerer (1990), Guth and Tietz (1990), Guth, Ockenfels, and Wendel (1991), Bolton (1991), Kachelmeier, Limberg, and Schadewald (1991), Kaplan and McCabe (1992), Camerer and Loewenstein (1993), Bolton, Katok, and Zwick (1993), Fehr, Kirchler, and Weichbold (1994), Weg and Smith (1993), Hoffman, McCabe, Shachat, and Smith (1994), Andreoni (1995), Cason and Mui (1995), and Camerer and Thaler (1995). For evidence from outside the laboratory, see

Examples of economic behavior induced by social goals are voluntary reductions of water-use during droughts, conservation of energy to help solve the energy crisis (as documented, for instance, in Train, McFadden, and Goett (1987)), and donations to public television stations. One context in which fairness has been studied is monopoly pricing (see, e.g., Thaler (1985) and Kahneman, Knetsch, and Thaler (1986a,1986b)). Might consumers see the conventional monopoly price as unfair, and refuse to buy at that price even when worth it in material terms? If this is the case, then even a profit-maximizing monopolist would price below the level predicted by standard economic theory. Finally, hundreds of researchers in psychology, industrial relations, and economics have investigated how equity, fairness, status-seeking, and other departures from self interest are important in employee behavior. Indeed, the massive psychological literature on equity theory was developed largely in the context of industrial relations; see, e.g., Adams (1963), Akerlof (1982), and Akerlof and Yellen (1990) for related arguments.[27]

The view that "social motivations"—concern by people for their effects on others—are unimportant in economic phenomena has diminished considerably in recent years.[28] In some domains, there have been prominent attempts (e.g., Becker (1981) on intra-family altruism) to incorporate social motivations into economic analysis. One form of social motivations on which economists have focused is that people are *altruistic* in the sense that they put positive value on the well-being of others. Roughly, this approach says that person 1 acts as if she is maximizing preferences of the form $U_1(x) \equiv (1\text{-}r)\cdot\Pi_1(x) + r\cdot\Pi_2(x)$, where $\Pi_1(x)$ is person 1's "material well-being" from outcome x, and $\Pi_2(x)$ is person 2's material well-being. By letting r be small, we can capture the idea that people are mostly self interested; but by assuming $r > 0$, we can investigate when and how concern for others affects behavior and welfare.[29]

Simple altruism does parsimoniously capture important phenomena, and is psychologically valid in many contexts. But there is a mass of psychological evidence—and, more recently, experimental economic evidence—that indicates it is often an importantly wrong model of social preferences.[30] To get a sense for how social

---

Finn and Lee (1972), Lord and Hohenfeld (1979), Blinder and Choi (1990), Lynn and Grassman (1990), Hollinger, Slora, and Terris (1992), Cowherd and Levine (1992), Levine (1993a, 1993b), Bewley and Brainard (1993), and Flinn (1995).

[27] For formal models of departures from self interest, see Farber (1981), Margolis (1982), Bolton (1991), Issac, Mathieu, and Zajac (1991), Huang and Wu (1992), Rabin (1993), Konow (1993), Ruffle (1995), Sally (1995), Meidinger (1996), and Rabin (1996); for other arguments about the economic implications of departures from self interest, see Akerlof (1982), Akerlof (1984), Kahneman, Knetsch and Thaler (1986a, 1986b), Lazear (1989), Akerlof and Yellen (1990), Kachelmeier, Limberg, and Schadewald (1991), Levine (1991), Romer (1992), Huang and Wu (1992), Shefrin and Statman (1993), and Fehr, Kirchler, and Weichbold (1994).

[28] There have, of course, always been economists (beginning with Smith and Walras) who have been highly attentive to other aspects of human nature. For the last several decades, however, virtually all teaching, research, and training in formal economics has been based on the assumption of pure self interest.

[29] See Collard (1975) and Margolis (1982) for examples economists' incorporating the role of altruism in economics.

[30] See Krebs (1970, 1982) for some psychological evidence on altruism and need-based helping motives. Even where simple altruism may adequately describe behavior (e.g., in donating to charity), psychological research may be of value for welfare economics. In particular, research has explored whether people who help others do so for

preferences differ from simple altruism, I turn to a (far from complete) account of what can be called "behavioral distributive justice": How do individuals choose to divide resources among themselves and others?[31] There are two aspects to this question: First, what do people, when disinterested, feel are proper rules for allocation? Second, to what degree do people sacrifice self interest for the sake of these principles? Very roughly, imagine that person 1's utility function takes the form $U_1 = (1-r)\cdot\Pi_1 + r\cdot W_1(\Pi_1, \Pi_2)$, where W_1 is person 1's view of the proper allocation, and Π_1 is (as above) her self-interested payoff. To understand the implications of fairness and justice for behavior, we need to know both the nature of the W_1 function, and the level of r.

To address the question of disinterested assessments, suppose two people together find $10 on the ground. How would the average person, in her role as a third party, decide to split the money between the two?[32] One answer, following from the simple-altruism perspective, is that the person who is poorer, or who can otherwise benefit most from the money, should get it. There is no doubt that people often consider comparative need in allocation decisions. There are similar norms about how to allocate goods whose usefulness is different for the two parties: We often find it appealing to allocate goods to maximize the benefit of each good.

A striking and robust finding, however, is that many people in many contexts do not find the "maximal-benefits" criterion attractive. One simple alternative norm is prevalent: "Pies" should be split 50-50. Ignoring issues of relative usefulness except in extreme cases, we often feel that goods should be divided equally. But the maximal-benefits criterion fares even worse than this. Many people feel goods should be allocated according to a "maximin" criterion which equalizes welfare improvements between the two people. That is, disinterested people seem often to maximize preferences of the form $W_0 \equiv \text{Min}[\Pi_1, \Pi_2]$, where Π_1 and Π_2 are the gains in utility from dividing resources. This norm typically implies that *more* than half the resources are allocated to the person who values those resources *less*.

Consider the following hypothetical situation that Yaari and Bar-Hillel (1984, p. 8) posed to 163 subjects:

"truly" altruistic reasons, in the sense that the actions they take will lower their experienced well-being; or whether they do so "only" to alleviate painful guilt, etc. Do you help a stranger when inconvenient because it is the right thing to do even though it makes you worse off? Or do you *like* bringing joy to others in the same way you like eating apples? Or do you know you will have unpleasant, guilt-induced nightmares if you don't help? These distinctions will be important in welfare analysis: Even if we think behavior is described by $\text{Max } U_1(x) \equiv (1-r)\cdot\Pi_1(x) + r\cdot\Pi_2(x)$, we must still resolve whether Person 1's experienced well-being is described by $U_1(x)$ or $\Pi_1(x)$. See Krebs (1970) for a relevant discussion.

[31] I use the term "behavioral distributive justice" to emphasize that I am reviewing evidence on how people actually feel about distributive justice, *not* normative or philosophical questions of what is the proper notion of distributive justice. Economics has a long history of assuming pure self interest among economic actors in the domain of positive economics, while debating the relative merits of different social welfare functions in the domain of normative economics. Consequently, it is sometimes hard to focus clearly on the unfamiliar question of how economic actors themselves judge proper social allocations without injecting our own views.

[32] I assume the money is found so as to consider the thought experiment that neither party *deserves* the money more than the other. Desert will obviously be relevant in many situations—and there is a massive psychological literature on "equity theory" which shows that people feel that those who have put more effort into creating resources have more claim on those resources. See, e.g., Berscheid, Boye, and Walster (1968), Brickman and Bryan (1976), Greenberg and Leventhal (1976), Greenberg (1982), Krebs (1982), Harris (1983), Camerer and MacCrimmon (1983), Cook and Yamagishi (1983), Mellers (1985), and Bigman (1991).

Q1: A shipment containing 12 grapefruits and 12 avocados is to be distributed between Jones and Smith. The following information is given, and is known also to the two recipients:

- Doctors have determined that Jones's metabolism is such that his body derives 100 milligrammes of vitamin F from each grapefruit consumed, while it derives no vitamin F whatsoever from avocado.

- Doctors have also determined that Smith's metabolism is such that his body derives 50 milligrammes of vitamin F from each grapefruit consumed and also from each avocado consumed.

- Both persons, Jones and Smith, are interested in the consumption of grapefruit and/or avocados only insofar as such consumption provides vitamin F – and the more the better. All the other traits of the two fruits (such as taste, calorie content, etc.) are of no consequence to them.

- No trades can be made after the division takes place.

How should the fruits be divided between Jones and Smith, if the division is to be just?

Grapefruits are worth more to Jones than to Smith, and avocados are worthless to Jones. The "socially efficient," metabolism-maximizing allocation, therefore, is for Jones to get all the grapefruits, and Smith to get all the avocados. Indeed, such an allocation would seem to accord somewhat to the fifty-fifty norms (though less so than the allocation where they both get six of each fruit). The "maximin allocation" would be to give eight grapefruits to Jones, and give the remaining grapefruits and all the avocados to Smith.

Subjects were given a menu of five different allocations, and asked to state which allocation they found the most "just."[33] Yaari and Bar-Hillel (1984, p. 10) report the percentage of respondents who chose each of the five allocations. The five allocations are denoted by the number of grapefruits and avocados to be distributed to Jones and Smith (denoted by their initials):

Distribution	% of respondents
(J:6-6, S:6-6)	8
(J:6-0, S:6-12)	0
(J:8-0, S:4-12)	82
(J:9-0, S:3-12)	8
(J:12-0, S:0-12)	2

The vast majority of respondents chose to equalize welfare gains (J:8-0, S:4-12), rather than maximize total welfare gains (J:12-0, S:0-12). Since the maximin solution is subtle, the results strongly suggest subjects thought about the problem, and did not merely choose some simple focal point.

Yaari and Bar-Hillel (1984, p. 11) then tested the robustness of the maximin criterion by posing a variant of the original question, where subjects were told that Smith derives only 20 milligrammes of Vitamin F from both fruits, rather than 50 milligrammes. The responses were as follows:

---

[33] Subjects were also asked to conjecture how Jones and Smith themselves would allocate the fruit, "on the assumptions that both recipients are committed to looking for a just division." Yaari and Bar-Hillel report that "the difference between the distribution of responses to these two variants are negligible."

Distribution	% of respondents
(J:6-6, S:6-6)	4
(J:4-0, S:8-12)	82
(J:6-0, S:6-12)	4
(J:9-0, S:3-12)	7
(J:12-0, S:0-12)	3

Subjects supported the maximin criterion just as strongly—now imposed at a greater cost to total social benefits. Indeed, Smith is now given more than half the grapefruits, though they are far less valuable to him or her than they are to Jones. Yaari and Bar-Hillel (1984, pp. 11) then push the limits of the maximin criterion, positing that, "Sooner or later,...[the maximin criterion] runs the risk of becoming morally unsound." They told subjects that Smith derives only 9.1 milligrammes of Vitamin F from both fruits, yielding the following responses:

Distribution	% of respondents
(J:6-6, S:6-6)	17
(J:2-0, S:10-12)	38
(J:6-0, S:6-12)	27
(J:8-0, S:4-12)	6
(J:12-0, S:0-12)	12

Thus, there are almost surely limits to the maximin criterion, and a far greater number of respondents now seemed willing to tolerate unequal welfare gains. Yet, even though grapefruits are 11 times more valuable to Jones than Smith, *still* only 12% of respondents felt that all 12 grapefruit should be given to Jones. Moreover, only 18% thought that Jones should get more than half the grapefruits while 38% thought Smith should get more than half.[34]

Yaari and Bar-Hillel (1984) study what disinterested people consider a proper allocation rule. To address the question of how people trade off self interest against justice, we need a situation where the allocator is *not* disinterested: If, for instance, an allocator must unilaterally choose how to divide money between herself and a second party, choices will depend both on what the person feels is a just allocation and on how much she values a just outcome relative to her self interest.

Andreoni and Miller (1996) consider such a situation of unilateral, interested choice.[35] They asked each subject to unilaterally allocate money between herself and a second party. Subjects were given various "exchange

---

[34] Another variable that contributes importantly to the use of the maximin criterion is whether the goods being allocated are perceived as *needs*, or merely *wants*. In a variant of the first question discussed above, Yaari and Bar-Hillel (1984, p. 12) asked subjects to suppose that Jones merely values grapefruit at $1.00 per pound and detests Avocados, and Smith values both fruits at $.50 per pound. Half the subjects still preferred an allocation within one grapefruit of the maximin allocation, but now 35% advocated the efficient allocation giving Jones all the grapefruits.

[35] Andreoni and Miller's and other monetary-stakes experiments are also useful for allaying concerns one might have regarding the Yaari and Bar-Hillel survey, which asked subjects to say what they considered the *just* allocation. If subjects considered "justice" to be only one component of a proper allocation rule, they may not have interpreted the question as meaning "How would you choose to allocate between these two parties?" Andreoni and Miller (1996) and bargaining experiments typically do not prompt subjects to evaluate allocations according to any criterion—but merely to make a choice.

rates" for allocating money; e.g., in some conditions subjects were told that for every dollar they gave up, the other (anonymous) party would get $3. Andreoni and Miller (1996, p. 10) found that over half the subjects behaved in a way significantly inconsistent with pure self interest. Among those whose behavior was most strongly disinterested, ten chose approximately the minimax allocation rule, and five chose approximately the dollar-maximizing allocation. While the authors note that these classifications are crude, and must be interpreted with caution, these results suggest that the minimax rule resonates with many people even when allocating money, and even when self interest is at stake.[36]

In moving from abstract, context-free allocation problems to everyday economic fairness judgments, things become significantly more complicated.[37] First, as elsewhere, reference levels play a huge role in the domain of fairness. Thaler (1980) and Kahneman, Knetsch, and Thaler (1986a,1986b) demonstrate that loss aversion plays a very strong role in people's notion of fairness; firms have more of an obligation not to hurt workers or customers relative to reference transactions than they have to improve the terms of trade. There is also evidence that people's general perceptions of fair behavior can adjust dramatically over time. Kahneman, Knetsch, and Thaler (1986b, p. 730) argue that, "Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction. Thus, the gap between the behavior that people consider fair and the behavior that they expect in the marketplace tends to be rather small." Franciosi, Kujal, Michelitsch, Smith, and Deng (1995) experimentally support this hypothesis by testing reactions to unfair price increases in a laboratory posted-offers market, showing that the role of fairness considerations in price-determination diminishes upon repetition.[38] In competitive spot markets, people may eventually come to believe that the prevailing market price is fair. Because adjustments of fairness judgments are not immediate, however, fairness considerations may help explain the sort of medium-run stickiness studied by macroeconomists.[39]

Finally, in attempting to capture behavioral findings with models of social preferences, it is important to note that it only makes sense to model fair-allocation decision by invoking social preferences if we assume that those preferences are defined over shares of the *given* pie people are considering. People implicitly consider equitable sharing over *changes* in total endowments, not total endowments themselves. Apparently, people generally have a

---

[36] Allowing proportions of a "pie" to have different monetary values to different parties has also been a theme in the experimental bargaining research. In these settings, disentangling self interested bargaining strategies from preferences for just allocations is quite difficult, but results seem similar to the findings of Andreoni and Miller (1996). See, e.g., Roth and Murnighan (1982) and Kagel, Kim, and Moser (1992).

[37] For examples of further research on perceptions of fairness and justice, see Brickman and Bryan (1975), Krebs (1982), Cook and Messick (1983), Cook and Yamagishi (1983), Kahneman, Knetsch and Thaler (1986a,1986b), Mellers and Hartka (1989), Loewenstein, Thompson, and Bazerman (1989), Miller (1992), Camerer and Loewenstein (1993), Mitchell, Tetlock, Mellers, and Ordonez (1993), and Bazerman, White, and Loewenstein (1995).

[38] See also Binmore, Morgan, Shaked, and Sutton (1991) for a related finding. But many other experiments find virtually no change in either behavior or perceptions of fairness over time; see, e.g., Fehr and Falk (1996).

[39] And it also bears emphasizing that the logical and empirical evidence that market outcomes are likely to correspond to the self-interested outcomes pertains solely to competitive spot markets. See e.g., Fehr, Kirchler, and Weichbold (1994) for an experimental study of labor markets where behavior never converges to the self-interested outcome.

one-pie-at-a-time conception of fair-division problems, rather than a more global notion. This is not an insurmountable tendency—if people are presented with several allocation problems together, they will likely attend to the overall implications of their several choices. But the fact remains that any attempt to capture behavioral norms of fairness and distributional justice with formal models of social preference must confront the "piecemeal" nature of these norms.[40]

Reciprocity and Attribution

The previous subsection considered evidence about social preferences defined over the allocations of goods. Psychological evidence indicates, however, that social preferences are not merely a function of consumption levels, or even changes in consumption levels: People do not seek uniformly to help other people, nor do they seek uniformly to share equitably. Rather, people's concerns regarding the experiences of other people depend on the behavior, motivations, and intentions of these other people. The same people who are altruistic towards deserving people are often indifferent to the plight of undeserving people, and often motivated to hurt those whom they believe to have behaved egregiously. If somebody is being nice to you or others, you are inclined to be nice to him. If somebody is being mean to you or others, you are inclined to be mean to him.

This "reciprocal" nature of preferences manifests itself in the distinction between simple altruism, as outlined earlier, and *reciprocal altruism*.[41] Consider the question of why people conserve water during a drought. Clearly they perceive that conservation contributes to the general good, which at a small cost is something they eagerly do. How might we model such preferences? First note that probably there are diminishing social benefits of conservation because the marginal social value of water is greater the less water there is. If other people conserve, it is less urgent for you to do so; if other people don't conserve, it is more urgent for you to do so. If you were a simple altruist, therefore, learning that others were not conserving should cause you to intensify your conservation efforts—if nobody else conserves, it becomes all the more urgent that you do. This prediction does not accord with

---

[40] Preferences defined over final wealth states cannot plausibly explain rules such as 50/50 sharing or the maximin criterion. With plausible levels of either inequality in or uncertainty about *initial* endowments entering any moderate size division-of-the-pie situation, *any* social welfare function defined with respect to overall consumption levels will yield all-or-nothing allocations: For each good, either one person or the other should get all of it. This is true even if people use the maximin criterion applied to overall consumption levels. The point is that the initial endowments are *very* unlikely to be such that any preference over final allocations will deem the marginal value of giving a good to one party to be virtually equal to the marginal value of giving it to the other party. For instance, using the maximin criterion over *total* wealth in distributing $10 found on the sidewalk will only lead to the money being shared if the initial wealth levels of the parties differ by less than $10.

[41] For experimental evidence related to reciprocal altruism, see Goranson and Berkowitz (1966), Frisch and Greenberg (1968), Schopler and Thompson (1968), Lerner and Lichtman (1968), Leventhal, Weiss, and Long (1969), Greenglass (1969), Nemeth (1970), Wilke and Lanzetta (1970), Greenberg, Block, and Silverman (1971), Savitsky and Babl (1976), Greenberg and Leventhal (1976), Akerlof (1984), Goetze and Orbell (1988), Hollinger, Slora, and Terris (1992), Stuckless and Goranson (1992), Fehr, Kirchler, and Weichbold (1994), and Croson (1995).

actual sentiments toward cooperation. People are more inclined to conserve energy or water if they think other people are doing their share, but *not* if they think that others are not doing their part. People reciprocate public spiritedness in others rather than counteract it.

Reciprocal altruism manifests itself most clearly in the prisoner's dilemma, illustrated in Figure 3:

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	4,4	0,6
	Defect	6,0	1,1

Figure 3

If it is common knowledge to the players that they are playing (Cooperate,Cooperate), then each player knows that the other is sacrificing her own material well-being in order to help her. So long as the material gains from defecting are not too large, (Cooperate, Cooperate) is likely to be an "equilibrium" given people's taste for reciprocal altruism. (Defect, Defect) is *also* likely to be an equilibrium, however, because if it is common knowledge that they are playing (Defect, Defect), then each player knows that the other is not sacrificing. The prisoner's dilemma illustrates that we cannot fully capture realistic behavior by invoking simple altruism: The prediction that Player 1 might cooperate if she believes the other will, but will never cooperate if she believes the other won't, implies that Player 1 must care differentially about helping (or hurting) Player 2 as a function of what she believes about Player 2's behavior.

The voluntary provision of public goods has been studied extensively by psychologists and experimental economists.[42] Many of these experiments hint that contributions towards public goods are not the result of simple altruism, though the evidence for reciprocal altruism is varied, and often indirect. When subjects in public-goods experiments are debriefed about their beliefs, there seems to be strong positive correlation between subjects'

[42] Laboratory experiments of public goods have been conducted by, among others, Marwell and Ames (1981), van de Kragt, Dawes, and Orbell (1983), Isaac, Walker, and Thomas (1984), Kim and Walker (1984), Isaac, McCue, and Plott (1985), Andreoni (1988a,1988b), and Isaac and Walker (1988a,1988b). For a survey of public goods experiments (not focused primarily on reciprocal altruism), see Ledyard (1995). The evidence from public-goods experiments is that people cooperate to a degree greater than would be implied by pure self-interest. Dawes and Thaler (1988) conclude that, for most experiments of one-shot public-good decisions in which the individually optimal contribution is close to 0%, the contribution rate varies between 40% and 60% of the socially optimal level. Many of these experiments are problematic in that their null hypothesis of completely self-interested behavior corresponds to zero contributions, where zero contributions is also the most extreme behavior subjects could exhibit. Therefore, all departures from full rationality are necessarily in the direction of "generous" behavior. Andreoni (1995) shows that, by a very conservative estimate, at least half the contributions to public goods are intentional rather than "errors." See also Keser (1996) for evidence that generous contributions are not merely an artifact of experimental design.

contribution levels and their beliefs about how others were contributing.[43] Other indirect evidence is that pre-decision communication greatly enhances cooperation.[44] One of the many reasons why communication might enhance contributions is that reciprocal-altruism preferences essentially turn public-goods situations into "coordination games," where high contributions are efficient equilibria, and low contributions are inefficient equilibria. Consequently, as in general coordination games, pre-game communication helps players coordinate on the efficient equilibria.[45]

The most direct evidence of reciprocal altruism in the prisoner's dilemma with which I am familiar comes from Shafir and Tversky (1992). When subjects were told that their anonymous partner in a Prisoners' Dilemma had cooperated, 16% also cooperated; when subjects were told that their partner did not cooperate, only 3% cooperated.

Reciprocity motives manifest themselves not only in people's refusal to cooperate with others who are being uncooperative, but also in their willingness to sacrifice their own well-being to hurt others who are being unfair. A consumer may refuse to buy a product sold by a monopolist at an "unfair" price, even if she hurts herself by foregoing the product. An employee who feels he has been mistreated by a firm may engage in costly acts of sabotage, perhaps to the point of violently retaliating against his employers. Members of a striking labor union may strike longer than is in their material interests because they are angry, or want to punish a firm for being unfair. These and many more examples make clear that no conception of social preferences that assumes people adhere universally to the "Pareto principle" (always preferring outcomes that make everyone better off) will be accurate.

This idea that people are motivated to retaliate when they feel they have been mistreated is a fairly obvious intuition, and well understood by psychologists.[46] More recently, it has been widely explored by experimentalists who have investigated many variants of the "ultimatum game."[47] The ultimatum game consists of two people splitting some fixed amount of money according to the following rules: A Proposer offers some division of (say) $10 to a Decider. If the Decider accepts, they split the money according to the proposal. If the Decider rejects,

---

[43] See, e.g., Croson (1995). Some psychologists have argued, however, that part of the correlation runs in the opposite direction than implied by the reciprocal-altruism hypothesis. Free riders may merely be justifying their actions by perceiving or claiming that others aren't contributing; or the correlation could be entirely spurious if it merely stems from the fact that those type of people who don't cooperate also believe others won't.

[44] See Sally (1995) for an excellent review.

[45] Indeed, the strong emphasis on the *in*ability of the two captured prisoners to communicate with each other in the standard description of the prisoner's dilemma indicates that we implicitly believe that the prisoner's dilemma really amounts to a coordination game. If defecting were truly a dominant strategy, pre-game communication would not matter.

[46] For research on retaliation, see, e.g. Berscheid, Boye, and Walster (1968), Kahneman, Knetsch, Thaler (1986), Ochs and Roth (1989), Guth and Tietz (1990), Friedland (1990), Akerlof and Yellen (1990), Carter and Sunderland (1992), Stuckless and Goranson (1992), Huang and Wu (1992), Weg and Smith (1993), Camerer and Thaler (1995), Flinn (1995), and Blount (1995).

[47] The ultimatum game was first developed by Guth, Schmittberger, and Schwarze (1982). For reviews of the (massive) literature developed since, see Thaler (1988), Guth and Tietz (1990), and Camerer and Thaler (1995). See also and Roth (1995) for a broader survey of bargaining experiments.

they both get nothing. The result of rational self-interest is clear: Proposers will never offer more than a penny, and the Decider will accept any offer of at least a penny. Experiments clearly refute such behavior: Even in one-shot settings, Deciders are willing to punish unfair offers by rejecting them, and Proposers tend to make fair offers. The decision by Proposers to make fair offers can come from at least two motivations: Proposers themselves may have a preference for being fair, or self-interested Proposers might correctly predict that Deciders will retaliate against unfair offers by rejecting them.

A crucial feature of the psychology of reciprocity is that perceived *intentions, motives,* and *volition* are fundamental to how people react to given behavior by others. People determine their dispositions towards others according to motives attributed to these others, not solely according to actions taken. When motivated by reciprocal (rather than simple) altruism, people differentiate between those who take a generous action by choice and those who are forced to do so. Demonstrating both the basic principle of reciprocity, and the role of volition, Goranson and Berkowitz (1966, p. 229) conducted an experiment in which confederates posing as subjects were in a position to help the true subjects fill out some worksheets:

> One third of the subjects were told that the [confederate] had voluntarily offered to help the subjects with her task (*voluntary-help* treatment). In this treatment, after the experimenter left the room the confederate entered and appropriately volunteered to work on one sheet. Another third of the subjects were told that the experimenter would *instruct* the [confederate] to come in and work on one of the subject's sheets (*compulsory-help* treatment). For these subjects the confederate came in and took one sheet. For the final third of the subjects, the experimenter said that the [confederate] *might* be willing to help the subject. However, in this condition the confederate came in and pointedly refused to help the subject (*help-refused* treatment).

When the subjects were later given an opportunity to assist the confederates, they reciprocated the earlier help, and did so significantly more when the earlier help was voluntary than when it was involuntary.[48]

Volition is also central to the propensity to retaliate against negative actions. This can be brought into focus by considering a simplified version of the ultimatum game, written on the left of Figure 4. The Proposer is given the choice between offering an \$8/\$2 split of \$10 or a \$5/\$5 split. The Decider then can either accept or reject the offer.

[48] For papers demonstrating the role of intentionality in willingness to reciprocate good or bad deeds, see Goranson and Berkowitz (1966), Lerner and Lichtman (1968), Schopler and Thompson (1968), Leventhal, Weiss, and Long (1969), Greenglass (1969), Wilke and Lanzetta (1970), Nemeth (1970), Greenberg, Block, and Silverman (1971), Greenberg and Frisch (1972), Carter and Sunderland (1992), and Blount (1995).

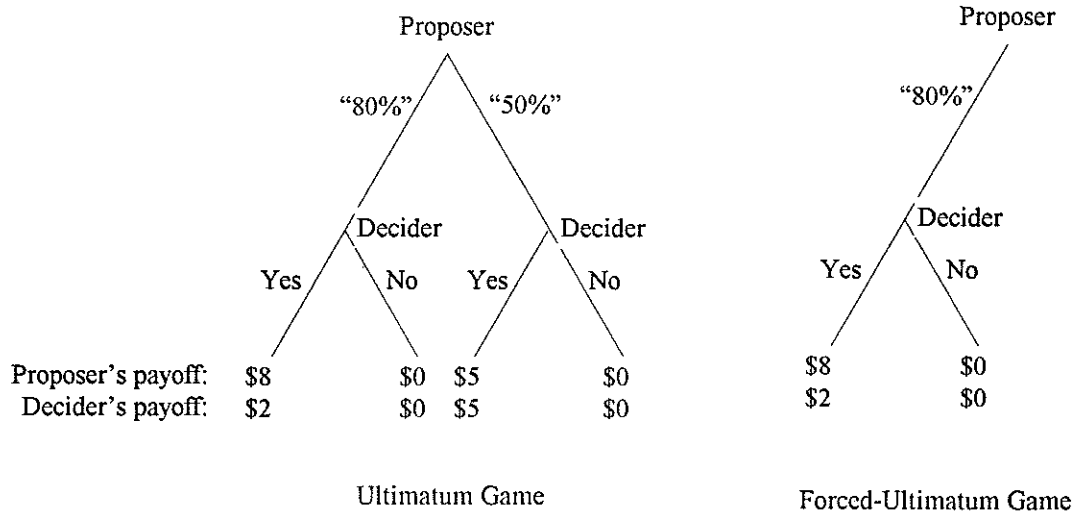

Figure 4

As noted, most people turn down mean offers in ultimatum games; in this version of the ultimatum game, it can be inferred that many Deciders would turn down \$8/\$2 offers. This behavior could be explained in part by an aversion to unequal divisions, due either to a principle of fairness, or to the Deciders' envy, where the Deciders prefer the outcome \$0/\$0 to \$8/\$2 because they don't like to walk away with less money than the Proposers.[49] Yet this does not seem an adequate explanation of the Deciders' motivation in turning down offers. Rather, it seems more likely that they don't like the idea of Proposer having *chosen* to make an unfair offer, and hence retaliate. This interpretation that rejection of unfair offers is an act of retaliation relies on the fact the Proposer *chose* the unfair offer. To highlight this issue, consider the variant of the ultimatum game on the right of Figure 4. In this "Forced-Ultimatum Game" the Decider knows that the Proposer has been *forced* to make an offer; that is, the Proposer is *only* allowed to choose the left branch of the Ultimatum Game. How would the Decider react in this situation? It is likely that she would be less prone to retaliate: The only reason to reject an offer in such a situation is out of pure aversion to the unequal outcome, not because of any blame for the unfair treatment by the Proposer.[50]

---

[49] See Bolton (1991) for such an interpretation. For research on the importance of envy in economics, see also Foley (1967), Brennan (1973), Varian (1974), Baumol (1987), Brenner (1987), Banerjee (1990), Elster (1991), Mitzkewitz and Nagel (1991), Kirchsteiger (1992), and Mui (1995).

[50] I am unaware of any experimental result on precisely such an example. A related series of experiments were conducted by Blount (1995), however. Subjects were asked about their willingness to accept offers in an ultimatum game, knowing that their chosen reaction would actually be used to determine their payoff as a function of offers made by anonymous other parties. One group of subjects was told that the "ultimatum" was coming from anonymous other students, and that their responses would determine the division between them and these anonymous other students. In a second variant, subjects were told that *a third party* (also an anonymous student) was to determine the offer made. In this variant, the "Proposer," who would be hurt by a subject's decision to reject an offer, did not participate in the offer, and the third party who made the offers would not be affected one way or another by the subject's decision. In a final variant, subjects were told that the offer would be generated *randomly* (by a computer-simulated roulette wheel). In one study, Blount (1995, p. 135) reports that "the mean minimum acceptable offers were, respectively, \$2.91, \$2.08, and \$1.20." That is, people did reject very low offers

Frank (1994, p.21), tells the following colorful story to illustrate the role of volition:

> There is an often told story of a boy who found two ripe apples as he was walking home from school with a friend. He kept the larger one for himself, and gave the smaller one to his friend.
>
> "It wasn't fair to keep the larger one for yourself", the friend replied.
>
> "What would you have done?" the first boy asked.
>
> "I'd have given you the larger one and kept the smaller one for myself," said the friend. To which the first boy responded, "Well, we each got what you wanted, so what are you complaining about?"

The punch line of this story is the obviously silly and counter-intuitive notion that the second boy's satisfaction with events should be judged solely by the allocation arising from those events. Yet this comical presumption is the basis of most attempts by economists to model "social preferences" defined solely over outcomes.

In attempting to capture these examples in formal models of social preferences, we can conceptualize volition in large part as a function of a person's choice of actions among those choices she perceives she has available. The role of intentions goes even further, however, when considering the choice of actions among those available. Suppose, for instance, you are eating dinner with parents and small child when all of a sudden your hands flail across the table and knock a pitcher of water all over the child. How do the parents react? If they thought your goal was to spill water on the child, they are probably angry. If they thought you were worried that pitcher was precariously perched next to the child, and that your flailing arms were an attempt, albeit uncoordinated, to *prevent* a spill, probably they are less angry.[51]

The story indicates that even "choice sets" don't capture all aspects of volition. Our beliefs about other peoples' motives depend on what we believe *their beliefs* are. Consider somebody who has been generous to you. Suppose that you think he was nice to you solely to get a much bigger favor from you in the future.[52] Then you do

---

even if computer or third-party generated (providing evidence that there is a component of pure "inequality aversion" in rejections of offers in the ultimatum game), but subjects were far less likely to turn down offers which were not the result of volition by the person who would be hurt by the rejections. (Blount reports that the difference in means between the first and second conditions were not statistically significant, however, and she conducted follow up studies to investigate under what circumstances people would be more likely to treat these conditions differently.)

[51] I have confirmed this hypothesis in a field study conducted in North London. The child's reaction depends on his stage of moral development. Piaget (1965) and Kohlberg (1963, 1983) posit that at a certain stage of moral development, children realize that not all good and bad outcomes are intentional, and that moral judgment depends on whether the consequences they induced were intentional or not. As Skolnick (1986, p. 394) puts it, "The more advanced stage [the second stage], called subjective or autonomous morality, develops in the middle years. During this stage children have a less rigid and absolute view. They recognize ... that moral judgments are made in terms of people's intentions, as well as standards of fairness and justice. Piaget believed that this second type of moral reasoning was not usually achieved before the ages of 12 or 13."

[52] For examples of this type of thought process, see Kahn and Tice (1973) and Charness (1996a). Kahn and Tice found that subjects' reactions to others' statements of intentions depended on whether they thought that those stating their good intentions knew that their intentions would be made known to the subjects. In Charness's (1996a) study, subjects bargained with other subjects whose previous behavior in simple divide-the-pie games they had observed. When subjects knew that their bargaining partners were generous previously, bargaining was more conciliatory. But across two experiments, it was shown that if the previous generosity was exhibited by subjects

not view his generosity to be as pure as if he had expected no reciprocity from you. To capture your reaction to generosity, therefore, we need to know what you believe he believes your reaction will be.[53]

The concern for intentions, not simply outcomes, is one component of the massive psychological literature on the distinction between "procedural justice" and "distributional justice" (see, e.g. Tyler (1988)). When our employer or our court system makes rulings against us, it is not always just a matter of whether we like the outcome, but whether we thought we got a "fair shake"—were decision makers *trying* to be just? While this issue of institutional intentions is not the only issue addressed in the literature on procedural justice, it does seem to explain some of the views people have regarding appropriate institutions for dispensing justice.

The role of reciprocity and volition appears in some important economic contexts. Akerlof (1982) posited that firms and workers can be thought to engage in "gift exchange," a view of social exchange emphasized in sociology and especially anthropology.[54] If a firm pays a higher wage to an employee, that employee is likely to reciprocate with higher quality work. Consequently, firms may give higher wages hoping workers will reciprocate with higher effort. Similarly, Akerlof (1982, 1984) and Akerlof and Yellen (1990) propose that "efficiency wages," above the market-clearing wages, will be paid to workers to induce higher effort by those workers.[55]

Fehr, Kirchsteiger, and Riedl (1993) tested this hypothesis in laboratory models of labor markets. Subjects were assigned roles as "firms" and "workers." Firms offered a wage, and workers responded by choosing an "effort" level; both the chosen wages and effort levels translated into real monetary rewards for the subjects in the obvious ways: A higher wage lowered firms' payoffs. Additionally, effort was costly, so that the money-maximizing worker effort level in response to any given wage was the minimum possible effort. However, most workers chose effort levels higher than their money-maximizing levels. Moreover, low wages induced little or no effort by workers. But they had a propensity to reciprocate high wage offers made by employers by providing high effort themselves, which decreased their money payoffs but increased total money payoffs, and rewarded firms for setting high wages.

What is the source of high effort levels by workers in response to high wages by firms? One hypothesis is that workers simply choose to share some of their additional wealth from higher wages with the firm. A second explanation relates to the attribution hypotheses discussed above: Workers may respond to the intentional generosity of firms with generosity. Charness (1996) conducts experiments that helps us differentiate these hypotheses, and which highlight the role of attribution in social behavior. In Fehr, Kirchler, and Weichbold (1994), it is clear to the worker-subjects that the firms choose wages of their own volition. Charness (1996)

---

who knew their behavior would be observed, bargaining was less conciliatory than when it appeared that the previous generosity was exhibited without any possible strategic advantage.

[53] To formalize these volitional aspects of fairness judgments, therefore, Rabin (1993) adopts the framework developed by Geanakoplos, Pearce, and Stacchetti (1989), who modify conventional game theory by allowing payoffs to depend on players' beliefs as well as their actions.

[54] For other examples discussing the role in labor economics of fairness and related issues, see Akerlof (1984), Akerlof and Yellen (1990), Baron (1988), Bishop (1987), and Levine (1991, 1993a, 1993b), and Romer (1992).

[55] Akerlof and Yellen posit that this can be based on perceived fairness by workers, providing an alternative explanation to the qualitatively similar theory of efficiency wages developed by Shapiro and Stiglitz (1984).

replicates this condition, but also conducts variants of the experiment where it is clear that firms do *not* choose wages of their own volition—wages are either chosen randomly, or by a "third party" (the experimenter). In these conditions, a high wage is not an act of kindness by a firm, and a low wage is not act of meanness; both are beyond a firm's control.

All conditions provided workers with the same monetary-payoff schedule as a function of wages and efforts. The results indicate that the high-wages-yields-high-effort reaction has both a "share-the-wealth" and an attribution element. Workers were substantially more likely to reward high wages with high effort and punish low wages with low effort, when the wages reflected the volition of the firm.[56]

3. Biases in Judgment

Economists have traditionally assumed that, when faced with uncertainty, people correctly form their subjective probabilistic assessments according to the laws of probability. But researchers have documented many systematic departures from rationality in judgment under uncertainty. A feel for this research is well expressed in Tversky and Kahneman (1974, page 1124), which serves as a *de facto* introduction to Kahneman, Slovic, and Tversky (1982), an influential volume containing many of the major contributions to behavioral research on heuristics and biases:

> How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.
>
> The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability.

Behavioral research on judgment under uncertainty teaches us that people are not nearly as rational as economists assume. But it teaches us much more. In this passage, the word "error" does not appear without the

word "systematic". This literature does not merely teach us that people aren't perfectly rational; it proposes specific ways that judgment departs from the rational model.[57]

While research on heuristics and biases is not as amenable to incorporation into mainstream formal economics as the material presented in the previous section, this literature already contains some clear and general insights. For the remainder of this section, I describe some of this research. I concentrate on just a few particular heuristics and biases, and then more quickly outline some others. I conclude by discussing briefly some of the evidence for how people do and don't learn to correct biases.[58]

The Representativeness Heuristic

We think a person is more likely to be a member of some group if that person is similar to the typical member of that group. If a man behaves more like a criminal (shifty eyes, etc.), then we think it is more likely he is a criminal. This simple intuition is, of course, captured by Bayes Law.[59]

Research on the *representativeness heuristic*, however, demonstrates that people tend to *over-use* "representativeness" in assessing probabilities. Bayes' Law tells us that our assessment of likelihoods should combine representativeness with base rates (the percentage of the population falling into various groups). Yet people under-use base-rate information in forming their judgments. If we see somebody who looks like a criminal, our assessment of the probability that he is a criminal tends to under-use knowledge about the percentage of people who are criminals. Similarly, if a certain medical test always comes out positive among people with a rare disease, and only occasionally among people without the disease, people will tend to exaggerate the likelihood of having the disease given a positive result. Given the rarity of the disease, the total number of false positives may be far greater than the number of true positives.

---

[56] The results comparing third party versus firm volition suggest one policy implication: government-forced wage increases are likely to increase effort less than voluntary increases.

[57] Economists should also take the passage's analogy to visual perception to heart: Perfect vision is not viewed by any scientific discipline as an imperative axiom of human nature, and nobody constructs *a priori* "methodological" reasons why visual illusions cannot be important in real-world situations.

[58] Despite emphasis here on how people depart from the notion of *perfect* rationality employed by economists, the heuristics-and-biases paradigm very much takes the perspective that people are intelligent and purposive in their decision-making. Simon (1961) and Williamson (1985, p. 30) have argued convincingly that the appropriate model for economists is that "human behavior is *intendly* rational but only *limitedly* so." In recent years, many economists, especially theorists developing models of evolution and long-run learning, have begun to relax the assumption of perfect rationality by going to the opposite extreme of making no presumption of rationality in novel situations. This research provides some useful insights into boundary conditions and long-run behavior, but will not get us very far in answering broad classes of real-world questions.

[59] Formally, if x is an attribute (shifty eyes) and A is a class of people, $p(x|A)$ is what percentage of people in class A have attribute x, and $p(A|x)$ is the likelihood that that somebody with attribute x is a member of group A, then Bayes's Law implies $\partial p(A|x)/\partial p(x|A) > 0$.

Tversky and Kahneman (1974, pp. 1124-1125) illustrate the representativeness heuristic using the following example:

> Subjects were shown brief personality descriptions of several individuals, allegedly sampled at random from a group of 100 professionals—engineers and lawyers. The subjects were asked to assess, for each description, the probability that it belonged to an engineer rather than to a lawyer. In one experimental condition, subjects were told that the group from which the descriptions had been drawn consisted of 70 engineers and 30 lawyers. In another condition, subjects were told that the group consisted of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. Specifically, it can be shown by applying Bayes' rule that the ratio of these odds should be $(.7/.3)^2$, or 5.44, for each description. In a sharp violation of Bayes' rule, the subjects in the two conditions produced essentially the same probability judgments. Apparently, subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer by the degree to which this description was representative of the two stereotypes, with little or no regard for the prior probabilities of the categories.

The subjects' subjective probabilities that a person was an engineer overwhelmingly reflected their beliefs whether the person's description was typical of an engineer rather than a lawyer. The population proportions had a statistically significant effect, but this effect was minuscule in comparison to the magnitude demanded by Bayes' law. The problem is not *per se* that subjects do not register the population proportions. Kahneman and Tversky (1973) show, for instance, that when given no description at all of an individual chosen randomly from the population, subjects chose the obvious probabilities (70-30 or 30-70) corresponding to the population base rates. Rather, subjects apparently felt base rates were not nearly as important as representativeness for assessing probabilities.

A more dramatic illustration of how little weight people put on base-rate probabilities is the effects of irrelevant information. Kahneman and Tversky (1973) report that when provided with a description of a man that is seemingly uninformative as to whether he is an engineer or a lawyer, people guessed that the probability that the man was an engineer was approximately 50% *irrespective* of whether the population was composed of 70% or 30% engineers. By assessing the odds at 50%, subjects clearly found the description uninformative about whether the person was a lawyer or an engineer; but once they got the description, they ignored the population base rates. Subsequent experiments showed that base rates are typically *not* ignored, and that many factors can get people to take base rates into account. But studies consistently indicate that base rates are generally under-used when non-diagnostic information is provided.

An even more striking illustration of the representativeness heuristic is the observed violation of the conjunction rule, a fundamental axiom of probability theory: The probability that somebody belongs to *both* categories A and B is less than or equal to the probability that she belongs to category B alone. Tversky and Kahneman (1982a) and Tversky and Kahneman (1983) demonstrate what they call the *conjunction effect*: When a description is representative of a person in category A but not of a person in category B, people often judge it more

likely that the description matches somebody who falls into both categories A and B than into category B alone. Tversky and Kahneman (1985, p. 297) illustrate this effect by providing subjects with the following description:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Subjects were then asked to rate the relative likelihood that eight different statements about Linda were true. Two statements on the list were "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement." Over 85% of subjects judged it more likely that Linda was both a bank teller and a feminist than that she was a bank teller. The description of Linda made her seem like a feminist, so that being a bank teller and a feminist seemed a more natural description, and thus more "representative" of Linda, than simply being a bank teller. Tversky and Kahneman (1982a) also elicit the conjunction effect in the prediction of real future events. For instance, subjects were asked to predict what would happen if Bjorn Borg, a tennis star then at his height, reached the finals of Wimbledon. Because good performance was more typical of Borg than bad performance, subjects thought it more likely that Borg would lose the first set but win the match than that he would lose the first set.

Further research sheds light on factors that mediate the relative weightings of base rates and representativeness. Bar-Hillel (1982) and Tversky and Kahneman (1982b, pp. 156-157) report that people become more attentive to base rates when they are perceived as "causal." Consider the following problem posed to subjects:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
> (a) 85% of the cabs in the city are Green and 15% are Blue.
> (b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
> What is the probability that the cab involved in the accident was Blue rather than Green?

Because the vast majority of cabs are green, despite the eyewitness report the correct posterior probability that the cab was blue is about 40%. By contrast, the median answer given by subjects was 80%, indicating that the base rate was virtually ignored in favor of the eyewitness's description. The experiment was then repeated with part (a) of the evidence replaced with the following, keeping everything else precisely the same:

> (a') Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

In this case, the median predicted probability of the cab being blue was 60%. The subjects were still not using the base rates enough, but they were using them substantially more than before. The degree to which base

rates were used to asses probabilities reflects the extent to which base rates seem causally related to the issues at hand.[60]

The Law of Small Numbers

A phenomenon related to the under-use of base rates is "the law of small numbers" (Tversky and Kahneman (1971)): People exaggerate how often to which a small group will closely resemble the parent population or underlying probability distribution that generates the group.[61] We expect even small classes of students to contain very close to the typical distribution of smart ones and personable ones. Likewise, we underestimate how often a good financial analyst will be wrong a few times in a row, and how often a clueless analyst will be right a few times in a row.

By the law of small numbers, people expect close to the same probability distribution of types in small groups as they do in large groups. For example, it is much less likely that at least 80% of 20 coin flips will come up heads than that at least 80% of 5 coin flips will come up heads (about 1% and 19%, respectively); however, people tend to view these as comparably likely. For example, Kahneman and Tversky (1982a, p.44) asked undergraduates the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.
>
> For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

Twenty-two percent of the subjects said that they thought that it was more likely that the larger hospital recorded more such days, and 56% said that they thought the number of days would be about the same. Only 22% of subjects answered correctly that the smaller hospital would report more such days. This is the same fraction as guessed exactly wrong. Apparently, the subjects simply did not see the relevance of the number of child births per day.

---

[60] It is tempting to view the representativeness heuristic as an alternative formula to Bayes' Law, where conditional probabilities are overweighted and base rates are underweighted. Such a formula would certainly facilitate its incorporation into economic models. Yet results on the conjunction fallacy, the role of irrelevant information, and causality all indicate that the representativeness heuristic cannot fully be conceived of in such formulaic terms. Any such formula must, for instance, necessarily quantify "no" information the same as "irrelevant" information, which dramatically violates the evidence presented above. For a sensible formalization as part of an empirical test (and confirmation) of the representiveness heuristic in a controlled economic experiment with real stakes, see Grether (1980, 1992).

[61] For research related to the law of small numbers, see Kahneman and Tversky (1972, 1982e), Tversky and Kahneman (1971, 1982a, 1982b), Gilovich, Vallone, and Tversky (1985), Tversky and Gilovich (1989a, 1989b), and Camerer (1989b).

While people believe in the law of small numbers, they apparently don't believe in the law of large numbers: While we exaggerate the resemblance of small samples to the overall population, we *underestimate* the resemblance that large samples will have to the overall population. Kahneman and Tversky (1972), for instance, found that subjects on average thought that there was a more than 1/10 chance that more than 750 of 1000 babies born on a given day would be male. The actual likelihood is (way) less than 1%. To overstate it a bit, people seem to have a universal probability distribution over sample means that is insensitive to the sample size. As Kahneman and Tversky (1972, p. 45) put it:

> The notion that sampling variance decreases in proportion to sample size is apparently not part of man's repertoire of intuitions. Indeed, misconceptions of the role of sample size occur frequently in everyday life. On the one hand, people are often willing to take seriously a result stated in percentages, with no concern for the number of observations, which may be ridiculously small. On the other hand, people often remain skeptical in the face of solid evidence from a large sample, as in the case of the well-known politician who complained bitterly that the cost of living index is not based on the whole population, but only on a large sample, and added, "Worse yet—a *random* sample."

The law of small numbers implies that people exaggerate the likelihood that a short sequence of flips of a fair coin will yield roughly the same number of heads as tails. What is commonly known as "the gambler's fallacy" is a manifestation of this bias: If a fair coin has not (say) come up tails for a while, then on the next flip it is "due" for a tails, because a sequence of flips of a fair coin ought to include about as many tails as heads. When the underlying probability distribution generating observed sequences is uncertain, the fallacy leads people to over-infer the probability distribution from short sequences. Because we exaggerate how likely it is that a bad financial analyst making three predictions will be wrong at least once, we exaggerate the likelihood that an analyst is good if she is right three times in a row.

This tendency to over-infer from short sequences, in turn, leads to misperception of *regression to the mean.* Because we read too much into patterns that depart from the norm, we don't expect that further observations will look more normal. As teachers, we exaggerate the extent to which one good or bad performance on a test is a sign of good or bad aptitude, so we don't expect exceptional performances to be followed by unexceptional performances as often as they are. (After a particularly outstanding performance, for instance, we *should* expect the next performance to be more normal.) Misunderstanding regression to the mean gives rise to spurious explanations for observed regression. When a student performs poorly on the midterm but well on the final, teachers infer that the student has worked harder; if the student performs well on a midterm but poorly on the final, teachers infer that the student has slacked off. Tversky and Kahneman (1974) give another example. Flight-training instructors observed that when they praised pilots for smooth landings, performance usually deteriorated on the next landing; but when they criticized pilots for poor landings, performance improved the next time. But *random* performance will lead to "deterioration" following a good landing and "improvement" following a poor landing. These flight instructors developed a wrong theory of incentives based on erroneous statistical reasoning.

Another implication of the law of small numbers is that people expect too few lengthy streaks in sequences of random events. As with regression to the mean, therefore, people tend to generate spurious explanations for long streaks that are determined by chance.[62] For instance, there is widespread belief in the "hot hand" in basketball—that particular basketball players are streak shooters who have "on" nights and "off" nights which cannot be explained by randomness. Gilovich, Vallone, and Tversky (1985) and Tversky and Gilovich (1989a, 1989b) have argued that the almost universally accepted phenomenon of the hot hand is almost surely non-existent in basketball. The "hot hand" myth seems partly explained by the perception that observed streaks are too long to be due to chance.[63]

Belief Perseverance and Confirmatory Bias

Once forming strong hypotheses, people often are less attentive to relevant new information supporting or contradicting their hypotheses. For instance, once you become convinced that one investment strategy is more lucrative than another, you may simply stop paying attention to the feedback you get. This can often lead pople to maintain a hypothesis formed on weak evidence, even if later evidence should lead them to reject earlier beliefs.

A particularly elegant demonstration of "anchoring" is found in Bruner and Potter (1964). About 90 subjects were shown blurred pictures that were gradually brought into sharper focus. Different subjects began viewing the pictures at different points in the focusing process, but the pace of the focusing process and final degree of focus were identical for all subjects. Strikingly, of those subjects who began their viewing at a severe-blur stage, less than a quarter eventually identified the pictures correctly, whereas over half of those who began viewing at a light-blur stage were able to correctly identify the pictures. Bruner and Potter (1964, p. 424) conclude that "Interference may be accounted for partly by the difficulty of rejecting incorrect hypotheses based on substandard cues." That is, people who use weak evidence to form initial hypotheses have difficulty correctly interpreting subsequent, better information that contradicts those initial hypotheses.[64]

This form of anchoring does not necessarily imply that people *misinterpret* additional evidence, only that they ignore additional evidence. While such anchoring is potentially quite important, psychological evidence reveals a stronger and more provocative phenomenon: People tend to *misread* evidence as *additional* support for initial hypotheses.[65] If a teacher initially believes that one student is smarter than another, she has the propensity to

---

[62] Spurious explanations also arise from the most basic small-numbers bias. For instance, sports analysts often interpret losses by heavily favored teams as evidence that the favored team was overconfident, took the game for granted, or was looking ahead to their next tough match-up, etc. A better interpretation is that good teams sometimes lose.

[63] For a related example, see Camerer (1989b).

[64] A similar experiment (Wyatt and Campbell (1951)) was cited by Perkins (1981) as one interpretation of the perspective that "fresh" thinkers may be better at seeing solutions to problems than people who have meditated at length on the problems, because the fresh thinkers are not overwhelmed by the "interference" of old hypotheses.

[65] For a formal model of confirmatory bias, see Rabin and Schrag (1996).

confirm that hypothesis when interpreting later performance.[66] Lord, Ross, and Lepper (1979, p. 2099) posited some of the underlying cognitive mechanisms involved in such propensities:

> ...there is considerable evidence that people tend to interpret subsequent evidence so as to maintain their initial beliefs. The biased assimilation processes underlying this effect may include a propensity to remember the strengths of confirming evidence but the weaknesses of disconfirming evidence, to judge confirming evidence as relevant and reliable but disconfirming evidence as irrelevant and unreliable, and to accept confirming evidence at face value while scrutinizing disconfirming evidence hypercritically. With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging "alternative interpretations." Indeed, they may even come to regard the ambiguities and conceptual flaws in the data *opposing* their hypotheses as somehow suggestive of the fundamental *correctness* of those hypotheses. Thus, completely inconsistent or even *random* data – when "processed" in a suitably biased fashion – can maintain or even reinforce one's preconceptions.

The most striking evidence for the confirmatory bias is a series of experiments demonstrating how providing the *same* ambiguous information to people who differ in their initial beliefs on some topic can move their beliefs *further apart.* To illustrate such polarization, Lord, Ross, and Lepper (1979) asked 151 undergraduates to complete a questionnaire that included three questions on capital punishment. Later, 48 of these students were recruited to participate in another experiment. Twenty-four of them were selected because their answers to the earlier questionnaire indicated that they were "'proponents' who favored capital punishment, believed it to have a deterrent effect, and thought most of the relevant research supported their own beliefs. Twenty-four were opponents who opposed capital punishment, doubted its deterrent effect and thought that the relevant research supported *their* views." These subjects were then asked to judge the merits of randomly selected studies on the deterrent efficacy of the death penalty, and to state whether a given study (along with criticisms of that study) provided evidence for or against the deterrence hypothesis. Subjects were then asked to rate, on 16 point scales ranging from -8 to +8, how the studies they had read moved their attitudes towards the death penalty, and how they had changed their beliefs regarding its deterrent efficacy. Lord, Ross, and Lepper (1979, pp. 2102-4) summarize the basic results (all of which hold with confidence $p < .01$) as follows:

> The relevant data provide strong support for the polarization hypothesis. Asked for their final attitudes relative to the experiment's start, proponents reported that they were *more* in favor of capital punishment, whereas opponents reported that they were *less* in favor of capital

[66] A related arena where the confirmation bias has been studied widely is in counselor judgments: Counselors in clinical settings tend to confirm original suppositions in their eventual judgments. If you are told ahead of time that an interviewee is combative, then both your conduct and your interpretation of his conduct during an interview may reinforce that supposition, even if he is in fact no more combative than the average person. See, e.g., Haverkamp (1993). There has also been extensive research on confirmatory bias in the interviewing process more generally; see, e.g., Dougherty, Turban, and Callender (1994) and Macan and Dipboye (1994). Research applying variants of confirmatory bias to other domains includes Arkes (1989) and Borum, Otto, and Golding (1993) to the law; Baumann, Deber, and Thompson (1991) to medicine; and Souter (1993) discusses the implications of overconfidence to business insurance.

> punishment ... Similar results characterized subjects' beliefs about deterrent efficacy. Proponents reported greater belief in the deterrent effect of capital punishment, whereas opponents reported less belief in this deterrent effect.[67]

Three types of evidence flows seem to lead to confirmatory bias. First, the ambiguity of evidence is widely recognized to be an important mediating factor in both confirmatory bias and overconfidence (see, e.g., Keren (1987) and Griffin and Tversky (1992)). Lord, Ross, and Lepper's (1979) study made this clear. Keren (1988) notes the lack of confirmatory bias in visual perceptions, and concludes that confirmatory tendency depends on some degree of abstraction and "discrimination" (i.e., the need for interpretation) not present in simple visual tasks. A primary mechanism of stereotype-maintenance is our tendency to interpret ambiguous behavior according to previous stereotype.[68] Similarly, a teacher can often interpret a question or answer by a student as either creative or just plain stupid; he will often interpret according to her previous hypothesis about the student's aptitude.

A second type of evidence flow that generates confirmatory bias is when people must interpret statistical evidence extended over time by assessing correlations among different phenomena. Nisbett and Ross (1980) argue that the inability to accurately identify such correlations (e.g., between hyperactivity and sugar intake, or between performance on exams and the time of day the exams are held) is one of the most robust shortcomings in human reasoning.[69] People often imagine correlations between events when no such correlation exists.[70] Jennings, Amadibile, and Ross (1982) argue that illusory correlation can play an important role in the confirmation of false

---

[67] Plous (1991) replicates the Lord, Ross, Lepper results in the context of judgments about the safety of nuclear technology. For other papers following on Lord, Ross, and Lepper (1979), see Fleming and Arrowood (1979), Jennings, Lepper, and Ross (1981), Hubbard (1984), Lepper, Ross, and Lau (1986); see also Miller, McHoskey, Bane, and Dowd (1993) for more mixed evidence regarding the Lord, Ross, and Lepper experiment. In the passage above, Lord, Ross, and Lepper posit that even professional scientists are susceptible to such same-evidence polarization. Indeed, many economists and other academics have probably observed how differing schools of thought interpret ambiguous evidence differently. A colleague saw the same model—calibrating the elasticity of demand facing a Cournot oligopolist as a function of the number of firms in an industry—described at the University of Chicago and at M.I.T. A Chicago economist derived the formula and said, "Look at how few firms you need to get close to infinite elasticities and perfect competition." An M.I.T. economist derived the same formula and said, "Look at how large n has to be before you get anywhere close to an infinite elasticity and perfect competition." These different schools each interpreted the same *mathematical formula* as evidence reinforcing their respective views. For related analysis in the scientific domain, see also Mahoney (1977).

[68] A vast literature explores the mechanisms by which people retain ethnic, gender and other group stereotypes. See, e.g., Hamilton and Rose (1980), Bodenhausen and Wyer (1985), Bodenhausen and Lichtenstein (1987), Stangor (1988), Stangor and Ruble (1989), and Hamilton, Sherman, and Ruvolo (1990).

[69] As Jennings, Amabile, and Ross (1982, p. 212) put it, "even the staunchest defenders of the layperson's capacities as an intuitive scientist ... have had little that was flattering to say about the layperson's handling of bivariate observation."

[70] Chapman and Chapman (1967, 1969, 1971) demonstrate that clinicians and laypeople often perceive entirely illusory correlation among (for instance) pictures and the personality traits of the people who drew the pictures. Stangor (1988) and Hamilton and Rose (1980) also discuss the role of illusory correlation in the context of confirmatory-like phenomena.

hypotheses, finding that people underestimate correlation when they have no theory of the correlation, but exaggerate correlation and see it where it is not when they have a preconceived theory of it.[71]

The final mode of information processing inducing confirmatory bias is selective collection or scrutiny of evidence. One form of "scrutiny-based" confirmatory bias is what I shall call *hypothesis-based filtering*.[72] While it is sensible to interpret ambiguous data according to current hypotheses, people tend to use the consequent "filtered" evidence inappropriately as further evidence for these hypotheses. If a student gives an unclear answer to an exam question, it is perfectly reasonable for a teacher to be influenced in his evaluation of the answer by his prior perceptions of that student's mastery of the material. However, after assigning differential grades to students according differential interpretation of comparable answers, it is a mistake to *then* use differential grades on the exam as *further* evidence of the differences in the students' abilities.[73] This sort of error is especially likely when the complexity and ambiguity of evidence *requires* the use of prior theories to interpret data and to decide what data to be attentive to.

Other Biases

I briefly outline a few more biases that might interest economists.[74] The first is *adjustment and anchoring*. Slovic and Lichtenstein (1971) demonstrate that, in forming numerical estimates of uncertain quantities, adjustments in assessments away from (possibly arbitrary) initial values are typically insufficient. Tversky and Kahneman (1974) provide the following example:

---

[71] Similarly, Redelmeier and Tversky (1996) argue illusory correlation may help explain the persistent belief that arthritis pain is related to the weather.

[72] Another mechanism can be defined as "positive test strategy": People tend to ask questions (of others, of themselves, or of data) that are likely to be true if their hypothesis is true—without due regard to the fact that they are likely to be true even if the hypothesis is *false*. See Einhorn and Hogarth (1978), Klayman and Ha (1987), Beattie and Baron (1988), Devine, Hirt, and Gehrke (1990), Hodgins and Zuckerman (1993), Friedrich (1993), and Zuckerman, Knee, Hodgins, and Miyake (1995). We are using this term a bit differently than we suspect psychologists would use it. So far as we know, the term was coined by Klayman and Ha (1987) to point out that much of what was put under the rubric of confirmatory bias could indeed be a rational form of hypothesis testing. Fischhoff and Beyth-Marom (1983, pp. 255-6) and Friedrich (1993) also point out that if people are fully aware that asking "soft" questions teaches them little about the truth of hypotheses, then no bias has occurred. While we feel research on the positive test strategy needs more careful calibration versus Bayesian updating, we believe that the evidence suggests that people do not fully appreciate how little they have learned about the validity of their hypotheses when asking soft questions. (Mehle, Gettys, Manning, Baca, and Fisher (1981), for instance, show that people with specified hypotheses for observed data tend to over-use such hypotheses to explain the data because they do not have "available" the many unspecified hypothesis that could also explain the data.)

[73] Lord, Ross, and Lepper (pp. 2106-7) note a similar distinction in reflecting on the bias in their experiment discussed above. They note that it is proper for people to differentially assess probativeness of different studies according to their current beliefs about the merits of the death penalty. The "sin" is in using their hypothesis-based interpretations of the strength of different studies as further support for their beliefs.

[74] For a more thorough introduction to this literature, see Kahneman, Slovic, and Tversky (1982), or, for an outstanding review of this material, and of individual decision-making more generally, see Camerer (1995).

> In a demonstration of the anchoring effect, subjects were asked to estimate various quantities, stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity, a number between 0 and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given number. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on estimates. For example, the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points. Payoffs for accuracy did not reduce the anchoring effect.

While this example is somewhat artificial, Tversky and Kahneman point out that anchoring can occur as a natural part of the assessment process itself.[75] If we ask an individual to construct a probability distribution for the level of the Dow Jones, her likely beginning point would be to estimate a median level. This value would likely then serve as an anchor for her further probability assessments. By contrast, if she were asked by somebody to construct the probability assessments by stating the likelihood of the Dow Jones exceeding a pre-specified value, she would likely anchor on this value. The two procedures, therefore, are likely to lead to different predictions, with the first procedure yielding a probability distribution more concentrated around the median than the second.

One of the most widely studied biases in the judgment literature is the *hindsight bias.*[76] Fischhoff (1975, p. 288) first proposed this bias by observing that "(a) Reporting an outcome's occurrence increases its perceived probability of occurrence; and (b) people who have received outcome knowledge are largely unaware of its having changed their perceptions [along the lines of (a)]." Combining these, the literature on the hindsight bias shows that people exaggerate the degree to which their beliefs before an informative event would be similar to their current beliefs. We tend to think we "knew it would happen all along." After a politician wins election, people label it as inevitable—and believe that they *always* thought it was inevitable.

One example of Fischhoff's (1975) original demonstration of this effect was to give subjects a historical passage regarding British intrusion into India and military interaction with the Gurkas of Nepal. Without being told the outcome of this interaction, some subjects were asked to predict the likelihood of each of four possible outcomes: (1) British victory, 2) Gurka victory, and military stalemates 3) with and 4) without a peace settlement). Four other sets of subjects were each *told* a different one of the four outcomes was the true one (the *real* true outcome is that the two sides fought to a stalemate without reaching a peace settlement). For all four outcomes, subjects on average guessed that they would have estimated the probability of the given event as about 15% more

---

[75] Moreover, many researchers have invoked adjustment and anchoring as a partial explanation for other observed biases; e.g., Fischhoff (1975) and Lopes (1982) on the hindsight bias discussed below, and Block and Harper (1991) on the confirmation bias, discussed in the previous section.

[76] For two excellent recent reviews of the hindsight bias, see Hawkins and Hastie (1990) and Christensen-Szalanski and Willham (1991). Christensen-Szalanski and Willham conduct a meta-analysis of the literature—aggregating the findings of 122 different studies, gathered through an unbiased procedure to test for whether the overall effect from the many experiments that exist confirm the bias. They conclude that the bias is very real. (They also argue that the effects are "small.")

likely than those subjects not told an outcome actually estimated it. The bias is that perceived beliefs clearly reflect to some degree the additional information they have; people don't sufficiently "subtract" information they currently have about an outcome in imagining what they would have thought without that information.[77]

A pervasive fact about human judgment is that people disproportionately weight *salient, memorable,* or *vivid* evidence even where they have better sources of information.[78] For instance, our assessment of the danger of a given city is likely to be too influenced by whether we personally know somebody who has been assaulted, even if we are familiar with much more relevant general statistics. Likewise, dramatic stories by people we know about difficulties with a brand of car are likely to be overly influential even if we are familiar, via *Consumer Reports*, with general statistics of the reliability of different brands. In both these cases and in many others, the more salient information should have virtually no influence on our beliefs in the face of much more pertinent statistical information. Tversky and Kahneman (1973a) discuss an example where the memorableness of events may lead to distorted perceptions. They consider clinicians' assessments of the likelihood that severe depression will lead to suicide. They note that incidents in which patients commit suicide are much more likely to be remembered by clinicians than are instances where patients do not commit suicide. In conjunction with the representativeness heuristic, this is likely to lead to an exaggerated assessment of the probability that depressed patients will commit suicide. Relatedly, Tversky and Kahneman (1973a, p. 178) note that dramatic anticipated events may also be more salient:

> Continued preoccupation with an outcome may increase its availability, and hence its perceived likelihood. People are preoccupied with highly desirable outcomes, such as winning the sweepstakes, or with highly undesirable outcomes, such as an airplane crash. Consequently, availability provides a mechanism by which occurrences of extreme utility (or disutility) may appear more likely than they actually are.

Finally, there is a mass of psychological research that finds people are prone towards overconfidence in their judgments. The vast majority of researchers argue that such overconfidence is pervasive, and most of the research concerns possible explanations (of which confirmatory bias discussed above is one).[79]

---

[77] The definition of hindsight bias regards people's perceptions of how *they themselves* would have answered a particular question absent information they now have. As economists, we are likely to care mostly about a person's beliefs about other people, not about herself. In general, it is hard to control experimentally for the fact that people have different information, and hard to isolate the hindsight bias when asking subjects what others would have believed absent certain information. Subjects may believe that others have different beliefs for a variety of reasons (e.g., you could believe that other people are not as smart as you). I feel, however, that the evidence suggests that we have a tendency to think that other people "should have known" as well; see, for instance, Hawkins and Hastie (1990, p.319).

[78] In Tversky and Kahneman's (1973) formulation: "[A] person is said to employ the *availability heuristic* whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind." For more general reviews of the role of salience and vividness, see Nisbett and Ross (1980, Chapters 3,4,6) and Fiske and Taylor (1991, Chapters 5,7).

[79] See, e.g., Oskamp (1982), Mahajan (1992), and Paese and Kinnaly (1993). An early paper arguing this is Fischhoff, Slovic, and Lichtenstein (1977), who also tested the robustness of overconfidence with monetary stakes rather than reported judgments. No decrease in overconfidence was found relative to the no-money-stakes

Do Learning and Expertise Eliminate Biases?

The conjecture that experience helps overcome biases often leads economists to doubt the relevance of laboratory evidence from inexperienced subjects. It is commonly argued that if important economic activity is performed by specialists and experts, or consists of tasks done repeatedly by the same individuals, the assumption of full rationality fares much better than some of the psychological evidence indicates.

Do experience, expertise, and learning substantially eliminate biases? Such conjectures are certainly reasonable, and such factors probably on average moderate biases. But the conjectures are a lot less correct than economists imagine. Kahneman and Tversky (1982a) and Tversky and Kahneman (1982a), for instance, present experiments with subjects who vary in their level of statistical sophistication, to test whether general knowledge of statistics reduces or eliminates observed biases. The results are surprisingly negative. More generally, the research leads to mixed conclusions about when and how learning takes place, but does not overall support the strong versions of the experts-get-things-right and in-the-real-world-people-learn hypotheses.[80]

Research also suggests we should use extreme caution in defining the relevant notion of learning, because many people who *do* learn general principles do not apply those principles in particular situations. Kahneman and Tversky (1982a, p. 495) call such errors *errors of applications*: "A failure in a particular problem is called an error of application if there is evidence that people know and accept a rule that they did not apply," and note that "An error of application is most convincingly demonstrated when a person, spontaneously or with minimal prompting, clutches his head and exclaims: 'How could I have missed that?'" Even if people learn the relevant statistical truths of their environment, they may continue to make errors in their judgments and decision-making in every single case.[81]

---

condition. There have, however, been criticisms of the evidence in support of overconfidence. Some of this research reflects a more general criticism of some of the research on judgment—that it identifies errors by people in artificial settings performing tasks chosen by researchers to be especially challenging, and that more realistic ("ecologically valid") environments are needed to evaluate human judgment. For instance, Bjorkman (1994) argues that overconfidence disappears in more realistic environments. See also Pfeifer (1994) for a related critique. Other papers challenging the "overconfidence" results include Tomassini, Solomon, Romney, and Krogstad (1982), Van Lenthe (1993), and Winman and Juslin (1993). See Soll (1996), however, for a counter-critique.

[80] See, for instance, several chapters in Kahneman, Slovic, and Tversky (1982), with titles such as "Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments," "Learning from Experience and Suboptimal Rules in Decision Making," "A Progress Report on the Training of Probability Assessors," and "Debiasing." See especially Einhorn and Hogarth (1978), who explore why people fail to learn from feedback they get over time.

[81] One fears that economists may sloppily interpret such head-clutching as evidence *for* the rationality hypothesis, rather than against it. But evidence that people see their errors when confronted with them does not boost the rationality assumption as economists use it. Our models use the rationality assumption as a *realized* feature of human behavior, not merely a human potential.

Griffin and Tversky (1992, p.431) reach a similar conclusion in the context of the overconfidence bias: Even when people understand the limits in their abilities to predict events accurately, they tend not to apply this general knowledge in calibrating the appropriate confidence in individual cases: [82]

> Thus, the overconfidence observed in average judgments of confidence need not apply to global judgments of expected accuracy. Indeed, Table 4 shows that estimated frequencies were substantially below the actual frequencies of correct prediction. In fact, the latter estimates were below chance for two of the three attributes. Similar results have been observed by other investigators (e.g., Gigerenzer et al., 1991; May, 1986; Sniezek & Switzer, 1989). Evidently, people can maintain a high degree of confidence in the validity of specific answers even when they know that their overall hit rate is not very high. This phenomenon has been called the "illusion of validity" (Kahneman & Tversky, 1973): people often make confident predictions about individual cases on the basis of fallible data (e.g., personal interviews or projective tests) even when they know that these data have low predictive validity (Dawes, Faust, & Meehl, 1989).

"Learning" sometimes tends to *exacerbate* errors. This was demonstrated in the context of confirmation bias discussed above. Relatedly, Griffin and Tversky (1992), address the relationship between expertise and overconfidence. When certain forms of predictability are high and when feedback takes the form of unambiguous statistical evidence, experts tend to have a pretty good sense of how accurate their predictions are. In such cases, experts not only know more, but are more realistic than laypersons about how much they know. But experts are often more susceptible to overconfidence than are laypersons. Griffin and Tversky (1992, p. 430) characterize such situations:[83]

> When predictability is reasonably high, experts are generally better calibrated than lay people. ... When predictability is very low, however, experts may be more prone to overconfidence than novices. If the future state of a mental patient, the Russian economy, or the stock market cannot be predicted from present data, then experts who have rich models of the system in question are more likely to exhibit overconfidence than lay people who have a very limited understanding of these systems. Studies of clinical psychologists (e.g., Oskamp, 1965) and stock market analysts (e.g., Yates, 1990) are consistent with this hypothesis.

---

[82] See also Kahneman and Lovallo (1993); similarly, see Baumann, Deber, and Thompson (1991) for evidence that people who are aware of their own accuracy overall are over-confident on a case-by-case basis. For a related phenomenon that is also related to the framing and context effects discussed in Section 4 below, see Redelmeier and Tversky (1990, 1992). Redelmeier and Tversky (1990, p. 1163) report that their "results indicate that physicians make different decisions when evaluating an individual patient than when considering a group of comparable patients ... From the individual as compared with the aggregate perspective, physicians are more likely to order an additional test, expend time directly assessing a patient, avoid raising some troubling issues, and recommend a therapy with a high probability of success but the chance of an adverse outcome. The discrepancy between the aggregate and individual perspectives demonstrated in these experiments cannot be attributed to differences in either medical information or economic incentives; hence it is difficult to explain on normative grounds."

[83] For other research related to determinants of expert overconfidence, see e.g. Keren (1987) and Braun and Yaniv (1992).

the main carriers of utility become not the absolute levels of consumption, but departures from our (new) reference level.[87]

People do not anticipate the degree of such adaptation, and thus exaggerate changes in utility caused by changes in their lives.[88] This suggests that the "decision-utility" aversion people have to losses is *not* consonant with "experienced utility." This realization, in turn, calls for a reexamination of the first topic of Section 2: Are loss aversion, the endowment effect, and other reference effects rational or irrational? If people experience losses relative to a status quo as quite unpleasant, then loss-averse behavior is rational, because people are correctly anticipating and avoiding unpleasant sensations. But Kahneman (1991, p. 143) argues that loss aversion is often a judgmental heuristic, and Tversky and Kahneman (1991) note that in decisions with significant long-run consequences, people should put less weight than they do on their initial experience of losses.[89]

In principle, of course, the remembered "loss" of an owned mug may carry over time, or in any event be substantial relative to the long-term utility consequences of owning the mug. But for other more extreme examples of loss aversion it is hard to believe that the "transition utility" ranks high relative to long-term utility. For instance, Thaler (1980, pp. 43-4) asked subjects each of the following two survey questions:

> (a) Assume you have been exposed to a disease which if contracted leads to a quick and painless death within a week. The probability you have the disease is .001. What is the maximum you would be willing to pay for a cure?
>
> (b) Suppose volunteers were needed for research on the above disease. All that would be required is that you expose yourself to a .001 chance of contracting the disease. What is the minimum payment you would require to volunteer for this program? (You would not be allowed to purchase the cure.)
>
> *The results.* Many people respond to questions (a) and (b) with answers which differ by an order of magnitude or more! (A typical response is $200 and $10,000.)

---

[87] Brickman, Coates, and Janoff-Bulman (1978) also interviewed twenty-nine paraplegics and quadriplegics about their happiness before their accidents, their current happiness, and their expected future happiness. The data here were much more equivocal; as we might expect, paraplegics felt less well off on average than they felt before, and rated their happiness as lower than did lottery winners or the control group. Moreover, they did *not* currently enjoy mundane activities more than lottery winners or the control group. The interviews, however, took place relatively soon after accidents, and there was evidence that accident victims' new perspectives led them to put more emphasis on and take a more positive view of mundane pleasures—they rated both past pleasure and anticipated future pleasure from such mundane activities slightly higher than either the lottery winners or the control group.

[88] E.g., in the simple model of reference-point adjustment discussed in Section 2, it may be that people systematically underestimate the parameter α.

[89] Indeed, other researchers invoke loss aversion more as an irrational rule of thumb than as a rational utility function. Benartzi and Thaler (1995) argue that the equity-premium puzzle can be explained by investors' aversion to short-term financial losses, even though they will not be spending their investment in the short term. Camerer, Babcock, Loewenstein, Thaler (1995) argue that New York taxi drivers decide when to quit driving for the day by a rule of thumb that says they should make sure to match their usual take for the day.

People charge heavy premiums for losses relative to their status quo, even when it is hard to imagine that the relevant transition utility is significant relative to the long-term utility (here, whether or not you live beyond a week).

An example of people misperceiving the utility consequences of their choices is Herrnstein and Prelec's (1992b) theory of *melioration*. Based on a mass of evidence for people and other animals, they argue that people tend to make current choices according to which choice *directly* yields the highest utility, without taking into account the choice's effect on the utilities from future choices. That is, people often ignore "internalities"—the effects one choice episode has on the utilities of later choice episodes.

Say you eat at one of two places every night—either *Blondie's* or *Fat Slice*. You enjoy Fat Slice more, but because you also enjoy variety, your utility each evening is as follows:

Utility from *Fat Slice* = 7 if you ate at *Blondie's* last night
Utility from *Fat Slice* = 5 if you ate at *Fat Slice* last night
Utility from *Blondie's* = 4 if you ate at *Fat Slice* last night
Utility from *Blondie's* = 3 if you ate at *Blondie's* last night

On any given day, no matter your recent eating pattern, you get higher utility from eating at Fat Slice than at Blondie's. Yet your utility-maximizing consumption program is to alternate between Fat Slice and Blondie's (thus alternating between payoffs of 7 and 4, for an average of 5.5) rather than eating all the time at Fat Slice (thus getting a payoff of 5 each period). Yet, because at each moment we tend to ask,"Which will yield me more pleasure—Fat Slice or Blondie's?", we may eat too often at Fat Slice.

Of course, we *do* often train ourselves, or learn over time, about the overall hedonic consequences of such "internalities." For example, one application of melioration is in predicting the overconsumption of addictive and other negative consumption goods (see Herrnstein and Prelec (1992b)). We are constantly warned by ourselves and others to realize the long-run consequences of smoking, or ingesting illicit drugs or licit cholesterol; likewise, we learn to engage in certain unpleasant activities because they enhance the utility of other activities. But the evidence suggests that we take too little account of the global utility effects of individual choices.[90]

Thus, people don't always correctly predict the utility consequences of the outcomes they achieve. One major source for predicting utility from future experiences is, of course, the recollected utility from comparable past experiences. We might presume that people will accurately predict their utility from familiar experiences. The research discussed above hints that this presumption is not accurate: Obtaining items such as mugs is presumably

---

[90] Herrnstein and Prelec (1992, pp. 257-8) argue that people are *not* necessarily always unaware of the shape of their global utility functions; rather, they may not understand the proper way to maximize given their preferences. The textbook approach of writing down a utility function for an overall consumption bundle, and asking our students to find the optimal bundle, obscures the fact that many real-life consumer choices take the form I have presented here (as marginal satisfaction from given activities). See Herrnstein and Prelec (1992, p. 236) for a nice discussion of how economists glide over the distinction between global and piecemeal maximization. See also Fehr and Zych (1994) for an experimental exploration of this distinction.

a fairly familiar experience; if we misperceive the long-run consequences of giving up minor consumer items, we may not have learned to assess correctly our everyday choices.

Additional research even more dramatically demonstrates systematic differences between people's experienced utility of episodes and their recollections of those episodes. Several recent experiments compare recollected utility to experienced utility for episodes extended over time, and rely on the periodic hedonic reports by subjects of their well-being at each point in time.[91] In evaluating the overall utility from such an extended episode, one must formulate criteria for "adding up" flows of experienced well-being. Arguably, the most reasonable normative evaluation for flows of utility is *temporal integration*—adding up well-being across different periods, giving equal weight to each point in time. A weaker criterion for comparing episodes is *temporal monotonicity*—that adding moments of pain to an otherwise unchanged experience decreases overall well-being, and that adding episodes of pleasure increases overall well-being. Armed with a means of comparing experienced well-being of episodes, we can then ask how people's own retrospective evaluations of episodes compare to their experienced well-being. Kahneman (1994, p. 29) offers the following summary of findings:

> The results of these studies support two empirical generalizations. (1) 'The Peak and End Rule': global evaluations are predicted with high accuracy by a weighted combination of the most extreme affect recorded during the episode and of the affect recorded during the terminal moments of the episode. Here again, as in the context of decision utility, the evaluations of particular moments appears to be used as a proxy for the evaluation of a more extended period of time. (2) Duration Neglect. The retrospective evaluation of overall or total pain (or pleasure) is not independently affected by the duration of the episode. In the colonoscopy study, for example, the duration of the procedure varied from 4 to 69 minutes in a sample of 101 patients. Surprisingly, these variations of duration had no significant effect on retrospective evaluations. The ratings of both patients and attending physicians were dominated by the intensity of pain at its worst, and by the intensity of discomfort during the last few minutes of the procedure. Duration neglect is not immutable, of course ... In general, however, affective peaks and endings are more salient than duration in the cognitive representation of events.

Of course, one must carefully consider the pain and pleasure associated with an experience before and after the actual experience; anticipation and recollection of pain, for instance, are clearly important influences on long-run utility, just as anticipation and recollection of a vacation are very significant in evaluating the overall well-being associated with vacations.[92]

Not knowing your own "experienced utility function" is obviously important for the welfare implications of choice, and the main lesson of this material is that economists ought recognize that people may not correctly predict what makes them happy. But I conclude this subsection with evidence that people misperceive their future *behavior*. Loewenstein and Adler (1995) consider people's predictions of their own future decisions. Their

---

[91] See Fredickson and Kahneman (1993), Kahneman, Fredrickson, Schreiber, and Redelmeier (1993), Kahneman and Varey (1991), Redelmeier and Kahneman (1993), and Varey and Kahneman (1992).

[92] Elster and Loewenstein (1992) discuss how economists typically ignore the strong role that remembered and anticipated experiences can play in experienced well-being. Brickman, Coates, and Janoff-Bulman (1978) also provide evidence for the strong role that remembered experiences play in the mind-set of accident victims.

experiment directly concerned the endowment-effect experiments of Kahneman, Knetsch, and Thaler (1990) discussed in Section 2. In their new version, some subjects were first asked to "imagine that we gave you a mug exactly like the one you can see, and that we gave you the opportunity to keep it or trade it for some money." All subjects were then given a mug, and were asked then to choose a minimal selling price (such that, they were truthfully told, an offer to buy would be implemented if and only if the buying price exceeded the stated price).

Before receiving the mugs, subjects on average predicted their own minimal selling price at $3.73. Once they had the mugs, however, their actual minimal selling price averaged $4.89. That is, subjects systematically underestimated the endowment effect. They behaved significantly differently than they had predicted about themselves *moments* earlier. Such a procedure is likely to underestimate the true degree of misperception, since people don't like to contradict recently expressed predictions of their own behavior. Indeed, subjects who had made no prediction averaged a much higher selling price of $5.62.

Elicitation Effects

They way choices are elicited can influence the choices people make. People often lack a stable preference ordering that has even minimal robustness to different ways of uncovering these preferences.

Research shows the pervasiveness of *framing effects*: two logically equivalent statements of a problem lead decision-makers to choose different options. Framing effects typically involve differing frames whose logical equivalence is neither totally transparent nor terribly obscure, and the ways in which decisions are influenced is typically predictable by the features emphasized by the frames. The most predictable influences of framing on choice relate to loss aversion and diminishing sensitivity, as outlined in Section 2 above. Because losses resonate with people more than gains, a frame that highlights the losses associated with a choice makes that choice less attractive. Similarly, a frame that exploits diminishing sensitivity by making losses appear small relative to the scales involved makes that choice more attractive. Tversky and Kahneman (1986, p. S254-5) give the following example of framing effects, taken from a study of medical decisions by McNeil, Pauker, Sox, and Tversky (1982):

> Respondents were given statistical information about the outcomes of two treatments of lung cancer. The same statistics were presented to some respondents in terms of mortality rates and to others in terms of survival rates. The respondents then indicated their preferred treatment. The information was presented [exactly] as follows.
>
> Problem 1 (Survival frame)
>
> Surgery: Of 100 people having surgery 90 live through the post-operative period, 68 are alive at the end of the first year and 34 are alive at the end of five years.
>
> Radiation Therapy: Of 100 people having radiation therapy all live through the treatment, 77 are alive at the end of one year and 22 are alive at the end of five years.
>
> Problem 1 (Mortality frame)

> Surgery: Of 100 people having surgery 10 die during surgery or the post-operative period, 32 die by the end of the first year and 66 die by the end of five years.
>
> Radiation Therapy: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year and 78 die by the end of five years.
>
> The inconsequential difference in formulation produced a marked effect. The overall percentage of respondents who favored radiation therapy rose from 18% in the survival frame (N = 247) to 44% in the mortality frame (N = 336). The advantage of radiation therapy over surgery evidently looms larger when stated as a reduction of the risk of immediate death from 10% to 0% rather than as an increase from 90% to 100% in the rate of survival. The framing effect was not smaller for experienced physicians or for statistically sophisticated business students than for a group of clinic patients.

This question is hypothetical, and real stakes might change the level of care which subjects give to the problem. (Controlled experiments of the issues above would be problematic.) But the fact that experienced physicians make the same mistakes suggests that similar patterns might play out in the real world. Similar framing effects were found in choices over lotteries with small monetary stakes.[93] More generally, it would be wrong to think that framing effects only matter when stakes are small, when the framing of the choices obscures the appropriate choice in an exceedingly subtle way, or when the decision context is hypothetical or non-economic. Tversky and Kahneman (1986, S261-2) cite some important real-world examples of framing effects:

> Thaler (1980) drew attention to the effect of labeling a difference between two prices as a surcharge or a discount. It is easier to forgo a discount than to accept a surcharge because the same price difference is valued as a gain in the former case and as a loss in the latter. Indeed, the credit card lobby is said to insist that any price difference between cash and card purchases should be labeled a cash discount rather than a credit surcharge. A similar idea could be invoked to explain why the price response to slack demand often takes the form of discounts or special concessions (Stigler and Kindahl 1970). Customers may be expected to show less resistance to the eventual cancellation of such temporary arrangements than to outright price increases. Judgments of fairness exhibit the same pattern (Kahneman, Knetsch, and Thaler (1986a)).
>
> Schelling (1981) has described a striking framing effect in a context of tax policy. He points out that the tax table can be constructed by using as a default case either the childless family (as is in fact done) or, say, the modal two-child family. The tax difference between a childless family and a two-child family is naturally framed as an exemption (for the two-child family) in the first frame and as a tax premium (on the childless family) in the second frame. This seemingly innocuous difference has a large effect on judgments of the desired relation between income, family size, and tax. Schelling reported that his students rejected the idea of granting the rich a larger exemption than the poor in the first frame but favored a larger tax premium on the childless rich than on the childless poor in the second frame. Because the exemption and the premium are alternative labels for the same tax differences in the two cases, the judgments violate invariance. Framing the consequences of public policy in positive or in negative terms can greatly alter its appeal.

---

[93] See, e.g., Tversky and Kahneman (1986, S264-5).

Money illusion also provides a good example of the importance of framing effects for economics. Kahneman, Knetsch, and Thaler (1986a), for instance, provide survey evidence that people are very attentive to nominal rather than real changes in wages and prices in assessing the fairness of firm behavior. For instance, a nominal wage increase of 5% in a period of 12% inflation offends people's sense of fairness less than a 7% decrease in a time of no inflation. More generally, people react more to decreases in real wages when they are also nominal decreases, and react negatively to nominal price increases even if they represent no increase in real prices.[94]

Framing effects can often be viewed as heuristic errors—people are boundedly rational, and the presentation of a choice may draw our attention to different aspects of a problem, leading us to make mistakes in pursuing our true, underlying preferences. As such, framing effects could have been included as a topic in Section 3. But sometimes framing effects cut more deeply to economists' model of choice—more than confusing people in pursuit of stable underlying preferences, the "frames" may in fact partially *determine* a person's preferences. The related phenomena of preference reversals and context effects raise even more fundamental doubts that choices reflect stable, well-defined preferences. Tversky and Thaler (1990, p. 209) quote two economists who experimentally confronted and failed to refute evidence on preference reversals:

> Taken at face value the data [showing preference reversals] are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency is deeper than the mere lack of transitivity or even stochastic transitivity. It suggests that no optimization principles of any sort lie behind the simplest of human choices and that the uniformities in human choice behavior which lie behind market behavior may result from principles which are of a completely different sort from those generally accepted (Grether and Plott, p. 623).

Tversky and Thaler (1990, p. 210) conclude their paper with a similarly strong statement:

> [P]eople do not possess a set of pre-defined preferences for every contingency. Rather, preferences are constructed in the process of making a choice or judgment. Second, the context and procedures involved in making choices or judgments influence the preferences that are implied by the elicited responses. In practical terms, this implies that behavior is likely to vary across situations that economists consider identical. For example, alternative auction mechanisms which are equivalent in theory might produce different outcomes if the auction procedures themselves influence bidding behavior.[95]

The specific area of research that prompted such conclusions is the study of "preference reversals." As Tversky and Thaler (1990, pp. 202-3) summarize it:

---

[94] See Shafir, Tversky, and Diamond (1996) for a further exploration of the psychology of money illusion.

[95] The hypersensitivity of "preferences" to the method of eliciting those preferences is a key issue in the emerging debate on "contingent valuation"; see, e.g., Kahneman and Knetsch (1992), Hausman (1993), Kahneman, Ritov, Jacowitz, and Grant (1993), Diamond and Hausman (1994), Kahneman and Ritov (1994), and McFadden (1994).

> For almost two decades, economists and psychologists have been intrigued by [an] inconsistency involving risky prospects. Subjects are first asked to choose between two gambles with nearly the same expected values. One gamble, called the H bet (for high chance of winning) has a high chance of winning a relatively small prize (say, 8/9 chance of win $4), while the other gamble, the L bet, offers a lower chance to win a larger prize (say, a 1/9 chance to win $40). Most subjects choose the H bet. Subjects are then asked to price each of the gambles. Specifically, they are asked to state the lowest price at which they would be willing to sell each gamble if they owned it. Surprisingly, most subjects put a higher price on the L bet. (In a recent study that used this particular pair of bets, for example, 71 percent of the subjects chose the H bet, while 67 percent priced L above H.) This pattern is called a preference reversal. Sarah Lichtenstein and Paul Slovic (1971, 1973) first demonstrated such reversals in a series of studies, one of which was conducted for real money with gamblers on the floor of the Four Queens Casino in Las Vegas...
>
> The preference reversal phenomenon raises an issue rarely discussed in economics: How is the notion of preference to be operationalized? We say that option A is preferred to option B if A is selected when B is available or if A has a higher reservation price than B. The standard analysis of choice assumes that these procedures give rise to the same ordering. This requirement—called procedural invariance—seldom appears as an explicit axiom, but it is needed to ensure that the preference relation is well defined.

Other research also raises doubts regarding the standard assumption that there exist stable preferences independent of the menu of choices. Simonson and Tversky (1992, p. 281) specifically take issue with the idea that consumer preferences among goods will be independent of their menu of choices. They provide examples where, because of such *context effects*, the addition of a new option to a set of choices may actually increase the proportion of consumers who choose one of the existing options. For example, the proportion of consumers who chose a particular model of microwave oven increased when a second, more expensive model was added to their choice set. Adding the more expensive model accentuated the inexpensiveness of the other model, and thus made it more attractive.[96] The more general principle that an unattractive option can enhance the attractiveness of other options is folk wisdom among salespersons. Simonson and Tversky (1992, p. 287) ran an experiment that illustrates this idea:

> [Subjects] were informed that some of them, selected randomly, would receive $6. They were further told that the winners would have the option of trading the $6 for a pen. Subjects were asked to examine the available pens and indicate whether they would like to trade the $6 for a pen. Later, 10% of the participants received either $6 or the pen they had chosen.
>
> In one version of the questionnaire, subjects were presented an elegant Cross pen. In the other version, subjects were given an additional option—a lesser known brand name that was selected specifically for its unattractiveness. ... Though [only 2% of the] subjects selected the less attractive pen, its inclusion in the offered set increased the percentage of respondents who preferred the more attractive Cross pen from 36% to 46% ($t = 1.5$, $p < .10$). This observation suggests that the tendency to pay cash for a good can be increased by the introduction of an inferior alternative.

---

[96] Subjects were first asked to look at a catalogue containing the prices and descriptions of all the relevant choices from which their eventual choice sets were drawn, so the results seem unlikely to be due to the information revealed by the descriptions of the offered sets.

Simonson and Tversky (1992, p. 292) note that people are often unaware that the menu of choices influences their decisions. But at other times, decision-makers explicitly "rationalize" their choices with references to their choice sets. For instance, people may state explicitly that a given choice is a compromise between two other choices. Such findings suggest an alternative to the utility-maximization framework that may help explain framing effects, preference reversals, and context effects. I turn to this interpretation in the next subsection.

Choices and Reasons

Rather than maximizing utility, research suggests that people may make choices in part by asking themselves whether they have a "reason" to choose one option over another.[97] Shafir, Simonson, and Tversky (1993, p. 13) describe this research program as follows:

> ... a focus on reasons seems closer to the way we normally think and talk about choices. When facing a difficult choice (e.g., between schools, or jobs) we try to come up with reasons for and against each option—we do not normally attempt to estimate their overall values. Second, thinking of choice as guided by reasons provides a natural way to understand the conflict that characterizes the making of decisions. From the perspective of reason-based choice, conflict arises when the decision-maker has good reasons for and against each option, or conflicting reasons for competing options. Unlike numerical values, which are easy to compare, conflicting reasons may be hard to reconcile. An analysis based on reasons can also accommodate framing effects (Tversky & Kahneman, 1986) and elicitation effects (Tversky, Sattath, & Slovic, 1988), which show that preferences are sensitive to the ways in which options are described (e.g., in terms of gains or losses), and to the methods through which preferences are elicited (e.g., pricing versus choice). These findings, which are puzzling from the perspective of value maximization, are easier to interpret if we assume that different frames and elicitation procedures highlight different aspects of the options and thus bring forth different reasons to guide decision.

As noted in the above passage, "reason-based choice" can help interpret some of the research in the previous subsection. One theme of Simonson (1989), Tversky and Simonson (1993), and Simonson and Tversky (1992), for instance, is that an option A is more likely to be chosen when there exists an option B that is clearly dominated by option A, apparently because the existence of the option B provides a "reason" to choose option A.

Other recent research establishes even more dramatically the role of reasons in decision making. For instance, Tversky and Shafir (1992b) demonstrate that when people are searching, they do not merely search to find a high-value option—as assumed in conventional search theory—but also seem to search for a "reason" to end their search and make one particular choice. Tversky and Shafir (1992b) conducted experiments where subjects were asked either to choose among existing options, or to engage in costly further search. In addition to a parallel

---

[97] Papers in this vein include Redelmeier and Shafir (1993), Shafir (1993), Shafir and Tversky (1992), Simonson (1989), Simonson, Carmon, and O'Curry (1994), Simonson, Nowlis, and Simonson (1993), Simonson and Tversky (1992), Slovic (1975), Tversky and Shafir (1992), and Tversky and Simonson (1993).

experiment where subjects were asked to choose among gambles for real monetary stakes, a hypothetical apartment-search problem was posed. Shafir, Simonson, and Tversky (1993, pp. 19-20) summarize it as follows:

> Subjects were presented choices between hypothetical student apartments. Some subjects received the following problem:
>
> Conflict:
>
> Imagine that you face a choice between two apartments with the following characteristics:
>
> x) $290 a month, 25 minutes from campus
> y) $350 a month, 7 minutes from campus
>
> Both have one bedroom and a kitchenette. You can choose now between the two apartments or you can continue to search for apartments (to be selected at random from the list you received). In that case, there is some risk of losing one or both of the apartments you have found.
>
> Other subjects received a similar problem except that option y was replaced by option x', to yield a choice between:
>
> Dominance:
> x) $290 a month, 25 minutes from campus
> x') $330 a month, 25 minutes from campus
>
> Note that in both pairs of problems the choice between x and y—the conflict condition—is nontrivial because the x's are better on one dimension and the y's are better on the other. In contrast, the choice between x and x'—the dominance condition—involves no conflict because the former strictly dominates the latter. Thus, while there is no obvious reason to choose one option over the other in the conflict condition, there is a decisive argument for preferring one of the two alternatives in the dominance condition.
>
> On average, subjects requested an additional alternative 64% of the time in the conflict condition, and only 40% of the time in the dominance condition ($p < .05$). Subjects' tendency to search for additional options, in other words, was greater when the choice among alternatives was harder to rationalize, than when there was a compelling reason and the decision was easy.

Why is the greater propensity to search further under the conflict condition than under the dominance condition inconsistent with the utility-maximization model? Clearly $U(x) > U(x')$ for all subjects; it is indeterminate whether subjects would assess $U(x) > U(y)$ or $U(y) > U(x)$. But if the decision to keep searching were based simply on some perceived continuation utility, V, then any subject who would choose to stop under the dominance condition, indicating that $U(x) \geq V$, would *also* choose to stop under the conflict condition, since she

could always choose x where $U(x) \geq V$, or choose y, if $U(y) > U(x)$, and do even better.[98] The aggregate data reject the value-maximization hypothesis.[99]

What explains the pattern? Tversky and Shafir argue that the choice in the conflict condition is difficult, and lacks a clear reason to choose x versus y—one is cheaper, the other is closer to campus, and it is hard to weigh these two attractive features. The dominance condition, on the other hand, yields an obvious choice—x is just as close as x′ and is cheaper. This is a simple, obvious "reason" to choose x. The experiment suggests that search behavior may be explained in part by the search for "reasons" rather than solely by the search for "value."

Tversky and Shafir (1992a) show that, even if the information will *not* affect their decision, people are likely to delay decisions until they learn information that may affect their *reason* for making the decision. Students were asked, under different hypothetical scenarios, whether they would like to buy a bargain package for a Hawaii vacation during winter break. Prior to the vacation, they would find out whether they had passed or failed an important exam; if they failed, they would have to re-take the exam after winter break. Students were given the choice to buy the vacation package, not buy the package, or pay a non-refundable $5 fee to delay the decision. Three conditions were run: The students were told that a) They would not know whether they passed the exam when they had to choose—but if they bought the "delay option," they *would* then know whether they passed before they ultimately had to decide. (b) They failed the exam, and would have to re-take after winter break. (c) They passed the exam.

Of those asked in condition (a), 32% chose to buy the package, 7% chose not to buy the package, and 61% chose to pay $5 to delay the decision. The delay option makes perfect sense: If your decision whether to take the vacation depends on whether or not you passed the exam, you may pay to wait to find out first. However, the results in conditions (b) and (c) rule out this explanation—only 31% in *each* group chose the delay option, and between 54% and 57% chose to buy the vacation (and, respectively, 16% and 12% chose not to take the package). That is, a majority of both passers and failers would choose to buy the vacation, yet without knowing, far fewer would choose to buy the vacation.

Tversky and Shafir argue that the key is that the *reason* for going on the vacation is very different depending on whether or not you pass—either you are going as a reward for a job well done, or to refresh yourself before retaking the exam. Like the apartment-search example, this example illustrates that people choose to defer decisions in part because they seek a clear-cut reason for their choice, not just to get the highest value out of their choice.

---

[98] The differences in behavior in different cases is unlikely to be due to perceived differences in information as a function of the initial choices; subjects were first given the entire set of 12 options apartments to review, and were told that further choices if they continued to search would be chosen at random from this menu.

[99] Redelmeier and Shafir (1995) report related results. Groups of physicians were randomly picked to respond to each of two hypothetical scenarios: A) Whether to prescribe one treatment in one category of treatments (e.g., medication) or one treatment in a second category (e.g., surgery), or B) Whether to prescribe one *of two* treatments in the first category, or one treatment in the second category. They found that more physicians chose the second

Motivated Cognition and Self-Serving Biases

A robust finding of psychological research is that people have "self-enhancing biases" in a range of matters. We are over-optimistic regarding our health and other aspects of our life (Weinstein (1980, 1984) and Kunda (1987)); we feel we are less vulnerable to risk than others (Perloff and Fetzer (1986)); and we are more responsible for our successes than we are for our failures (Miller (1976)). We think that we are superior to others in all sorts of ways (Klein and Kunda (1993)): We are better at controlling risk (Klein and Kunda (1994)), better drivers (Svenson (1981)), and more ethical (Messick, Bloom, Boldizar, and Samuelson (1985), Liebrand, Messick, and Wolters (1986), Tyson (1990), and Morgan (1993)).

The source of self-enhancing biases has been much debated among psychologists. While at first blush such biases would seem to come from such emotions as wishful thinking or anxiety reduction, the cognitive revolution in social psychology has persistently challenged such "motivational" interpretations. Biases are motivated if we have them because our thinking is affected by what we want to be true; biases are unmotivated if they are solely artifacts of heuristic errors, unaffected by our motivations. Self-serving biases may be unmotivated. We may believe we are better-than-average drivers not because we find this an attractive belief, but because the precautions we take are more salient to us than are the precautions other people take. Many researchers have shown that some self-enhancing biases can be explained in part by the type of unmotivated cognitive errors reviewed in Section 3.[100]

But there is strong evidence that self-serving biases are often motivated.[101] The beliefs we form are intertwined with the preferences we have over those beliefs. For instance, people who take risks may deceive themselves into thinking their behavior isn't risky to avoid anxiety.[102]

---

category in Scenario B—apparently because the decision felt less arbitrary than choosing one of two, hard-to-distinguish choices.

[100] For a good example of this perspective, see Miller and Ross (1975), and especially Nisbett and Ross (1980). Some biases can not be explained on motivational grounds. For instance, an often-cited paper by Ross and Sicoly (1979) found that subjects tended to believe that they started more fights than did their spouses; this belief is a bias, but goes against the motivation to believe that we are superior to others (the authors explain the finding in terms of the greater salience or memorability of having started a fight yourself).

[101] See, e.g., Miller (1976), Bradley (1978), Zuckerman (1979), Kunda (1987, 1990, 1992), Pyszczynski and Greenberg (1987), Wood (1989), Sanitioso and Kunda (1991), Klein and Kunda (1993), and Mackie and Hamilton (1993). While attempts by economists (see, Akerlof and Dickens (1982), Dickens (1986), Akerlof (1989), Rabin (1994), and Montgomery (1994)) to incorporate biased information processing have generally invoked the concept "cognitive dissonance", cognitive dissonance, as both a buzz word and a unifying framework, has fallen out of favor among psychologists in recent years; recent strands of psychological research with similar themes are presented under the more general rubric of motivated cognition. (See Kunda (1992) for a discussion of the relationship between earlier cognitive-dissonance research and motivated-cognition research.)

[102] Or, often, simply exaggerate the risks less than people who don't engage in the activity. There are many papers claiming that cognitive dissonance is in evidence merely by finding such correlation between beliefs and behavior, without particularly worrying that (as economists would immediately hypothesize) the causation runs in the opposite direction. Smokers may believe smoking is less dangerous than non-smokers because those who are most worried about the health risks don't start smoking. For a somewhat cleaner laboratory test of whether people self-protectively interpret evidence of dangers, see Kunda (1987).

The literature on selective exposure examines how people avoid information that they find unappealing, and demonstrates that people often prefer to avoid or distort information that challenges comfortable beliefs.[103] This contrasts with the standard economic presumption that people care about information only for its instrumental value, and hence always prefer more of it to less. If the headline indicates that an article will change your beliefs in ways you don't like, you may avoid reading the article.

Are people aware of their own information-processing strategies and biases? Bayesian information processing demands that if an agent sees enough of a signal to judge the implication of the full signal, she must update her beliefs in expected terms even if she does not choose to see the remainder of the signal. Much research explores how people avoid signals they don't like (e.g., that people avoid reading a favorable article about a politician they don't like) without addressing this paradox. More recently, however, some research has addressed the issue. People may "subconsciously" screen out unattractive signals. Gur and Sackiem (1979, p. 162), for instance, cleverly establish that people deceive themselves by screening from their consciousness images that they find unpleasant, apparently recognizing at some level the true content of the signal. They first note that people who have low self-esteem are known to dislike to "confront themselves":

> Recent studies have shown that experimental manipulations of cognitive discrepancy influence selective exposure to the self. [Researchers] (cited in Duval & Wicklund, 1972, pp. 16-21) found that subjects who received prior negative false feedback about themselves departed from a room in which they were confronted with a mirror sooner than subjects who received prior positive false feedback or subjects who were not confronted with a mirror. ... [Researchers] found that self-confronted subjects who received positive false feedback concerning a bogus test of creativity emitted more first person pronouns, when asked to guess English translations of foreign words, than subjects who received negative false feedback or subjects who were not self-confronted. The results of these studies indicate that after experiences of negative feedback, self-esteem is lowered, and confrontation with the self becomes more aversive.

They use this presumption to test whether people with low self-esteem would avoid recognizing themselves. They gave two groups of subjects (false) feedback on how well they did on a short multiple-choice test they were told was designed to measure intelligence.[104] Then, (replicating an earlier study) subjects were asked to listen to a quick series of recorded voices including their own (recorded earlier) and guess which voices were their own. Subjects who had been given the false negative feedback were less likely to misidentify other voices as their own, and more likely to misidentify their own voices as not their own. Using physiological measurements, Gur and Sackiem establish that subjects "at some level" recognized their own voices; apparently, when they found their own voices unpleasant, they were able to hear it, and then immediately screen it from their consciousness.[105]

---

[103] See Cotton (1985) for a review.

[104] Subjects were actually given accurate feedback on how many answers they got right or wrong—but one group was given a very hard test, the other a very easy test, so that performance was a misleading indicator of aptitude.

[105] Another mechanism for selective exposure is probably salience- or memory-based—we know that if we read a positive magazine article about somebody we don't like, the positive images will stick with us; the momentary decision not to read an article will not stick with us. More generally, experimental evidence indicates that we can

Motivational factors have also been implicated in research on the just-world hypothesis, blaming the victim, and victim derogation. We believe that life is fairer than it is, and that people get what they deserve. This tendency is at its nastiest when it leads us to blame victims who might more reasonably be considered deserving of compassion. It has been hypothesized that we blame victims to alleviate our guilt for having caused their problems, or to let ourselves off the hook in helping.[106] We may also prefer to believe in a just world because it is scary to believe innocent people are victims—we might be next.[107]

In Section 2, I posited that people have a "preference" for fairness. Yet one form of self-serving bias calls into question whether concerns for fairness are necessarily simple "preferences." A prominent pattern in research on norms of fairness is that, in case-by-case applications of general norms, our sense of fairness is often tilted towards that which serves our own self-interest (see Leventhal and Anderson (1970), Kelley and Thibaut (1978) and Messick and Sentis (1983)). The frequent attempts we make to mold our sense of fairness to suit our self interest suggests that "being fair" is in some sense not a component of preferences in the same qualitative way as self interest is.[108]

Self-serving biases in fairness judgments arise in day-to-day economic, social, and legal interactions, and some recent research suggests that such fairness biases may importantly exacerbate legal and economic disputes. Loewenstein, Issacharoff, Camerer, and Babcock (1993), for instance, demonstrated that subjects given roles (and real stakes) in a hypothetical legal dispute were apt to interpret evidence self-servingly. Subjects manifestly cared about the fairness of the outcome, but, when reading a description of the case, their assessments of fairness depended on which role (plaintiff or defendant) they knew they were going to play in the ensuing negotiations. The researchers show that the biased reading increased the costliness of disputes.[109]

If people process information motivationally, then the timing of information arrival may be very important. Whether we get information before or after we know our "preferences" can heavily influence how we process the information. For instance, Loewenstein, Issacharoff, Camerer, and Babcock (1993), in a variant of the experiment discussed in the previous paragraph, informed some subjects before reading the case whether they were defendants or plaintiffs. Other subjects were not told of their role until after they read the case (but before they began

---

be very unaware of how we process information. See Nisbett and Wilson (1977) for a review of the general evidence that we are often unaware of how we have come to our conclusions.

[106] E.g., Lerner and Matthews (1967) experimentally demonstrated that subjects who were led to believe they were responsible for a bad thing happening to another person tended to derogate that other person.

[107] See Lerner and Miller (1978) for an excellent discussion of research on the motivations behind the just-world hypothesis.

[108] Rabin (1995) further discusses psychological evidence along these lines, and develops a formal model of biases in ethical perceptions that attempts to account for these features.

[109] For other research on self-serving biases in negotiations, see Thompson (1990), Thompson and Loewenstein (1992), Babcock and Olsen (1992), Kramer, Newton, and Pommerenke (1993), Camerer and Loewenstein (1993), Babcock, Wang, and Loewenstein (1996), and Babcock and Loewenstein (forthcoming). (Camerer and Knez (1995) contains related results.)

negotiating). In this latter group, dispute costs were far lower. Without knowing where their self-interest lay, subjects read the evidence disinterestedly, and could not quickly bias themselves when it came time to negotiate.[110]

Time-Variant Preferences

People like to experience rewards soon and to delay costs until later. We procrastinate on tasks that involve immediate costs and delayed rewards (mowing the lawn) and do things soon when they involve immediate rewards and delayed costs (seeing a movie). Economists traditionally model such tastes by assuming that people discount streams of utility over time exponentially. An important qualitative feature of exponential discounting is that it implies that a person's intertemporal preferences are time-consistent: A person feels the same about a given intertemporal tradeoff no matter when she is asked.[111]

Casual observation, introspection, and psychological research all suggest that the assumption of time-consistency is importantly wrong.[112] People have a short-term tendency to pursue their immediate well-being in a way that their "long-run selves" do not appreciate. While today we feel that it is best that we not overeat tomorrow, tomorrow we tend to overeat; while today we feel we should write a referee report tomorrow, tomorrow we tend to put it off. More generally, when considering tradeoffs between two future moments, we give stronger relative weight to the earlier moment as it gets closer.

The preference for immediate gratification in large part reflects psychological findings that, to a close approximation, people discount future streams of payoffs hyperbolically rather than exponentially.[113] As Loewenstein (1996, p. 279) summarizes it:

> The non-exponential discounting perspective has been bolstered by findings from hundreds of experiments showing that humans and a wide range of other animals display hyperbolic discount functions of the type predicted to produce impulsive behavior (see, e.g., Chung and Herrnstein, 1967; Mazur, 1987). Many experiments with animals, and a small number with humans, have also demonstrated the types of temporally based preference reversals that are implied by hyperbolic discounting.

---

[110] For a related finding in the role of timing of information on willingness to punish unfairness, see Boles and Messick (1990).

[111] Moreover, under some innocuous ancillary stationarity assumptions, exponential discounting is the *only* form of discounting that leads to time-consistent intertemporal preferences.

[112] Psychological research showing preferences are not time-consistent includes Chung and Herrnstein (1967), Ainslie (1974, 1975, 1987, 1991, 1992), Ainslie and Herrnstein (1981), Thaler (1981), Funder and Block (1989), Hoch and Loewenstein (1991), Ainslie and Haslam (1992a, 1992b), and Loewenstein and Prelec (1992).

[113] Ainslie (1992) observes that hyperbolic discounting is connected to formulations of a "matching law" of intertemporal choice (see Chung and Herrnstein (1967) and Ainslie and Herrnstein (1981)), which is also invoked in the formulation of the theory of melioration discussed earlier. See Loewenstein and Prelec (1992) for a generalized hyperbolic discount function.

Kirby and Herrnstein (1995), for instance, asked subjects to state their preferences among a series of pairs, in each case choosing between a smaller, earlier reward and a larger, later reward. Subjects were (truthfully) told that one of their choices would be implemented. In two experiments with monetary rewards, 23 of 24 subjects "consistently reversed their choices from the smaller, earlier reward to the later, larger reward as the delay to both rewards increased." (The 24th subject never reversed her choice, always preferring the smaller, earlier reward.) Both the monetary stakes and the delays were substantial—subjects received an average of about $21.50, with an average delay of about 2½ weeks.[114] Kirby and Herrnstein interpret these data as consistent with (generalized) hyperbolic discounting, and inconsistent with the standard economics model of exponential discounting.

The most relevant feature of hyperbolic discounting is that, at every moment, a person discounts near-term incremental delays more severely than she discounts distant-future incremental delays. Hence, a person's preferences today over her future delays are different than her future preferences over those same delays, so that preferences are *not* time consistent. A small set of economists and psychologists have over the years proposed formal models of such time-*variant* preferences.[115] Phelps and Pollak (1968) capture the taste for immediate gratification with a simple, non-hyperbolic, two-parameter model that slightly modifies exponential discounting. Let u_t be the instantaneous utility a person gets in period t. Then her intertemporal preferences at time t, U^t, can be represented by the following utility function:

$$\text{For all t, } U^t(u_t,u_{t+1},\ldots,u_T) \equiv (\delta)^t \cdot u_t + \beta \cdot \sum_{\tau=t+1}^{T} (\delta)^\tau \cdot u_\tau$$

If $\beta = 1$, then these preferences are simply (the discrete version of) exponential discounting. But for $\beta < 1$, these preferences capture in a parsimonious way the type of time-inconsistent preferences so widely observed.[116]

To see how these preferences capture the preference for immediate gratification, suppose that you had a choice between doing 7 hours of an unpleasant task on April 1, versus spending 8 hours to complete the same task on April 15.[117] Assume that your instantaneous disutility from doing work is simply the number of hours of work—$u_t(7) = -7$ and $u_t(8) = -8$ for all t—and that $\delta = 1$, so that there is no "consistent" discounting (at least not over the very short durations that we are considering here). Suppose, however, that $\beta = .8$ for a two-week delay: You are willing to suffer a given loss in utility two weeks hence for a gain in utility today that is 80% as large.

---

[114] These numbers are calculated from the data presented by Kirby and Herrnstein (1995, p. 85-6).

[115] For early economics papers on time-inconsistent discounting, see Strotz (1955), Koopmans (1960), Phelps and Pollak (1968), Pollak (1968), and Goldman (1979, 1980). See also Elster (1977), Schelling (1978), Thaler and Shefrin (1981), Akerlof (1991), Laibson (1994), and O'Donoghue and Rabin (1996). For economics papers modeling time-variant tastes more generally, see Koopmans (1960), Pollak (1970, 1978), Peleg and Yaari (1973), Hammond (1976a,1976b), Yaari (1977), Pessemier (1978), and Glazer and Weiss (1992).

[116] The parameter δ determines how "time-consistently patient" a person is, just as in exponential discounting.

[117] The task could be completing your taxes—on April 15 you would have to take the half-hour extra ride to the post office to mail your returns, whereas on April 1 you could simply mail it without hassle on your way to work the next day.

Now suppose that April 1 has arrived and you are considering whether or not to work. You can experience a disutility of -7 by working today, or experience a discounted utility of $.8 \cdot (-8) = -6.4$ by delaying the work until 2 weeks from now. You will, therefore, delay work. Contrast this with what your decision would be if, instead of choosing when to work on April 1, you are told by your boss that you must decide on February 1. Now, because from February 1 you discount *both* dates by β, you will choose to work 7 hours on April 1 rather than 8 hours on April 15. From the February 1 point of view, you find procrastinating in April an undesirable thing. For the exact same problem, your choice on April 1 is different than your choice on February 1. Irrespective of its specific prediction, exponential discounting would predict that your choice would be the same whether you made that choice on February 1 or April 1.

The preferences I have hypothetically attributed to you resemble most people's preferences. On April 1, most of us are apt to put off the work until April 15, even if it means a little more work. Virtually no one would choose the delay if asked on February 1.

To examine dynamic choice given time-variant preferences, researchers have converged on a simple modeling strategy: For each point in time, a person is modeled as a separate "agent" who chooses her current behavior to maximize her current long-run preferences, whereas each of her future selves, with her own preferences, will choose her future behavior to maximize *her* preferences. On one level, this idea of multiple selves—that a single human does not have unified preferences which she maximizes—is a radical departure from the utility-maximization framework.[118] But because this conceptualization of intertemporal choice uses familiar economics tools (dynamic game theory), it is ready-made for adoption by economists interested in improving the behavioral realism of our models.

The behavior predicted by models of time-variant preferences often differs dramatically from the behavior predicted by the exponential model. The most notorious examples are efforts at *self control*: Because you may not like the way you will behave in the future, you often scheme to control yourself by restricting your future options. Consider again the work example. Instead of your boss telling you that you *must* choose on February 1 when to work, suppose now she gives you three options: You commit to do the task on April 1; you choose to promise her to do the task on April 15; or you *wait* until April 1 and *then* choose on which day to do the task. Which would you choose? The advantage of waiting is manifest: By not precluding either of your options, if there are *any* uncertainties that may be resolved between now and April, the flexibility you have retained may be valuable. Yet we sometimes engage in behavior precisely *to* restrict our own future flexibility. If there were few uncertainties, you might want to commit on February 1 to the April 1 date. Given your current preference to do the task earlier, you wish to restrict your future self from procrastinating. More generally, researchers have explored many *self-commitment* devices we employ to limit our future choices. Such self-commitment devices include alcohol clinics and fat farms from which you cannot check out, not owning a television, contributing to a "Christmas Club" from which you are not allowed to withdraw money until Christmas, or buying only small packages of enticing foods so

[118] For a broader perspective on "multiple-self" models of human agency, see Elster (1985).

that you won't over-eat when you get home. More subtly, you may try to control yourself through a variety of internal "rules" (e.g., never drink alcohol), even if you have no external mechanisms of self control.[119]

Attempts to control our own future behavior indicate an awareness that we may not behave as we would wish to behave. This raises the question of how aware people are of their time-inconsistency. Simply because you find your propensity to over-respond to salient desires disagreeable does not preclude your having "rational expectations" about this propensity. You may *wish* your future self wouldn't behave the way she will, but you may still predict it. On the other hand, you may naively believe that your preferences in the future will match your current preferences. If today you prefer not to overeat tomorrow, you may naively believe that you will feel the same way when facing an enticing bowl of ice cream tomorrow. If on February 1 you prefer working on April 1 to April 15, you may believe you'll feel the same way in April.

People who are fully aware of their future self-control problems can be labeled *sophisticated*, and people who are fully unaware that they will have a self-control problem are *naïve*.[120] Since we ubiquitously assume that people have perfect foresight, it is natural for economists studying time-inconsistent preferences to assume sophistication. But beyond this general habit, there is strong evidence of foresight in the context of time-variant preferences. As noted above, some degree of sophistication is implied by the fact that people want to commit themselves to smaller choice sets: If you were naive, you would never worry that your tomorrow self might choose an option you don't currently like, and therefore you would find committing yourself unattractive. On the other hand, it does appear that people underestimate the degree to which their future behavior will not match their current preferences over future behavior. This accords with the evidence discussed in the first subsection of Section 4, that people often incorrectly predict their own future preferences: As with predicting the effects of changes in reference points, here too knowing your future preferences means that you know your preferences *won't* accord with your current preference. As in Loewenstein and Adler (1995), people may not realize how changes in perspective will change their preferences.[121] For example, people will repeatedly not have the "will power" to forego tempting foods or quit smoking while predicting that *tomorrow* they will have this will power.

---

[119] For analyses of self-control in humans, see Ainslie (1975,1987,1992), Ainslie and Haslam (1992b), Funder and Block (1989), Glazer and Weiss (1992), Hoch and Loewenstein (1991), Laibson (1994, 1995), Schelling (1978,1984,1992a), Shefrin and Thaler (1992), Thaler (1980), Thaler and Shefrin (1981), and Wertenbroch (1993). Ainslie (1974) explores similar issues in pigeons.

[120] Strotz (1955) and Pollak (1968) carefully lay out these two assumptions (and coined the terms), but do not much consider the implications of one versus the other. More recent papers have assumed one or the other, without attempting to justify the choice with behavioral evidence. For instance, Akerlof (1991) assumes naive beliefs, while Laibson (1994,1995) assumes sophisticated beliefs. O'Donoghue and Rabin (1996) consider both, and explicitly contrast the two, but likewise do not provide behavioral evidence for either.

[121] I am not aware of behavioral evidence that calibrates the degree of sophistication, so that this remains merely an impression. Loewenstein (1996, pp. 281-2) reaches a comparable conclusion (stated in Proposition 5) as to the evidence that people may be naïve, also in part based on the phenomena described in the first subsection of this section. Yet he too draws the conclusion indirectly from some other psychological findings, and does not discuss any data that directly address the question.

Whether they are sophisticated or naïve, people's time-inconsistent propensity for immediate gratification is important in a variety of economic realms.[122] Such preferences may be important to consumption and savings decisions because the benefits of current consumption are immediate, whereas the increased future consumption that saving allows is delayed. Several researchers have developed models of time inconsistency in savings behavior, where it is generally understood that people have a higher propensity to consume than is in their long-run self interest.[123] Self-control problems are also clearly important in the demand for addictive goods.[124] Addiction involves pursuit today of a salient desire that has future costs. Addictive goods are but one context where there are negative internalities (a concept developed by Herrnstein and Prelec (1992 Chap. 10), reviewed above), where consumption of a good today causes harm tomorrow. Many implications of time-inconsistent preferences are likely to be similar for non-addictive goods that exhibit negative internalities, such as fatty food and the many harmful but non-addictive drugs.

The role of self-control in purchasing decisions is well-known and much studied by marketing experts. Naughty goods are sold in small packages, because people tend not to buy large packages of such goods, knowing they have a tendency to over-consume. Marketing consumer goods is also a realm where the type of non-temporal salience and impulsiveness ignored in this paper (e.,g., having goods enticingly packaged, aromatically vivid, and prominently displayed at checkout counters) play an immense role.[125]

Some problems arise with the welfare economics of time-variant preferences. If a person's preferences disagree at different times, a change in a consumption profile may help some selves while hurting other selves.

---

[122] We might also ask, for given preferences and in a given context, how sophisticated beliefs compare to naïve beliefs. It is intuitive that sophisticated people should fare better than naïve people, since being aware of future self-control problems should help you avoid them. This intuition is valid if costs are salient, because sophistication heightens your awareness of how costly delay is, whereas naiveté can lead you to repeatedly procrastinate under the false belief that you will do it tomorrow. Moreover, these delays can be extremely damaging to your long-run well-being, even if the incremental cost of each decision to delay is very small. See Akerlof (1991) and O'Donoghue and Rabin (1996) for formalizations of this claim.

An implicit result in some research, explored directly by O'Donoghue and Rabin (1996), is that sometimes naiveté can be better than sophistication. Suppose you have a resource that will grow in value the longer you resist consuming it. If you are sophisticated, you know you will behave poorly in the future if you wait, and, because this makes you (properly) skeptical of the benefits of waiting, you are more tempted to consume the object today. This can lead to "unwinding" of the sort often associated with the finitely repeated prisoners' dilemma: In the end, you will give in to temptation and grab a prize too soon; because you realize this, near the end you will cave in a little sooner than if you thought you would resist temptation in the end; realizing this you will cave in a little sooner, etc. Naiveté always prevents such unwinding: By being over-optimistic that you will behave yourself in the future, you think that the benefits of foregoing consumption now are greater.

[123] See, for instance, Strotz (1955), Phelps and Pollak (1968), Pollak (1968), Thaler and Shefrin (1981), Shefrin and Thaler (1988, 1992), Bernheim (1994), Laibson (1994, 1995), and Thaler (1994).

[124] Becker, Grossman and Murphy (1991, 1994) and Becker and Murphy (1988) have proposed models of "rational addiction," where people with time-consistent intertemporal preferences get addicted because they choose to. These models insightfully formalize the essence of (bad) addictive goods: Consuming more of the good today decreases overall utility but increases marginal utility for consumption of the same good tomorrow. However, these models *a priori* rule out time-inconsistency and self-control issues. For more realistic papers on addiction, see Herrnstein and Prelec (1992, Chap. 13), Heyman (1994), Orford (1985), and Schelling (1992b).

The natural approach in most situations is the "long-run perspective": What would you wish now (if you were fully informed) about your profile of future behavior? Some economists are more agnostic: They have employed (as economists are wont to do) "Pareto-comparisons," where one outcome for a person is deemed unambiguously better than another only if it is better from the point of view of *all* of a person's preferences. This sometimes allows comparisons of outcomes: All of your selves, even your addicted selves, are likely to agree that a lifetime of smoking is worse than a lifetime of non-smoking; they simply disagree about who should bear the brunt of quitting. But as with Pareto comparisons in general, reasonable and powerful welfare comparisons are often proscribed. Therefore, we need to posit "reasonable intrapersonal welfare functions" to usefully study the welfare consequences of different outcomes.[126]

5. Discussion and Conclusion

In this section, I briefly address some of the reasons and rationalizations that economists continue to express for ignoring psychological research. I then conclude with some thoughts on how psychological research might be integrated into economics.

Should We Continue to Ignore Psychological Research?

Economists have (until very recently) been fundamentally dismissive of laboratory evidence. A major complaint has been that the stakes involved are generally too small to be used for predicting real-world behavior; indeed, some psychological research consists of hypothetical choices in which subjects report what they *would* do in particular situations, or of questionnaires where subjects are asked to give answers without being provided with extrinsic incentives for correct answers. The "stakes debate" has, if anything, been exacerbated by the advent of experimental economics, which has at times deemed the use of financial incentives a methodological imperative. The view that responses elicited from subjects will be informative only if substantial financial stakes are involved is an *empirical hypothesis*. It is also a testable hypothesis. And it is, by and large, a false hypothesis. Camerer (1995), in by far the most comprehensive and integrated review of psychological and economic experimental

---

[125] For discussions of self control in the context of consumer choice, see Thaler (1980), Rook (1987), Hoch and Loewenstein (1991), and Wertenbroch (1993).

[126] See Schelling (1984) for a thoughtful discussion of some related issues. While I think that on balance the long-run self is "right" in terms of experienced utility, there are exceptions. For instance, people may be too inattentive to the pain induced by following through on plans that they deem proper for their long-run selves.

evidence on individual decision-making, finds *very* few instances where financial-stakes replications of psychological experiments have reversed the inferences about theory drawn from those experiments.

While there are some areas (e.g., departures from self interest) where answers to hypothetical questions are likely to be misleading, our goal should be to develop a psychologically sophisticated sense of *when* facts about people can be accurately elicited in low-stakes contexts, rather than putting forth a methodological proclamation that such elicitation procedures are inadequate. As the list of examples (preference reversals, ultimatum-game sharing) lengthens where psychological findings are robust to the scale of financial stakes, perhaps we can converge to the point where in certain domains we accept the presumptive validity of low-stakes or hypothetical-question evidence.[127]

A second common critique of psychology experiments is that participants are amateurs at the tasks they face, and are not given ample opportunity to learn. If nearly all important economic activity involves people experienced with the decisions they face, then the heavy reliance by psychologists on relatively inexperienced subjects is (it is argued) problematic. This critique has some legitimacy in some contexts. Yet psychologists do study the effects of learning and expertise; recall that the evidence presented at the end of Section 3 shows that some departures from rationality are as strong or stronger among experienced as among inexperienced subjects. Indeed, many psychologists feel that the type of repetitive environments experimental economists create in the laboratory are *unrealistically* conducive to learning.[128]

These differing intuitions about the environments most common in the real world raise an important issue: Even if it were true that "sufficient" stakes and "sufficient" learning eliminate irrationalities, it is tenuous to argue that only the behavior of experienced people engaging in large-stakes activities is economically important. The sum-total of "low-stakes" transactions (e.g., retail purchases of cola) is a big part of the economy, and the behavior of "novices" (e.g., new home buyers) is something that economists should be curious about.[129] Assertions that economics ought concern itself mostly with repetitive, large-stakes transactions warrant more careful examination than they have received.[130]

---

[127] And in those cases where on average people are *more* rational and their behavior is *more* self-interested when stakes are large, economists should avoid the substantive or rhetorical leap to the conclusion that they are *nearly perfectly* rational or *nearly fully* self-interested. It should be stressed that attempts so far to push stakes high enough to *even significantly decrease* experimental departures from rationality and self interest have been surprisingly unsuccessful. For instance, Cameron (1995) conducted experiments in Indonesia on the ultimatum game with stakes that were one fourth the annual expenditures of the average participant. No reduction relative to typical experimental results in the propensity to reject a given percentage offer were found. See also Fehr and Tougareva (1996).

[128] Camerer (1995, p. 600) notes that "many psychologists think stationary replication overstates the frequency, speed, and clarity of feedback the world actually provides."

[129] See Smith and Walker (1993) for a nice articulation of this point.

[130] Partly in reaction to the increased recognition that full rationality does not accurately describe people, there has recently been a renewed interest in game theory and microeconomics on learning and long-run dynamics without strong rationality assumptions. (See, e.g., Fudenberg and Levine (1993), Ellison (1993), Kalai and Lehrer (1993), Young (1993), Kandori, Malaith, and Rob (1993), and Ely (1995).) There is, of course, a long tradition in economics of invoking "the long run" as the relevant case where our assumptions and models are more likely to

A third argument against the relevance of psychological findings is that competitive markets will pervasively eliminate the effects of departures from rationality, self interest, or other familiar assumptions. This argument is logically and empirically wrong. Akerlof and Yellen (1985), for instance, show that departures from pure rationality or pure self interest can have important market consequences even when they leave individuals with virtually no incentive to alter their own conduct. Haltiwanger and Waldman (1985) explore which types of markets are most likely to reduce irrationalities, and which are likely to *exacerbate* them. In all but a limited class of financial and spot markets, it is simply not clear how a general principle that markets eliminate irrationalities would operate. In financial markets, there is some logic and empirical evidence that arbitrage opportunities and other market forces mitigate many types of irrationalities; but even in these settings, the evidence is far from overwhelming.[131]

Another problem with invoking the markets-negate-irrationalities hypothesis is, of course, that much important economic activity is *not* mediated fully by competitive markets. Most economists believe that there are important departures from perfect competition, and the primary approaches to fields such as labor economics, industrial organization, and macroeconomics do not assume that all economic activity is conducted through perfectly competitive markets.

A final problem with the assertion that markets wipe out unfamiliar psychological facts is that its presumed implication—that we ought to ignore these psychological facts—is a *non sequitur*. Even if we believe virtually all economic activity (in developed capitalist societies) is mediated by competitive markets, we should recall that a major lesson of economics—traditionally, the *core* lesson—has been that competitive markets are often superior to other institutions in allocating resources. In light of such comparative institutional analysis, if competitive markets are uniquely prone to eradicating the implications of unfamiliar psychological facts, then clearly we should very much care about these facts. If (say) markets correct irrational biases that other institutions don't, then this fact should be given a prominent place on the list of virtues of market allocation. If (say) markets de-sensitize people to concerns for fairness or eradicate its implications, then this shouldn't tell us we can ignore social motivations. It should tell us the *opposite*: We need to be highly attentive to how markets modify the implications of social motivations before we can feel we've made a fair comparison of markets to other allocation mechanisms.

Many psychological findings, whatever their true economic implications, may not be very compatible with mainstream economic methods. As such, economists have "methodological" objections to incorporating

---

hold, and (therefore?) the case we should focus on. Here is not the forum for debating whether the long run is the dominant case to which economics ought address itself. Instead, I invoke the authority of Keynes (1924, p. 88): "Now 'in the long run' this [way of summarizing the quantity theory of money] is probably true....But this *long run* is misleading guide to current affairs. *In the long run* we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is long past the ocean is flat again."

[131] For empirical investigations challenging the markets-negate-irrationalities hypothesis in financial markets, see also DeBondt and Thaler (1985, 1987, 1990); and for more theoretical investigations, see De Long, Shleifer, Summers, Waldmann, (1989, 1990a, 1990b, 1991) and Summers (1993). For experimental tests, see Camerer (1987, 1992a) and Camerer, Loewenstein, and Weber (1989).

psychological research into our discipline. The perspective of this essay has been that the simplicity and tractability of hypotheses about people are necessarily criteria by which we select our formal assumptions. Yet economists often seem to use tractability to justify indifference to the realism of our assumptions. I recollect a parable being used to make fun of economists' taste for easy methods of investigation: A drunk man is looking for his wallet at night under a street lamp. Asked whether he lost the wallet near the street lamp, he replies no, he lost it elsewhere; but he is searching for it under the street lamp because it is easier to see there.[132]

A second methodological roadblock to revising familiar assumptions is the view that we should maintain our familiar set of assumptions until somebody comes up with a single alternative set of assumptions which can, at one fell swoop, replace it. Until we find an alternative general theory proven to "deliver the goods" by making strong and precise predictions, it is argued, we should maintain our current general theory.[133] It is counterproductive to demand that explorations of alternatives immediately rank as high in these factors as does the current theory, which has been fleshed out and worked over for decades or even centuries.

A final type of roadblock to incorporating greater psychological realism into economics is that many arguments are *selectively* invoked against unfamiliar assumptions. Even economists who rarely assume perfect competition in their own research often invoke competitive markets as the test for whether alternative assumptions matter. And even economists who would never dismiss a standard model for being too complicated will invoke "parsimony" as a reason to dismiss alternative assumptions.[134]

---

[132] The drunk's strategy is silly when he *knows* the wallet isn't near the street lamp. But when the whereabouts of the wallet are uncertain, even a very sober man would look *first* under a street lamp for his lost wallet, and might do so even if he thinks it is relatively unlikely that he lost the wallet there. Likewise, it is sensible to begin our search for truths about the economy by exploring those avenues that are well-lit by our methods of investigation—to first explore the implications of assumptions that rate highly in terms of consistency and generality, even if other assumptions better match human behavior. But economists seem simply not to be making the tractability-versus-accuracy tradeoff in an appropriate way. Harless and Camerer (1994, p. 1286), in discussing the current state of alternatives to expected utility theory inspired by behavioral research, note that "an historical parallel described by Stigler (1950 [pp. 393-4]) may be instructive for those who cling to EU: 'Economists long delayed in accepting the generalized utility function because of the complications in its mathematical analysis, although no one (except Marshall) questioned its realism...Manageability should mean the ability to bring the theory to bear on specific economic problems, not ease of manipulation. The economist has no right to expect of the universe he explores that its laws are discoverable by the indolent and the unlearned. The faithful adherence for so long to the additive utility function strikes one as showing at least a lack of enterprise.'"

[133] One variant of this attitude is "reactionary Kuhnianism". Thomas Kuhn (1970) hypothesized that sciences tend to form general paradigms that shape our thinking; anomalies inconsistent with the current paradigm tend to be de-emphasized or go unrecognized by scientists. As these anomalies accumulate and become incontrovertible, eventually there is a "paradigm shift" to account for such anomalies. Kuhn's hypothesis is an attempt at description, not prescription. I sense, however, that some researchers find this view of science *normatively* appealing, feeling licensed to wait for a full-fledged "paradigm shift" before being obligated to pay attention to facts that don't fit their current theories.

[134] The ubiquitous invocation of parsimony is especially frustrating given that so often economists respond to *simple* psychological hypotheses by constructing baroque alternative explanations of phenomena relying exclusively on traditional economic assumptions. Often, these alternative explanations lack not only parsimony, but also predictiveness and empirical implementability. When confronted with the simple and intuitive hypothesis, for instance, that people reject unfair deals because they dislike unfair deals, economists instead insist that

How We Can Use Psychological Findings

I have in this essay reviewed psychological research mostly with an eye to the question of whether it has a positive role to play in understanding economic activity. But in some contexts the main lesson of psychological research may simply be the destructive one that we should not put much faith in our traditional assumptions and methods, even if psychology provides no useful alternative. In policy analysis, for instance, behavioral evidence can caution us against pursuing some costly policy premised on conventional but dubious economic assumptions.

Consider the current debate over contingent valuation. Both Federal agencies and the courts have been using surveys of citizens to gather evidence about how people value such non-market goods as environmental pollution or protection of endangered species. This method is sometimes used to help determine damage settlements in law suits, such as the case against Exxon for the *Valdez* oil spill. Research discussed in Section 4 indicates that it is very problematic to infer people's preferences from any particular elicitation procedure, because people's expressed preferences are likely to be highly sensitive to the elicitation procedure itself. While psychological research helps suggest alternative elicitation procedures, *one* lesson of this research may simply be that we should not have faith that any such procedure will meaningfully uncover preferences. Diamond and Hausman (1994, p. 63) reach such a conclusion:

> Thus, we conclude that current contingent valuation methods should not be used for damage assessment or for benefit cost analysis.
>
> It is impossible to conclude definitely that surveys with new methods (or the latest survey that has been done) will not pass internal consistency tests. Yet, we do not see much hope for such success. This skepticism comes from the belief that the internal consistency problems come from an absence of preferences, not a flaw in survey methodology.

Even when psychological research does not yield insights that lend themselves to general conclusions, the research can often play a constructive role by improving economics on an *ad hoc* basis, helping us interpret patterns of behavior context by context.[135] Consider, the research on framing effects discussed in Section 4. It is likely to be hard for mainstream economics to incorporate such research in very broad terms—the very sensitivity of outcomes to particular frames suggests that it will be difficult to reach universal and rigorous conclusions from this research. But awareness of framing effects is likely to help with economic analysis at a more *ad hoc* level, as suggested by many of the economic examples cited in Section 4.

---

complicated reputations stories (which predict that everything can be an equilibrium) be disproved before the simple, intuitive, and parsimonious psychological explanation can be accepted.

[135] When economists describe a model, assertion, or assumption as "*ad hoc*," we all know it is being condemned. The tenth edition of Merriam Webster's Collegiate Dictionary defines the phrase as follows: "1a: concerned with a particular end or purpose <an ad hoc investigating committee> b: formed or used for specific or immediate problems or needs <ad hoc solutions> 2: fashioned from whatever is immediately available: improvised <large ad hoc parades and demonstrations>." None of these definitions has the negative connotations economists associate with the phrase.

One particular use of psychological research has clearly been the main focus of this essay. It is *empirical assumption-making*: We should attempt to replace some of the current assumptions in economics with assumptions built from the systematic patterns of behavior identified by psychological research. Whatever the advantages and disadvantages of rigorously deriving conclusions from formally stated assumptions, our conclusions are likely to be more realistic if our assumptions have a better empirical base. Psychological research has accumulated enough general insights for us to convert these insights into tractable assumptions in formal economics. The material presented in Section 2 is especially ripe for this process. And, as noted at the end of Section 4, simple formal models of time-inconsistent preferences are already here. Developing tractable formalizations of the findings discussed in Section 3 and other parts of Section 4 is likely to be harder; but as some of the citations in those sections indicate, even these show promise.[136]

Psychologists, and now experimental economists, often contend that controlled laboratory experimentation is a more valid source of insights into human behavior than uncontrolled field studies.[137] Conversely, economists have traditionally contended that field evidence provides more insight than laboratory evidence. Many of the tensions between the disciplines seem to come from these faiths in different modes of empirical research. In some domains (e.g., departures from self interest) we must find ways to test experimentally-generated hypotheses outside the laboratory. Indeed, the scale of monetary and other stakes and the degree of experience involved in some economic situations cannot possibly be replicated in the laboratory. Perhaps one of the most fruitful avenues for productive behavioral research is for researchers to use the laboratory to generate hypotheses, and then to use field evidence to test these hypotheses.[138] As reviewed in Section 2, for instance, many researchers, backed both by general analysis and specific laboratory studies, have argued that fairness, equity, and related considerations are important for labor economics in general, and unemployment in particular. Theories of unemployment cannot be adequately tested directly in the laboratory. But armed with experimental evidence about what motivates people, economists can test in the field the theories of unemployment implied by these motivations.

---

[136] The "empirical assumption-making" approach should definitely be employed in the realm of welfare economics. The research identifying departures from rationality should caution us against a simplistic "revealed preference" approach to welfare economics, and consequently should help us formulate some more sophisticated methods for inferring the welfare implications of choice. A research program along these lines was suggested by the material presented in the first subsection of Section 4. The challenge to conventional welfare economics is starkest in some of the other topics in Section 4; what becomes of welfare economics if, for instance, people don't have preferences independent of a choice context?

[137] Every general psychology text contains an introductory chapter exalting the experimental method and treating controlled experiments as a fundamental feature of the discipline, often with the connotation that controlled experiments are what define a discipline as scientific.

[138] See Babcock and Loewenstein (forthcoming) for discussion of some research attempting to combine laboratory and field empirics.

Feedback

The theme of this essay is that there should be more interaction between economists and psychologists. My emphasis throughout has been on how economists would benefit from paying more attention to the insights developed by psychologists. But, just as psychologists should tell economists how true our assumptions are, economists should provide feedback to psychologists about what facets of behavior are most likely to have economic implications—and also tell psychologists where we need to have more precise formulations of behavioral hypotheses. In such a way, economists can play an important role in improving the economic relevance of psychological research.

My own initial attempts at research, for instance, have led me to believe that the economic implications of a person's cognitive biases often will depend crucially upon how much others are aware that she suffers from the bias. Do investors who hire a money manager believe that the money manager suffers from overconfidence? Does a principal hiring an employee believe that the employee is prone to the confirmation bias? It could be that people are well calibrated regarding the flaws in others' judgment, or it may be that people are unaware of others' biases. By the logic of economic models that involve multiple agents, these distinctions are likely to matter. If nobody anticipates the biases of others, then perhaps economic outcomes will be dramatically affected by these biases, but economic institutions won't be designed in response to the biases. If the bias is anticipated by others, on the other hand, institutional design may be altered.[139]

While the question of "social awareness" of cognitive biases may turn out to be central to economic applications, the relevant psychological literature is not conclusive, nor formulated in a way that is most helpful to economists.[140] It is natural for psychologists not to have focused on this issue, if its importance is manifest only in specific economic contexts. As economists start to incorporate insights from psychology, however, it will be natural for us to put forward lists of such issues that are of special importance to us.

---

[139] One possibility is that people may exploit the biases of others, and design contracts to exploit them. Another possibility is that people will wish to counteract the biases of others, and design incentives so as to avoid errors.

[140] There is some related literature; see, e.g., Ross (1987) and Paese and Kinnaly (1993).

REFERENCES

Adams, J. S. (1963). "Wage Inequities, Productivity and Work Quality." Industrial Relations 3: 9-16.

Ainslie, G. (1975). "Specious reward: A behavioral theory of impulsiveness and impulse control." Psychological Bulletin 82(4): 463-496.

Ainslie, G. (1987). "Self-reported tactics of impulse control." International Journal of the Addictions 22(2): 167-179.

Ainslie, G. (1991). "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves." American Economic Review 81(2): 334-340.

Ainslie, G. (1992). Picoeconomics: The strategic interaction of successive motivational states within the person. Cambridge University Press; New York, NY, US.

Ainslie, G. and N. Haslam (1992a). Hyperbolic discounting. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation, New York, NY, US: 57-92.

Ainslie, G. and N. Haslam (1992b). Self-control. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation, New York, NY, US: 177-209.

Ainslie, G. and R. J. Herrnstein (1981). "Preference reversal and delayed reinforcement." Animal Learning & Behavior 9(4): 476-482.

Ainslie, G. W. (1974). "Impulse control in pigeons." Journal of the Experimental Analysis of Behavior 21(3): 485-489.

Akerlof, G. (1982). "Labor Contracts as Partial Gift Exchange." Quarterly Journal of Economics 97(NEovember): 543-569.

Akerlof, G. and W. Dickens (1982). "The Economic Consequence of Cognitive Dissonance." American Economic Review 72: 307-319.

Akerlof, G. A. (1984). "Gift Exchange and Efficiency-Wage Theory: Four Views." American Economic Review 74(2): 79-83.

Akerlof, G. A. (1989). "The Economics Of Illusion." Economics and Politics 1(1 (Spring)): 1-15.

Akerlof, G. A. (1991). "Procrastination and Obedience." American Economic Review 81(2): 1-19.

Akerlof, G. A. and J. L. Yellen (1985). "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?" American Economic Review 75(4): 708-720.

Akerlof, G. A. and J. L. Yellen (1990). "The fair wage-effort hypothesis and unemployment." Quarterly Journal of Economics 105(2): 255-283.

Allais, M. (1953). "Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Américaine." Econometrica 21: 503-546.

Allais, M. (1979). The So-Called Allais Paradox and Rational Decisions Under Uncertainty. The Expected Utility Hypothesis and the Allais Paradox. M. Allais and O. Hagen, Dordrecht: Reidel.

Anderson, C. A., M. R. Lepper and L. Ross (1980). "Perseverance of social theories: The role of explanation in the persistence of discredited information." Journal of Personality & Social Psychology 39(1-6): 1037-1049.

Andreoni, J. (1988a). "Privately Provided Public Goods in a Large Economy: The Limits of Altruism." Journal of Public Economics 35: 57-73.

Andreoni, J. (1988b). "Why Free Ride? Strategies and learning in public goods experiments." Journal of Public Economics 37: 291-304.

Andreoni, J. and J. H. Miller (1995). "Auctions with Artificial Adaptive Agents." Games and Economic Behavior 10: 39-64.

Andreoni, J. and J.H. Miller (1996) "Giving According to GARP: An Experimental Study of Rationality and Altruism." Working Paper: Social Systems Research Institute, University of Wisconsin at Madison.

Arrow, K. J. (1974). Essays in the Theory of Risk Bearing, North Holland.

Asare, S. K. (1992). "The Auditor's Going-Concern Decision: Interaction of Task Variables and the Sequential Processing of Evidence." Accounting Review 67(2): 379-393.

Babcock, L. and G. Loewenstein (forthcoming). "Explaining Bargaining Impasse: The Role of Self-Serving Biases." Journal of Economic Perspectives.

Babcock, L., G. Loewenstein, S. Issacharoff and C. Camerer (1995). "Biased judgments of fairness in bargaining." American Economic Review 85(5): 1337-1343.

Babcock, L., X. Wang and G. Loewenstein (1996). Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias. Quarterly Journal of Economics 111(1): 1-19.

Babcock, L. C. and C. A. Olson (1992). "The Causes of Impasses in Labor Disputes." Industrial Relations 31(2): 348-360.

Banerjee, A. (1990). "Envy." Economic Theory and Policy: Essays in Honor of Dipak Banerjee. Eds.: B. Dutton et al. Oxford: Oxford University Press.

Bar-Hillel, M. (1982). Studies of Representativeness. Judgment under Uncertainty: Heuristics and Biases. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press: 69-83, Chap. 5.

Baron, J. N. (1988). "The Employment Relation as a Social Relation." Journal of the Japanese and International Economies 2(4): 492-525.

Battalio, R. C., J. H. Kagel and K. Jiranyakul (1990). "Testing between Alternative Models of Choice under Uncertainty: Some Initial Results." Journal of Risk and Uncertainty 3(1): 25-50.

Baumann, A. O., R. B. Deber and G. G. Thompson (1991). "Overconfidence among physicians and nurses: The "micro-certainty, macro-uncertainty" phenomenon." Social Science & Medicine 32(2): 167-174.

Baumol, W. J. (1987). Superfairness: Applications and Theory. Cambridge, MA: MIT Press.

Bazerman, M. H., S. B. White and G. F. Loewenstein (1995). "Perceptions of fairness in interpersonal and individual choice situations." Current Directions in Psychological Science 4(2): 39-43.

Beattie, J. and J. Baron (1988). "Confirmation and matching biases in hypothesis testing." Quarterly Journal of Experimental Psychology: Human Experimental Psychology 40A(2): 269-297.

Becker, G. S., M. Grossman and K. M. Murphy (1991). "Rational Addiction and the Effect of Price on Consumption." American Economic Review 81(2): 237-41.

Becker, G. S. and K. M. Murphy (1988). "A Theory of Rational Addiction." Journal of Political Economy 96(4): 675-700.

Becker, M. A. (1981). A Treatise on the Family. Cambridge, MA, Harvard University Press.

Becker, S. W. and F. O. Brownson (1964). "What price ambiguity? On the role of ambiguity in decision-making." Journal of Political Economy 72: 62-73.

Berlin, I. (1994). The magus of the north: J. G. Hamann and the origins of modern irrationalism, Farrar, Straus, and Giroux.

Benartzi, S. and R. H. Thaler (1995). "Myopic loss aversion and the equity premium puzzle." Quarterly Journal of Economics 110(1): 73-92.

Bernasconi, M. and G. Loomes (1992). "Failures of the reduction principle in an Ellsberg-type problem." Theory & Decision 32(1): 77-100.

Bernheim, B. D. (1994). Personal Saving, Information, and Economic Literacy: New Directions for Public Policy. Tax Policy for Economic Growth in the 1990's, American Council for Capital Formation: 53-78.

Berscheid, E., D. Boye and E. Walster (1968). "Retaliation as a means of restoring equity." Journal of Personality & Social Psychology 10(4): 370-376.

Betz, A. L., K. M. Gannon and J. J. Skowronski (1992). "The moment of tenure and the moment of truth: When it pays to be aware of recency effects in social judgments." Social Cognition 10(4): 397-413.

Bewley, T. and W. Brainard (1993). A Depressed Labor Market, as Explained by Participants.

Bewley, T. F. (1986). Knightian decision theory: Part I, Cowles Foundation.

Bigman, D. (1991). "Fairness and Efficiency in a Collective Enterprise." Journal of Economic Behavior and Organization 15(1): 115-135.

Binmore, K., P. Morgan, A. Shaked and J. Sutton (1991). "Do People Exploit Their Bargaining Power? An Experimental Study." Games and Economic Behavior 3: 295-322.

Bishop, J. (1987). "The Recognition and Reward of Employee Performance." Journal of Labor Economics 5(4 (part 2)): S36-S56.

Bjorkman, M. (1994). "Internal cue theory: Calibration and resolution of confidence in general knowledge." Organizational Behavior & Human Decision Processes 58(3): 386-405.

Blinder, A. S. and D. H. Choi (1990). "A Shred of Evidence on Theories of Wage Stickiness." The Quarterly Journal of Economics 1990(November): 1003-1015.

Block, R. A. and D. R. Harper (1991). "Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis." Organizational Behavior & Human Decision Processes 49(2): 188-207.

Blount, S. (1995). "When Social Outcomes Aren't Fair: The effect of causal attributions on preferences." Organizational Behavior and Human Decision Processes 63(2): 131-144.

Bodenhausen, G. V. and M. Lichtenstein (1987). "Social stereotypes and information-processing strategies: The impact of task complexity." Journal of Personality and Social Psychology 52(5): 871-880.

Bodenhausen, G. V. and R. S. Wyer (1985). "Effects of stereotypes in decision making and information-processing strategies." Journal of Personality & Social Psychology 48(2): 267-282.

Boles, T. L. and D. M. Messick (1990). Accepting Unfairness: Temporal Influence on Choice. Contemporary Issues in Decision Making. K. Borcherding, O. I. Larichev and D. M. Messick. Amsterdam, North Holland: 375-389.

Bolton, G. E. (1991). "A Comparative Model of Bargaining: Theory and Evidence." The American Economic Review 81(5): 1096-1136.

Bolton, G. E., E. Katok and R. Zwick (1993). Dictator Game Giving: Rules of Fairness versus Random Acts of Kindness, Penn State University; University of Pittsburgh.

Border, K. C. and U. Segal (1994). "Dynamic Consistency Implies Approximately Expected Utility Preferences." Journal of Economic Theory 63(2): 170-88.

Bowman, D., D. Minehart and M. Rabin (1995). "Loss Aversion in a Consumption/Saving Model."

Bradley, G. W. (1978). "Self-serving biases in the attribution process: A reexamination of the fact or fiction question." Journal of Personality and Social Psychology 36(1): 56-71.

Braun, P. A. and I. Yaniv (1992). "A case study of expert judgment: Economists' probabilities versus base-rate model forecasts." Journal of Behavioral Decision Making 5(3): 217-231.

Brenner, G. (1973). "Pareto Desirable Redistribution: The Case of Malice and Envy." Journal of Public Economics 2: 173-183.

Brenner, R. (1987). Rivalry: In Business Science, and among Nations, Cambridge University Press.

Brickman, P. and J. H. Bryan (1975). "Moral judgment of theft, charity, and third-party transfers that increase or decrease equality." Journal of Personality & Social Psychology 31(1): 156-161.

Brickman, P. and J. H. Bryan (1976). "Equity versus equality as factors in children's moral judgments of thefts, charity, and third-party transfers." Journal of Personality & Social Psychology 34(5): 757-761.

Brickman, P., D. Coates and R. Janoff-Bulman (1978). "Lottery winners and accident victims: Is happiness relative?" Journal of Personality & Social Psychology 36(8): 917-927.

Bruner, J. S. and M. C. Potter (1964). "Interference in Visual Recognition." Science 144(April 24, 1964): 424-425.

Camerer, C. F. (1987). "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." American Economic Review 77(5): 981-997.

Camerer, C. F. (1989). "An Experimental Test of Several Generalized Utility Theories." Journal of Risk and Uncertainty 2(1): 61-104.

Camerer, C. F. (1990). Behavioral game theory. Insights in decision making: A tribute to Hillel J. Einhorn. M. H. Robin, University of Chicago Press, Chicago, IL, US: 311-336.

Camerer, C. F. (1992a). "The rationality of prices and volume in experimental markets. Special Issue: Decision processes in negotiation." Organizational Behavior & Human Decision Processes 51(2): 237-272.

Camerer, C. F. (1992b). Recent tests of generalized utility theories. Utility theories: Measurement and applications. W. Edwards, Cambridge University Press.

Camerer, C. (1995). Individual Decision Making. Handbook of Experimental Economics. J. Kagel and A. E. Roth, Princeton University Press: 587-703.

Camerer, C., L. Babcock, G. Loewenstein and R. Thaler (1995). A Target Income Theory of Labor Supply: Evidence from Cab Drivers, California Institute of Technology, Carnegie-Mellon University, University of Chicago.

Camerer, C. and M. Knez (1995). Coordination, Business Practices, and the Boundaries of Firms, California Institute of Technology and University of Chicago.

Camerer, C. and M. Knez (1995). First-Mover Advantage without Observability in Ultimatum Bargaining Games, Division of Social Sciences, California Institute of Technology; University of Chicago, Graduate School of Business.

Camerer, C., G. Loewenstein and M. Weber (1989). "The Curse of Knowledge in Economic Settings: An Experimental Analysis." Journal of Political Economy 97(5): 1232-1254.

Camerer, C. and R. Thaler (1995). "Anomalies: Ultimatums, Dictators, and Manners." Journal of Economic Perspectives 9(2): 209-219.

Camerer, C. and R. Thaler (1995). "Correspondence." Journal of Economic Perspectives 9(4): 239-240.

Camerer, C. F. and T.-H. Ho (1994). "Violations of the Betweenness Axiom and Nonlinearity in Probability." Journal of Risk and Uncertainty 8(2): 167-96.

Camerer, C. F. and R. Karjalainen (1994). Ambiguity and nonadditive probability in noncooperative game experiments. Models and Experiments on Risk and Rationality. M. Machina and B. Munier, Kluwer Academic Publishers.

Camerer, C. F. and H. C. Kunreuther (1989). "Experimental markets for insurance." Journal of Risk and Uncertainty 2: 265-300.

Camerer, C. F. and G. Loewenstein (1993). Information, fairness, and efficiency in bargaining. Psychological perspectives on justice: theory and applications. J. Baron and B. A. Mellers. New York, Cambridge University Press: 155-179.

Camerer, C. F. and K. R. MacCrimmon (1983). Underground and Overpaid: Equity Theory in Practice. Equity Theory: Psychological and Sociological Perspectives. D. M. Messick and K. S. Cook, Praeger: 295-326.

Cameron, L., (1995). "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." Working Paper #345: Industrial Relations Section, Princeton Univ., July 1995.

Caporael, L. R., R. M. Dawes, J. M. Orbell and A. J. C. van de Kragt (1989). "Selfishness examined: Cooperation in the absence of egoistic incentives." Behavioral and Brain Sciences 12: 683-739.

Carter, M. and M. Sunderland (1992). A Bargaining Experiment, Department of Economics, University of Canterbury.

Cason, T. and V.-L. Mui (1995). Fairness and Group Polarization in the Team Dictator Game, Department of Economics, University of Southern California.

Chapman, L. J. and J. P. Chapman (1967). "Genesis of popular but erroneous psychodiagnostic observations." Journal of Abnormal Psychology 72(3): 193-204.

Chapman, L. J. and J. P. Chapman (1969). "Illusory correlation as an obstacle to the use of valid psychodiagnostic signs." Journal of Abnormal Psychology 74(3): 271-280.

Chapman, L.J. and J. P. Chapman (1982) "Test results are what you think they are." Judgment under uncertainty: Heuristics and biases. Eds.: D. Kahneman, P. Slovic and A. Tversky. New York: Cambridge University Press: 239-248.

Charness, G. (1996a). "Pre-Screening and Dispute Costs: An Experimental Investigation." Mimeo, University of California at Berkeley, May.

Charness, G. (1996b). "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation." Mimeo, University of California at Berkeley, August.

Chew, S. H. and L. G. Epstein (1989). "A Unifying Approach to Axiomatic Non-expected Utility Theories." Journal of Economic Theory 49(2): 207-40.

Christensen-Szalanski, J. J. and C. F. Willham (1991). "The Hindsight Bias: A Meta-Analysis." Organizational Behavior & Human Decision Processes 48(1): 147-168.

Chung S-H. and R. J. Herrnstein (1967). "Choice and delay of reinforcement." Journal of the Experimental Analysis of Behavior 10: 67-74.

Cohen, D. and J. Knetsch (1990). Judicial Choice and Disparities Between Measures of Economic Values, Simon Fraser University Working Paper.

Collard, D. (1975). "Edgeworth's Propositions on Altruism." Economic Journal 85(338): 355-60.

Conlisk, J. (1989). "Three Variants on the Allais Example." American Economic Review 79(3): 392-407.

Conlisk, J. (1996). "Why Bounded Rationality?" Journal of Economic Literature, 34(June): 669-700ε.

Cook, K. S. and D. M. Messick (1983). Psychological and Sociological Perspectives on Distributive Justice: Convergent, Divergent, and Parallel Lines. Equity Theory: Psychological and Sociological Perspectives. D. M. Messick and K. S. Cook, Praeger: 1-12.

Cook, K. S. and T. Yamagishi (1983). Social Determinants of Equity Judgments: The Problem of Multidimensional Input. Equity Theory: Psychological and Sociological Perspectives. D. M. Messick and K. S. Cook, Praeger: 95-126.

Coombs, C. H. and L. C. Huang (1976). "Tests of the betweenness property of expected utility." Journal of Mathematical Psychology 13(3): 323-337.

Cotton, J. L. (1985). Cognitive dissonance in selective exposure. Selective Exposure to Communication. D. Zillmann and J. Bryant. Hillsdale, Lawrence Erlbaum Associates: 11-33.

Coursey, D. L., J. L. Hovis and W. D. Schulze (1987). "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value." The Quarterly Journal of Economics 102: 679-690.

Cowherd, D. M. and D. I. Levine (1992). "Product Quality and Pay Equity Between Lower-level Employees and Top Management: An Investigation of Distributive Justice Theory." Administrative Science Quarterly 37: 302-320.

Cox, J. C. and S. Epstein (1989). "Preference Reversals Without the Independence Axiom." American Economic Review 79(3): 408-426.

Croson, R. T. A. (1995). Expectations in Voluntary Contributions Mechanisms, Wharton School, Univ. of Penn.

Curley, S. P. and J. F. Yates (1985). "The Center and Range of the Probability Interval as Factors Affecting Ambiguity Preferences." Organizational Behavior & Human Decision Processes 36(2): 273-287.

Curley, S. P. and J. F. Yates (1989). "An empirical evaluation of descriptive models of ambiguity reactions in choice situations." Journal of Mathematical Psychology 33(4): 397-427.

Curley, S. P., J. F. Yates and R. A. Abrams (1986). "Psychological Sources of Ambiguity Avoidance." Organizational Behavior & Human Decision Processes 38(2): 230-256.

Davis, D. and T.C. Brock (1975). "Use of first person pronouns as a function of increased objective self-awareness and performance feedback. Journal of Experimental Psychology 11: 381-388.

Dawes, R.M., D. Faust and P. E. Meehl (1989). "Clinical versus actuarial judgment." Science 243: 1668-1674.

Dawes, R. M. and R. H. Thaler (1988). "Anomalies: cooperation." Journal of Economic Perspectives 2(3 (Summer 1988)): 187-197.

DeBondt, W. F. M. and R. H. Thaler (1985). "Does the Stock Market Overreact?" Journal of Finance 40(3): 793-808.

De Bondt, W. F. M. and R. H. Thaler (1987). "Further Evidence on Investor Overreaction and Stock Market Seasonality." Journal of Finance 42(3): 557-581.

De Bondt, W. F. M. and R. H. Thaler (1990). "Do Security Analysts Overreact?" American Economic Review 80(2): 52-57.

Dekel, E. (1986). "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom." Journal of Economic Theory 40(2): 304-18.

Dekel, E. (1989). "Asset Demands without the Independence Axiom." Econometrica 57(1): 163-69.

De-Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann (1989). "The Size and Incidence of the Losses from Noise Trading." Journal of Finance 44(3): 681-96.

De-Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann (1990a). "Noise Trader Risk in Financial Markets." Journal of Political Economy 98(4): 703-38.

De-Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann (1990b). "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." Journal of Finance 45(2): 379-95.

De-Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann (1991). "The Survival of Noise Traders in Financial Markets." Journal of Business 64(1): 1-19.

Devine, P. G., E. R. Hirt and E. M. Gehrke (1990). "Diagnostic and confirmation strategies in trait hypothesis testing." Journal of Personality & Social Psychology 58(6): 952-963.

Diamond, P. (1994). "Testing the Internal Consistency of Contingent Valuation Surveys." Journal of Environmental Economics and Management 30: 337-347.

Diamond, P. A. and J. A. Hausman (1994). "Contingent Valuation: Is Some Number Better than No Number?" Journal of Economic Perspectives 8(4): 45-64.

Dickens, W. T. (1986). "Crime and Punishment Again: The Economic Approach with a Psychological Twist." Journal of Public Economics 30: 97-107.

Dillard, J. F., N. L. Kauffman and E. E. Spires (1991). "Evidence Order and Belief Revision in Management Accounting Decisions." Accounting, Organizations & Society 16(7): 619-633.

Dougherty, T. W., D. B. Turban and J. C. Callender (1994). "Confirming first impressions in the employment interview: A field study of interviewer behavior." Journal of Applied Psychology 79(5): 659-665.

Dow, J. and S. R. d. Costa Werlang (1992). "Excess Volatility of Stock Prices and Knightian Uncertainty." European Economic Review 36(2,3): 631-638.

Dow, J. and S. R. d. C. Werlang (1992). "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio." Econometrica 60(1): 197-204.

Dow, J. and S. R. d. C. Werlang (1994). "Nash equilibrium under Knightian uncertainty: Breaking down backward induction." Journal of Economic Theory 65(2): 305-324.

Duesenberry, J. S. (1952). Income, Saving and the Theory of Consumer Behavior, Harvard University Press.

Duval, S. and R.A. Wicklund. A Theory of Objective Self-Awareness, Academic Press.

Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. Nations and Households in Economic Growth. P. A. David and M. W. Reder, Academic Press: 89-125.

Einhorn, H. and R. Hogarth (1978). "Confidence in Judgment: Persistence of the Illusion of Validity." Psychological Review 85(5): 395-416.

Einhorn, H. J. and R. M. Hogarth (1987). Decision Making Under Ambiguity. Rational Choice: The Contrast Between Economics and Psychology. R. M. Hogarth and M. W. Reder, University of Chicago Press: 41-66.

Ellison, G. (1993). "Learning, Local Interaction, and Coordination." Econometrica 61: 1047-1071.

Ellsberg, D. (1961). "Risk, ambiguity and the Savage axioms." Quarterly Journal of Economics 75: 643-679.

Elster, J. (1977). "Ulysses and the sirens: A theory of imperfect rationality." Social Science Information 16(5): 469-526.

Elster, J. (1985). "Sadder but wiser? Rationality and the emotions." Social Science Information 24(2): 375-406.

Elster, J. (1991). Envy in Economic Life. Strategy and Choice. Zeckhauser, MIT Press.

Elster, J. (1992). Intertemporal Choice and Political Thought. Choice over Time. G. Loewenstein and J. Elster, Russell Sage Foundation: 35-56.

Epstein, L. G. and M. Le Breton (1993). "Dynamically consistent beliefs must be Bayesian." Journal of Economic Theory 61(1): 1-22.

Elster, J. and G. Loewenstein (1992). Utility From Memory and Anticipation. Choice Over Time. J. Elster and G. Loewenstein. 213-234, Ch. 9.

Epstein, L. G. and T. Wang (1994). "Intertemporal asset pricing under Knightian uncertainty." Econometrica 62(2): 282-322.

Ely, J. (1995). "Local Conventions." Mimeo, University of California at Berkeley, October.

Farber, H. S. (1981). "Splitting-the-Difference in Interest Arbitration." Industrial and Labor Relations Review 35(1): 70-77.

Fehr, E. and A. Falk (1996). "Wage Rigidities in a Competitive Incomplete Contract Market: An Experimental Investigation." Institute for Empirical Economic Research, University of Zurich.

Fehr, E. S. Gaechter, G. Kirchsteiger (1996). "Reciprocity as a Contract Enforcement Device: Experimental Evidence." Mimeo (May).

Fehr, E., E. Kirchler, and A. Weichbold (1994). "When Social Forces Remove the Impact of Competition: An Experimental Investigation." Mimeo, Department of Economics, University of Technology, Vienna; Institute of Psychology, Department of Applied Psychology, University of Vienna; Department of Economics, Institute for Advanced Studies, Vienna, (July).

Fehr, E., G. Kirchsteiger and A. Riedl (1993). "Does fairness prevent market clearing? An experimental investigation." Quarterly Journal of Economics 108(2): 437-459.

Fehr, E., G. Kirchsteiger, and A. Riedl (1996). "Involuntary Unemployment and Non-Compensating Wage Differentials in an Experimental Labour Market." The Economic Journal 106(434): 106-121.

Fehr, E. and E. Tougareva (1996). "Do High Monetary Stakes Remove Reciprocal Fairness? Experimental Evidence From Russia." Institute for Empirical Research in Economics, Zurich; Academy of Sciences, Institute of Psychology, Moscow.

Fehr, E. and P. Zych (1994). "The Power of Temptation: Irrationally Myopic Excess Consumption in an Addiction Experiment." Mimeo (September).

Fellner, W. (1961). "Distortion of subjective probabilities as a reaction to uncertainty." Quarterly Journal of Economics 75: 670-694.

Fershtman, C. (1993). On the Value of Incumbency: Managerial Reference Point and Loss Aversion. Discussion Paper, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Finn, R. H. and S. M. Lee (1972). "Salary equity: Its determination, analysis, and correlates." Journal of Applied Psychology 56(4): 283-292.

Fischhoff, B. (1975). "Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty." Journal of Experimental Psychology: Human Perception & Performance 104(1): 288-299.

Fischhoff, B. and R. Beyth-Marom (1983). "Hypothesis Evaluation from a Bayesian Perspective." Psychological Review 90(3): 239-260.

Fischhoff, B., P. Slovic and S. Lichtenstein (1977). "Knowing with certainty: The appropriateness of extreme confidence." Journal of Experimental Psychology: Human Perception & Performance 3(4): 552-564.

Fischburn, P. C. (1988). Non-linear preference and utility theory, Johns Hopkins Press.

Fischburn, P. C. (1989). "Generalizations of expected utility theories: A survey of recent proposals." Annals of Operations Research 19: 3-28.

Fischburn, P. C. and G. A. Kochenberger (1979). "Two-Piece Von Neumann-Morgenstern Utility Functions." Decision Sciences 10(4): 503-518.

Fiske, S. T. and S. E. Taylor (1984). Social Cognition. Reading, MA, Addison-Wesley.

Fleming, J. and A. J. Arrowood (1979). "Information processing and the perseverance of discredited self-perceptions." Personality & Social Psychology Bulletin 5(2): 201-205.

Flinn, J. (1995). Steamed at Starbucks. San Francisco Examiner. San Francisco: B1, B3.

Foley, D. (1967). "Resource Allocation and the Public Sector." Yale Economic Essays 7: 45-98.

Fox, C. R., B. A. Rogers and A. Tversky (1995). Options Traders Exhibit Subadditive Decision Weights, Stanford University.

Fox, C. R. and A. Tversky (1995). "Ambiguity aversion and comparative ignorance." Quarterly Journal of Economics 110(3): 585-603.

Franciosi, R., P. Kujal, R. Michelitsch, V. Smith and G. Deng (1995). "Fairness: effect on temporary and equilibrium prices in posted-offer markets." Economic Journal 105: 938-950.

Frank, R. H. (1989). "Frames of Reference and the Quality of Life." American Economic Review 79(May): 80-85.

Frank, R. H. (1992). Frames of reference and the intertemporal wage profile. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 371-382.

Frank, R. H. (1994). Local status, fairness, and wage compression revisited, Department of Economics, Cornell University.

Frank, R. H. and R. M. Hutchens (1993). "Wages, seniority, and the demand for rising consumption profiles." Journal of Economic Behavior & Organization 21(3): 251-276.

Fredrickson, B. L. and D. Kahneman (1993). "Duration neglect in retrospective evaluations of affective episodes." Journal of Personality & Social Psychology 65(1): 45-55.

Friedland, N. (1990). "Attribution of control as a determinant of cooperation in exchange interactions." Journal of Applied Social Psychology 20(4, Pt 1): 303-320.

Friedrich, J. (1993). "Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena." Psychological Review 100(2): 298-319.

Frisch, D. and J. Baron (1988). "Ambiguity and rationality." Journal of Behavioral Decision Making 1(3): 149-157.

Frisch, D. M. and M. S. Greenberg (1968). "Reciprocity and intentionality in the giving of help." Proceedings of the 76th Annual Convention of the American Psychological Association 3: 383-384.

Fudenberg, D. and D. K. Levine (1993). "Self-Confirming Equilibrium." Econometrica 61(3): 523-545.

Funder, D. C. and J. Block (1989). "The role of ego-control, ego-resiliency, and IQ in delay of gratification in adolescence." Journal of Personality & Social Psychology 57(6): 1041-1050.

Gardenfors, P. and N.-E. Sahlin (1988). Unreliable probabilities, risk taking, and decision making. Cambridge University Press, New York, NY, US.

Geanakoplos, Pearce and Stacchetti (1989). "Psychological Games." Games and Economic Behavior 1(1): 60-79.

Gigerenzer, G., U. Hoffrage and H. Kleinbolting (1991). "Probabilistic mental models: A Brunswikian theory of confidence." Psychological Review 98(4): 506-528.

Gilboa, I. (1987). "Expected Utility with Purely Subjective Non-Additive Probabilities." Journal of Mathematical Economics 16(1): 65-88.

Gilboa, I. and D. Schmeidler (1989). "Maxmin Expected Utility with Non-Unique Prior." Journal of Mathematical Economics 18(2): 141-153.

Gilovich, T., R. Vallone and A. Tversky (1985). "The Hot Hand in Basketball: On the Misperception of Random Sequences." Cognitive Psychology 17: 295-314.

Glazer, J. and A. Weiss (1992). Conflicting Preferences and Voluntary Restrictions on Choices, Tel Aviv University, Boston University.

Goetze, D. and J. M. Orbell (1988). "Understanding and Cooperation in Social Dilemmas." Public Choice 57(June 1988): 275-279.

Goldman, S. M. (1979). "Intertemporally Inconsistent Preferences and the Rate of Consumption." Econometrica 47(3): 621-626.

Goldman, S. M. (1980). "Consistent Plans." Review of Economic Studies 47: 533-537.

Goranson, R. E. and L. Berkowitz (1966). "Reciprocity and Responsibility Reactions to Prior Help." Journal of Personality and Social Psychology 3: 227-232.

Greenberg, J. (1982). Approaching Equity and Avoiding Inequity in Groups and Organizations. Equity and Justice in Social Behavior. J. Greenberg and R. L. Cohen, Academic Press: 389-435.

Greenberg, J. and G. S. Leventhal (1976). "Equity and the use of overreward to motivate performance." Journal of Personality & Social Psychology 34(2): 179-190.

Greenberg, M. S., M. W. Block and M. A. Silverman (1971). "Determinants of helping behavior: Person's rewards versus other's costs." Journal of Personality 39(1): 79-93.

Greenberg, M. S. and D. M. Frisch (1972). "Effect of Intentionality on Willingness to Reciprocate." Journal of Experimental Social Psychology 8: 99-111.

Greenglass, E. R. (1969). "Effects of prior help and hindrance on willingness to help another: Reciprocity or social responsibility." Journal of Personality & Social Psychology 11(3): 224-231.

Grether, D. M. (1980). "Bayes' rule as a descriptive model: The representativeness heuristic." Quarterly Journal of Economics 95: 537-57.

Grether, D. M. (1981). Financial incentive effects and individual decision making, California Institute of Technology.

Grether, D. M. (1992). "Testing Bayes rule and the representativeness heuristic: Some experimental evidence." Journal of Behavior and Organization 17: 31-57.

Griffin, D. and A. Tversky (1992). "The weighing of evidence and the determinants of confidence." Cognitive Psychology 24(3): 411-435.

Gul, F. (1991). "A Theory of Disappointment Aversion." Econometrica 59(3(May 1991)): 667-686.

Gur, R. C. and H. A. Sackeim (1979). "Self-Deception: A Concept in Search of a Phenomenon." Journal of Personality and Social Psychology 37(2): 147-169.

Guth, W., P. Ockenfels and M. Wendel (1991). Efficiency by Trust in Fairness? - Multiperiod Ultimatum Bargaining Experiments With an Increasing Cake, J. W. Goethe Universität Frankfurt/ Main.

Guth, W., R. Schmittberger and B. Schwarze (1982). "An Experimental Analysis of Ultimatum Bargaining." Journal of Economic Behavior and Organization 3: 367-388.

Guth, W. and R. Tietz (1990). "Ultimatum Bargaining Behavior: A survey and comparison of experimental results." Journal of Economic Psychology 11(3): 417-449.

Haltiwanger, J. and M. Waldman (1985). "Rational expectations and the limits of rationality: an analysis of heterogeneity." American Economic Review 75(3): 326-340.

Hamilton, D. L. and T. L. Rose (1980). "Illusory correlation and the maintenance of stereotypic beliefs." Journal of Personality & Social Psychology 39(5): 832-845.

Hamilton, D. L., S. J. Sherman and C. M. Ruvolo (1990). "Stereotype-based expectancies: Effects on information processing and social behavior." Journal of Social Issues 46(2): 35-60.

Hammond, P. J. (1976). "Changing Tastes and Coherent Dynamic Choice." Review of Economic Studies 43: 159-173.

Hammond, P. J. (1976). "Endogenous Tastes and Stable Long-Run Choice." Journal of Economic Theory 13: 329-340.

Harless, D. W. and C. F. Camerer (1994). "The predictive utility of generalized expected utility theories." Econometrica 62(6): 1251-1289.

Harris, R. J. (1983). Pinning Down the Equity Formula. Equity Theory: Psychological and Sociological Perspectives. D. M. Messick and K. S. Cook, Praeger: 207-242.

Hartman, R. S., M. J. Doane and C.-K. Woo (1991). "Consumer Rationality and the Status Quo." Quarterly Journal of Economics 106(1): 141-162.

Haugtvedt, C. P. and D. T. Wegener (1994). "Message order effects in persuasion: An attitude strength perspective." Journal of Consumer Research 21(1): 205-218.

Hausman, J. A. (1993). Contingent Valuations: A Critical Assessment. Amsterdam: North Holland Press.

Haverkamp, B. E. (1993). "Confirmatory bias in hypothesis testing for client-identified and counselor self-generated hypotheses." Journal of Counseling Psychology 40(3): 303-315.

Hawkins, S. A. and R. Hastie (1990). "Hindsight: Biased judgments of past events after the outcomes are known." Psychological Bulletin 107(3): 311-327.

Heath, C. and A. Tversky (1990). Preference and Belief: Ambiguity and Competence in Choice Under Uncertainty. Contemporary Issues in Decision Making. K. Borcherding, O. I. Larichev and D. M. Messick. Amsterdam, North-Holland: 93-123.

Helson, H. (1964). Adaptation level theory: An experimental and systematic approach to behavior, Harper & Row.

Herrnstein, R. J. and D. Prelec (1991). "Melioration: A Theory of Distributed Choice." Journal of Economic Perspectives 5(Summer): 137-156.

Herrnstein, R. J. and D. Prelec (1992a). Melioration. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 235-263. Ch. 10.

Herrnstein, R. J. and D. Prelec (1992b). A theory of addiction. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 331-357. Ch. 13.

Hey, J. D. and C. Orme (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data." Econometrica 62(6): 1291-1326.

Heyman, G. M. (1994). Addiction, Choice or Disease? An Empirical Review and Theoretical Synthesis, Harvard University.

Hoch, S. J. and G. Loewenstein (1991). "Time-inconsistent Preferences and Consumer Self-Control." Journal of Consumer Research 17: 492-507.

Hodges, J. L. and E. L. Lehman (1952). "The use of previous experience in reaching statistical decisions." Annals of Mathematical Statistics 23: 396-407.

Hodgins, H. S. and M. Zuckerman (1993). "Beyond selecting information: Biases in spontaneous questions and resultant conclusions." Journal of Experimental Social Psychology 29(5): 387-407.

Hoffman, E., K. McCabe, K. Shachat and V. Smith (1994). "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior 7(3): 346-380.

Hogarth, R. M. and H. J. Einhorn (1990). "Venture Theory: A Model of Decision Weights." Management Science 36(7): 780-803.

Hogarth, R. M. and H. J. Einhorn (1992). "Order effects in belief updating: The belief-adjustment model." Cognitive Psychology 24(1): 1-55.

Hogarth, R. M. and H. Kunreuther (1985). "Ambiguity and Insurance Decisions." American Economic Review 75(2): 386-390.

Hogarth, R. M. and H. Kunreuther (1989). "Risk, ambiguity and insurance." Journal of Risk and Uncertainty 2: 5-35.

Hogarth, R. M. and H. Kunreuther (1990). Pricing insurance and warranties: Ambiguity and correlated risks, University of Chicago, Center of Decision Research.

Hollinger, R. C., K. B. Slora and W. Terris (1992). "Deviance in the fast-food restaurant: Correlates of employee theft, altruism, and counterproductivity." Deviant Behavior 13(2): 155-184.

Huang, P. H. (1993). Integrating Feeling and Thinking in Economic Actors, Economics Department, Tulane University.

Huang, P. H. and H.-m. Wu (1992). "Emotional Responses in Litigation." International Review of Law and Economics 12: 31-44.

Hubbard, M. (1984). "Impression perseverance: Support for a cognitive explanation." Representative Research in Social Psychology 14(1): 48-55.

Ibrahim, M. and M. Shehata (1993). "Information sequence, response mode and asset valuation judgment." Managerial Finance 19(5): 11-27.

Isaac, R. M. and J. M. Walker (1988a). "Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism." Economic Inquiry 26(2): 585-608.

Isaac, R. M. and J. M. Walker (1988b). "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." The Quarterly Journal of Economics 103(1): 179-199.

Issac, R. M., D. Mathieu and E. E. Zajac (1991). "Institutional framing and perceptions of fairness." Constitutional Political Economy 2(3): 329-370.

Issac, R. M., J. McCue and C. Plott (1985). "Public goods provision in an experimental environment." Journal of Political Economy 26: 51-74.

Jennings, D. L., T. M. Amabile and L. Ross (1982). Informal Covariation Assessment: Data-Based versus Theory-Based Judgments. Judgment under Uncertainty: Heuristics and Biases. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press: 211-230, Chap. 15.

Jennings, D. L., M. R. Lepper and L. Ross (1981). "Persistence of impressions of personal persuasiveness: Perseverance of erroneous self-assessments outside the debriefing paradigm." Personality & Social Psychology Bulletin 7(2): 257-263.

Kachelmeier, S. J., S. T. Limberg and M. S. Schadewald (1991). "Fairness in markets: A laboratory investigation." Journal of Economic Psychology 12(3): 447-464.

Kagel, J. H., C. Kim and D. Moser (1992). "Fairness" in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs, Department of Economics, University of Pittsburgh; Joseph M. Katz Graduate School of Business, University of Pittsburgh.

Kagel, J. H., D. N. MacDonald and R. C. Battalio (1990). "Tests of "Fanning Out" of Indifference Curves: Results from Animal and Human Experiments." American Economic Review 80(4): 912-921.

Kagel, J. and A. Roth, Eds. (1995). Handbook of Experimental Economics. Princeton, Princeton University Press.

Kahn, A. and T. Tice (1973). "Returning a Favor and Retaliating Harm: The Effects of Stated Intentions and Actual Behavior." Journal of Experimental Social Psychology 9: 43-56.

Kahneman, D. (1991). "Judgment and decision making: A personal view." Psychological Science 2(3): 142-145.

Kahneman, D. (1992). "Reference points, anchors, norms, and mixed feelings. Special Issue: Decision processes in negotiation." Organizational Behavior & Human Decision Processes 51(2): 296-312.

Kahneman, D. (1994). "New Challenges to the Rationality Assumption." Journal of Institutional and Theoretical Economics 150(1): 18-36.

Kahneman, D., B. L. Fredrickson, C. A. Schreiber and D. A. Redelmeier (1993). "When more pain is preferred to less: Adding a better end." Psychological Science 4(6): 401-405.

Kahneman, D. and J. L Knetsch (1992). "Valuing Public Goods: The Purchase of Moral Satisfaction." Journal of Environmental Economics and Management 22: 57-70.

Kahneman, D., J. Knetsch and R. Thaler (1991). "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." Journal of Economic Perspectives 5(Winter): 193-206.

Kahneman, D., J. Knetsch and R. Thaler (1986a). "Fairness and the Assumptions of Economics." Journal of Business 59: S285-S300.

Kahneman, D., J. Knetsch and R. Thaler (1986b). "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." American Economic Review 76: 728-741.

Kahneman, D., J. L. Knetsch and R. Thaler (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." Journal of Political Economy 98: 1325-1348.

Kahneman, D., J. L. Knetsch, R. H. Thaler, H. Kunreuther, R. D. Luce and R. A. Shweder (1986). "Fairness and the Assumptions of Economics/Comments." Journal of Business 59(4): S285-S300,S329-S354.

Kahneman, D. and D. Lovallo (1993). "Timid choices and bold forecasts: A cognitive perspective on risk taking." Management Science 39(1): 17-31.

Kahneman, D. and I. Ritov (1994). "Determinants of Stated Willingness to Pay for Public Goods -- A Study in the Headline Method." Journal of Risk and Uncertainty 9(1): 5-38.

Kahneman, D., I. Ritov, K. E. Jacowitz and P. Grant (1993). "Stated willingness to pay for public goods: A psychological perspective." Psychological Science 4(5): 310-315.

Kahneman, D., P. Slovic and A. Tversky, Eds. (1982). Judgment under uncertainty: Heuristic and biases, Cambridge University Press.

Kahneman, D. and R. Thaler (1991). "Economic Analysis and the Psychology of Utility: Applications to Compensation Policy." American Economic Review (Proceedings) 81: 341-352.

Kahneman, D. and A. Tversky (1973). "On the Psychology of Prediction." Psychological Review 80: 237-251.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk." Econometrica 47(March): 263-291.

Kahneman, D. and A. Tversky (1982a). "Subjective probability: A judgment of representativeness." Judgment under uncertainty: Heuristics and biases. D. Kahneman, P. Slovic, and A. Tversky. New York: Cambridge University Press: 32-47, Ch. 3.

Kahneman, D. and A. Tversky (1982b). "Variants of uncertainty." Cognition 11(2): 143-157.

Kahneman, D. and A. Tversky (1983). "On the study of statistical intuitions." Judgment under uncertainty: Heuristics and biases. Eds.: D. Kahneman, P. Slovic and A. Tversky. New York: Cambridge University Press: 493-508.

Kahneman, D. and A. Tversky (1983). "Can irrationality be intelligently discussed?" Behavioral & Brain Sciences 6(3): 509-510.

Kahneman, D. and A. Tversky (1984). "Choices, values, and frames." American Psychologist 39(4): 341-350.

Kahneman, D. and C. Varey (1991). Notes on the psychology of utility. Cambridge University Press, Cambridge, MA, US.

Kalai and Lehrer (1993). "Rational Learning Leads to Nash Equilibria." Econometrica 61(5): 1019-1045.

Kandori, M., G. Mailath and R. Rob (1993). "Learning, Mutation, and Long Run Equilibria in Games." Econometrica 61: 29-56.

Kaplan, T. R. and K. A. McCabe (1992). Experiments on Market Power in Three-Player Ultimatum Games, Economics Department, University of Minnesota; Carlson School of Management, University of Minnesota.

Kelly, H. H. and J. W. Thibaut (1978). Interpersonal relations: A theory of interdependence, John Wiley & Sons.

Keren, G. (1987). "Facing Uncertainty in the Game of Bridge: A Calibration Study." Organizational Behavior & Human Decision Processes 39(1): 98-114.

Keren, G. (1988). "On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies." Acta Psychologica 67(2): 95-119.

Keser, C. (1996). "Voluntary Contributions to a Public Good When Partial Contribution is a Dominant Strategy." Economics Letters 50(3): 359-366.

Keynes, J. M. (1924). Monetary Reform, Harcourt, Brace and Co.

Kirby, K. and R. J. Herrnstein (1995). "Perference Reversals Due to Myopic Discounting of Delayed Reward." Psychological Science 6(2): 83-89.

Kirchsteiger, G. (1992). The Role of Envy in Ultimatum Games, Institute for Advanced Studies, Vienna.

Klayman, J. and Y.-w. Ha (1987). "Confirmation, disconfirmation, and information in hypothesis testing." Psychological Review 94(2): 211-228.

Klein, W. M. and Z. Kunda (1993). "Maintaining self-serving social comparisons: Biased reconstruction of one's past behaviors." Personality & Social Psychology Bulletin 19(6): 732-739.

Klein, W. M. and Z. Kunda (1994). "Exaggerated self-assessments and the preference for controllable risks." Organizational Behavior & Human Decision Processes 59(3): 410-427.

Klibanoff, P. (1993). Dynamic Choice with Uncertainty Aversion, Massachusetts Institute of Technology.

Klibanoff, P. (1993). Uncertainty Aversion in Game Theory, Massachusetts Institute of Technology.

Knetsch, J. L. (1989). "The Endowment Effect and Evidence of Nonreversible Indifference Curves." American Economic Review 79: 1277-1284.

Knetsch, J. L. (1992). "Preferences and Nonreversibility of Indifference Curves." Journal of Economic Behavior & Organization 17(1): 131-139.

Knetsch, J. L. and J. A. Sinden (1984). "Willingness to Pay and Compensation demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." Quarterly Journal of Economics 99(August): 507-521.

Kohlberg, L. (1963). "The Development of Children's Orientations Toward a Moral Order: Pt. 1. Sequence in the Development of Moral Thought." Vita Human 6: 11-33.

Kohlberg, L. (1983). The Philosophy of Moral Development. New York, Harper and Row.

Konow, J. (1993). Toward a General Theory of Economic Fairness, Department of Economics, Loyola Marymount University.

Koopmans, T. (1960). "Stationary ordinal Utility and Impatience." Econometrica 28: 287-309.

Kramer, R. M., E. Newton and P. L. Pommerenke (1993). "Self-enhancement biases and negotiator judgment: Effects of self-esteem and mood." Organizational Behavior & Human Decision Processes 56(1): 110-133.

Krebs, D. L. (1970). "Altruism--an examination of the concept and a review of the literature." Psychological Bulletin 73(4): 258-302.

Krebs, D. L. (1982). "Prosocial Behavior, Equity, and Justice." Equity and Justice in Social Behavior. Eds.: J. Greenberg and R. Cohen, New York: Academic Press.

Kuhn, T. S. (1970). The Structure of Scientific Revolutions. Second edition. Chicago: University of Chicago Press.

Kunda, Z. (1987). "Motivated inference: self-serving generation and evaluation of causal theories." Journal of Personality and Social Psychology 53(4): 636-647.

Kunda, Z. (1990). "The case for motivated reasoning." Psychological Bulletin 108(3): 480-498.

Kunda, Z. (1992). "Can dissonance theory do it all?" Psychological Inquiry 3(4): 337-339.

Kunda, Z. and B. Sherman-Williams (1993). "Stereotypes and the construal of individuating information." Personality & Social Psychology Bulletin 19(1): 90-99.

Laibson, D. (1994). Essays in Hyperbolic Discounting. Economics, MIT.

Laibson, D. (1995). Hyperbolic Discount Functions, Undersaving, and Savings Policy, Harvard University.

Lazear, E. P. (1989). "Pay Equity and Industrial Politics." Journal of Political Economy 97(31): 561-580.

Ledyard, J. O. (1995). "Public Goods: A Survey of Experimental Research." Handbook of Experimental Economics. Eds.: J. Kagel and A. E. Roth. Princeton: Princeton University Press, pp. 111-194.

Lepper, M. R., L. Ross and R. R. Lau (1986). "Persistence of inaccurate beliefs about the self: Perseverance effects in the classroom." Journal of Personality & Social Psychology 50(3): 482-491.

Lerner, M. J. and R. R. Lichtman (1968). "Effects of Perceived Norms on Attitudes and Altruistic Behavior Toward A Dependent Other." Journal of Personality & Social Psychology 9(3): 226-232.

Lerner, M. J. and G. Matthews (1967). "Reactions to suffering of others under conditions of indirect responsibility." Journal of Personality and Social Psychology 5(3): 319-325.

Lerner, M. J. and D. T. Miller (1978). "Just world research and the attribution process: looking back and ahead." Psychological Bulletin 85(5): 1030-1051.

Leventhal, G. S. and D. Anderson (1970). "Self-interest and the maintenance of equity." Journal of Personality & Social Psychology 15(1): 57-62.

Leventhal, G. S. and J. T. Bergman (1969). "Self-depriving behavior as a response to unprofitable inequity." Journal of Experimental Social Psychology 5(2): 153-171.

Leventhal, G. S., J. W. Michaels and C. Sanford (1972). "Inequity and Interpersonal Conflict: Reward Allocation and Secrecy about Reward As Methods of Preventing Conflict." Journal of Personality and Social Psychology 23(1): 88-102.

Leventhal, G. S., T. Weiss and G. Long (1969). "Equity, reciprocity, and reallocating rewards in the dyad." Journal of Personality & Social Psychology 13(4): 300-305.

Levine, D. I. (1991). "Cohesiveness, productivity, and wage dispersion." Journal of Economic Behavior and Organization 15: 237-255.

Levine, D. I. (1993a). "Fairness, Markets, and Ability to Pay: Evidence from Compensation Executives." American Economic Review 83(5): 1241-1259.

Levine, D. I. (1993b). "What Do Wages Buy?" Administrative Science Quarterly 38(3): 462-483.

Lichtenstein, S. and P. Slovic (1971). "Reversals of preference between bids and choices in gambling decisions." Journal of Experimental Psychology 89(1): 46-55.

Liebrand, W. B. G., D. M. Messick and F. J. M. Wolters (1986). "Why we are fairer than others: a cross-cultural replication and extension." Journal of Experimental Social Psychology 22: 590-604.

Loewenstein, G. (1992). The fall and rise of psychological explanations in the economics of intertemporal choice. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 3-34.

Loewenstein, G. (1996). "Out of Control: Visceral Influences on Behavior." Organizational Behavior and Human Decision Processes 65(2).

Loewenstein, G. and D. Adler (1995). "A bias in the prediction of tastes." Economic Journal 105: 929-937.

Loewenstein, G. and J. Elster, Eds. (1992). Choice over time, Russell sage foundation.

Loewenstein, G., S. Issacharoff, C. Camerer and L. Babcock (1993). "Self-Serving Assessments of Fairness and Pretrial Bargaining." Journal of Legal Studies 22(January): 135-159.

Loewenstein, G. and D. Prelec (1992). "Anomalies in Intertemporal Choice: Evidence and an Interpretation." Quarterly Journal of Economics 107(2): 573-597.

Loewenstein, G. and N. Sicherman (1991). "Do Workers Prefer Increasing Wage Profiles?" Journal of Labor Economics 9(1): 67-84.

Loewenstein, G. F., L. Thompson and M. H. Bazerman (1989). "Social utility and decision making in interpersonal contexts." Journal of Personality & Social Psychology 57(3): 426-441.

Loomes, G. and R. Sugden (1982). "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty." Economic Journal 92(368): 805-824.

Loomes, G. and R. Sugden (1987). "Some Implications of More General Form of Regret Theory." Journal of Economic Theory 41(2): 270-87.

Loomes, G. and R. Sugden (1987). "Testing for Regret and Disappointment in Choice Under Uncertainty." Economic Journal 97(Co): 118-129.

Lopes. L. L. (1982). "Towards a procedural theory of judgment." Technical Report No. 17. Madison: University of Wisconsin, Wisconsin Human Information Processing Program.

Lord, C. G., L. Ross and M. R. Lepper (1979). "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." Journal of Personality & Social Psychology 37(11): 2098-2109.

Lord, R. G. and J. A. Hohenfeld (1979). "Longitudinal Field Assessment of Equity Effects on the Performance of Major League Baseball Players." Journal of Applied Psychology 64(1): 19-26.

Luce, R. D. and L. Narens (1985). "Classification of concatenation measurement structures according to scale type." Journal of Mathematical Psychology 29(1): 1-72.

Luchins, A. S. and E. H. Luchins (1986). "Primacy and recency effects with descriptions of moral and immoral behavior." Journal of General Psychology 113(2): 159-177.

Lynn, M. and A. Grassman (1990). "Restaurant Tipping: An Examination of Three Rational Explanations." Journal of Economic Psychology 11(1990): 169-181.

Macan, T. H. and R. L. Dipboye (1994). "The effects of the application on processing of information from the employment interview." Journal of Applied Social Psychology 24(14): 1291-1314.

MacCrimmon, K. R. and S. Larsson (1979). Utility theory: Axioms versus paradoxes. The expected utility hypothesis and the Allais Paradox. M. Allais and O. Hagen. Dordrecht, Holland, D. Riedel: 333-409.

Machina, M. J. (1982). ""Expected Utility" Analysis without the Independence Axiom." Econometrica 50(March): 227-324.

Machina, M. (1987). "Choice under uncertainty: Problems solved and unsolved." Journal of Economic Perspectives(1): 121-54.

Machina, M. J. (1994). Non-Expected Utility and the Robustness of the Classical Insurance Paradigm, Department of Economics, University of California, San Diego.

Mackie, D. M. and D. L. Hamilton, Eds. (1993). Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception. San Diego, CA, Academic Press, Inc; Harcourt Brace Jovanovich Publishers.

Mahajan, J. (1992). "The Overconfidence Effect in Marketing Management Predictions." Journal of Marketing Research 29(3): 329-342.

Mahoney, M. (1977). "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." Cognitive Therapy and Research 1(2): 161-175.

Mannix, E. A. and G. F. Loewenstein (1994). "The effects of interfirm mobility and individual versus group decision making on managerial time horizons." Organizational Behavior & Human Decision Processes 59(3): 371-390.

Margolis, H. (1982). Selfishness, Altruism, and Rationality: A Theory of Social Choice. Chicago, University of Chicago Press.

Markowitz, H. (1952). "The Utility of Wealth." Journal of Political Economy 60: 151-58.

Maxwell, G. and R. E. Ames (1981). "Economists Free Ride, Does Anyone Else?" Journal of Public Economics 15: 295-310.

Mazur, J. E. (1987). "An adjustment procedure for studying delayed reinforcement." M. L. Commons, J. E. Mazur, J. A. Nevins and H. Rachlin. Quantitative Analysis of Behavior: The Effect of Delay and of Intervening Events on Reinforcement Value. Hillsdale NJ: Erlbaum.

McFadden, D. (1994). "Contingent Valuation and Social Choice." American Journal of Agricultural Economics 76(4): 689-708.

McGlothlin, W. H. (1956). "Stability of Choices among Uncertain Alternatives." American Journal of Psychology 69: 604-615.

McNeil, B. J., S. G. Pauker, H. C. Sox, Jr. and A. Tversky (1982). "On the Elicitation of Preferences For Alternative Therapies." New England Journal of Medicine 306(May 27, 1982): 1259-1262.

Mehle, T., C. F. Gettys, C. Manning, S. Baca and S. Fisher (1981). "The availability explanation of excessive plausibility assessments." Acta Psychologica 49(2): 127-140.

Meidinger, C. (1996). Equity, Fairness Equilibria and Coordination in the Ultimatum Game. Paris, Centre National de la Recherche Scientifique: Laboratoire de Microéconomie Appliquée.

Mellers, B. (1985). "A reconsideration of two-person inequity judgments: A reply to Anderson." Journal of Experimental Psychology: General 114(4): 514-520.

Mellers, B. and E. Hartka (1988). "Fair selection decisions." Journal of Experimental Psychology: Human Perception & Performance 14(4): 572-581.

Messick, D. M., S. Bloom, J. P. Boldizar and C. D. Samuelson (1985). "Why we are fairer than others." Journal of Experimental Social Psychology 21: 480-500.

Messick, D. M. and K. Sentis (1983). Fairness, Preference, and Fairness Biases. Equity Theory: Psychological and Sociological Perspectives. D. M. Messick and K. S. Cook. New York, Praeger: 61-94.

Miller, A. G., J. W. McHoskey, C. M. Bane and T. G. Dowd (1993). "The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change." Journal of Personality & Social Psychology 64(4): 561-574.

Miller, D. T. (1976). "Ego involvement and attributions for success and failure." Journal of Personality & Social Psychology 34(5): 901-906.

Miller, D. T. and M. Ross (1975). "Self-Serving Bias in the Attribution of Causality: Fact or Fiction?" Psychological Bulletin 82(2): 213-225.

Miller, G. (1992). Ethics and the new game theory. Ethics and agency theory. N. E. Bowie and R. E. Freeman. New York, Oxford University Press: 117-126.

Mitchell, G., P. E. Tetlock, B. A. Mellers and L. D. Ordonez (1993). "Judgments of social justice: Compromises between equality and efficiency." Journal of Personality & Social Psychology 65(4): 629-639.

Mitzkewitz, M. and R. Nagel (1991). Envy, Greed, and Anticipation in Ultimatum Games with Incomplete Information: An Experimental Study, University of Bonn, London School of Economics.

Montgomery, J. D. (1994). "Revisiting Tally's Corner: Mainstream Norms, Cognitive Dissonance, and Underclass Behavior." Rationality and Society 6(4): 462-488.

Morgan, R. B. (1993). "Self- and co-worker perceptions of ethics and their relationships to leadership and salary." Academy of Management Journal 36(1): 200-214.

Mui, V.-L. (1995). "The Economics of Envy." Journal of Economic Behavior and Organization 26: 311-336.

Nemeth, C. (1970). "Bargaining and reciprocity." Psychological Bulletin 74(5): 297-308.

Nisbett, R. and L. Ross (1980). Human Inference: Strategies and Shortcomings of Social Judgment, Prentice Hall.

Nisbett, R. E. and T. D. Wilson (1977). "Telling more than we can know: Verbal reports on mental processes." Psychological Review 84(3): 231-259.

Ochs, J. and A. E. Roth (1989). "An Experimental Study of Sequential Bargaining." American Economic Review 79(3 (June, 1989)): 355-384.

O"Donoghue, T. and M. Rabin (1996). "Doing It Now or Later." Mimeo, University of California at Berkeley, May.

Orford, J. (1985). Excessive Appetites: A Psychological View of Addictions, John Wiley & Sons.

Oskamp, S. (1965). "Overconfidence in Case-Study Judgments. Judgment under Uncertainty: Heuristics and Biases." The Journal of Consulting Psychology 29: 261-265.

Oskamp, S. (1982). Overconfidence in Case-Study Judgments. Judgment under Uncertainty: Heuristics and Biases. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press: 287-293, Chap. 20.

Paese, P. W. and M. Kinnaly (1993). "Peer input and revised judgment: Exploring the effects of (un)biased confidence." Journal of Applied Social Psychology 23(23): 1989-2011.

Peake, P. K. and D. Cervone (1989). "Sequence anchoring and self-efficacy: Primacy effects in the consideration of possibilities." Social Cognition 7(1): 31-50.

Pears, D. (1974). Freud, Sartre and Self-Deception. Freud. R. Wollheim. New York, Anchor Books: 97-112.

Pei, B. W., S. A. Reed and B. S. Koch (1992). "Auditor Belief Revisions in a Performance Auditing Setting: An Application of the Belief-Adjustment Model." Accounting, Organizations & Society 17(2): 169-183.

Peleg, B. and M. E. Yaari (1973). "On the Existence of a Consistent Course of Action when Tastes are Changing." Review of Economic Studies 40: 391-401.

Perkins, D. N. (1981). The Mind's Best Work, Harvard University Press.

Perloff, L. S. and B. K. Fetzer (1986). "Self-Other Judgments and Perceived Vulnerability to Victimization." Journal of Personality and Social Psychology 50(3): 502-510.

Pfeifer, P. E. (1994). "Are we overconfident in the belief that probability forecasters are overconfident?" Organizational Behavior & Human Decision Processes 58(2): 203-213.

Phelps, E. S. and R. A. Pollak (1968). "On Second-best National Saving and Game-equilibrium Growth." Review of Economic Studies 35: 185-199.

Piaget, J. (1965). The Moral Judgment of the Child. New York, Macmillan Publishing.

Pigou, A. C. (1951). "Some Aspects of Welfare Economics." American Economic Review 43(3): 287-302.

Plous, S. (1991). "Biases in the assimilation of technological breakdowns: Do accidents make us safer?" Journal of Applied Social Psychology 21(13): 1058-1082.

Pollak, R. A. (1968). "Consistent Planning." Review of Economic Studies 102: 201-208.

Pollak, R. A. (1970). "Habit Formation and Dynamic Demand Functions." Journal of Political Economy 78(4 July/August): 745-763.

Pollak, R. A. (1978). "Endogenous Tastes in Demand and Welfare Analysis." American Economic Review 68(2 May): 374-379.

Pratt, J. (1964). "Risk Aversion in the Small and the Large." Econometrica 22: 122-136.

Prelec, D. (1979). Decreasing Impatience: Definition and Consequences: WP 90-015, Harvard Business School.

Pyszczynski, T. and J. Greenberg (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. Advances in experimental social psychology, Vol. 20. B. Leonard, Academic Press, Inc, San Diego, CA, US: 297-340.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." American Economic Review 83(5): 1281-1302.

Rabin, M. (1994). "Cognitive dissonance and social change." Journal of Economic Behavior and Organization 23(1994): 177-194.

Rabin, M. (1995). "Moral Preferences, Moral Constraints, and Self-Serving Biases." Mimeo, University of California at Berkeley, August.

Rabin, M. (1996). Bargaining, Fairness and Efficiency, University of California at Berkeley.

Rabin, M. and J. Schrag (1996). "First Impressions Matter: A Model of Confirmatory Bias." Mimeo. University of California at Berkeley, Emory University, August.

Redelmeier, D. and D. Kahneman (1993). The pain of invasive medical procedure: patients' real time and retrospective evaluations of colonoscopy, University of California, Berkeley.

Redelmeier, D. A. and E. Shafir (1995). "Medical Decision Making in Situations that Offer Multiple Alternatives." Journal of the American Medical Association 273(4): 302-305.

Redelmeier, D. A. and A. Tversky (1990). "The discrepancy between medical decisions for individual patients and for groups." New England Journal of Medicine 322: 1162-1164.

Redelmeier, D. A. and A. Tversky (1992). "On the framing of multiple prospects." Psychological Science 3(3): 191-193.

Redelmeier, D. A. and A. Tversky (1996). "On the belief that arthritics pain is related to the weather." Proceedings of the National Academy of Science USA, Vol. 93, pp. 2895-2896, April 1996.

Roberts, H. V. (1963). "Risk, ambiguity and the Savage axioms: Comment." Quarterly Journal of Economics 77: 327-36.

Romer, D. (1992). Why do firms prefer more able workers?, Department of economics, UC Berkeley.

Rook, D. W. (1987). "The Buying Impulse." Journal of Consumer Research 14: 189-199.

Ross, L. D., M. R. Lepper, F. Strack and J. Steinmetz (1977). "Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood." Journal of Personality & Social Psychology 35(11): 817-829.

Ross, L. (1987). The problem of construal in social inference and social psychology. A distinctive approach to psychological research: The influence of Stanley Schachter. R. E. N. Neil E. Grunberg, Judith Rodin, Jerome E. Singer, Eds., Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US: 118-150.

Ross, M. and F. Sicoly (1979). "Egocentric biases in availability and attribution." Journal of Personality & Social Psychology 37(3): 322-336.

Roth, A. E. (1995). Bargaining Experiments. Handbook of Experimental Economics. J. Kagel and A. Roth. Princeton, Princeton University Press: 253-348.

Roth, A. E. and J. K. Murnighan (1982). "The role of Information in Bargaining: An Experimental Study." Econometrica 50: 1123-1142.

Ruffle, B. J. (1995). Modeling Emotions: A Game-Theoretic Approach, Department of Economics, Princeton University.

Ryder, H. E. and G. M. Heal (1973). "Optimal Growth with Intertemporally Dependent Preferences." Review of Economic Studies 40: 1-33.

Sally, D. (1995). On Sympathy, Johnson Graduate School of Management, Cornell University.

Samuelson, P. (1961). "The St. Petersburg Paradox as a Divergent Double Limit." International Economic Review 1: 31-37.

Samuelson, W. and R. Zeckhauser (1988). "Status Quo Bias in Decision Making." Journal of Risk and Uncertainty 1: 7-59.

Sanitioso, R. and Z. Kunda (1991). "Ducking the collection of costly evidence: Motivated use of statistical heuristics." Journal of Behavioral Decision Making 4(3): 161-176.

Sarin, R. K. and M. Weber (1993). "Effects of ambiguity in market experiments." Management Science 39(5): 602-615.

Sarin, R. K. and R. L. Winkler (1992). "Ambiguity and decision modeling: A preference-based approach." Journal of Risk and Uncertainty 5: 839-407.

Savage, L. J. (1954). The Foundations of Statistics, Wiley: New York.

Savitsky, J. C. and J. Babl (1976). "Cheating, intention, and punishment from an equity theory perspective." Journal of Research in Personality 10(1): 128-136.

Schapiro, C. and J. E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." American Economic Review 74 (June): 433-444.

Schelling, T. C. (1978). "Egonomics, or the Art of Self-Management." American Economic Review 68(2): 290-294.

Schelling, T. C. (1981). "Economic Reasoning and the Ethics of Policy." Public Interest 63: 37-61.

Schelling, T. C. (1984). Choice and Consequence: Perspectives of an Errant Economics, Harvard University Press.

Schelling, T. C. (1992a). Self-command: a new discipline. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 167-176.

Schelling, T. C. (1992b). Addictive Drugs: The Cigarette Experience. Science. 255: 430-433.

Schmeidler, D. (1989). "Subjective probability and expected utility without additivity." Econometrica 57: 571-87.

Schopler, J. and V. Thompson (1968). "Role of attribution in mediating the amount of reciprocity for a favor." Journal of Personality and Social Psychology 10(3): 243-250.

Segal, U. (1987). "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach." International Economic Review 28(1): 175-202.

Segal, U. and A. Spivak (1990). "First Order versus Second Order Risk Aversion." Journal of Economic Theory 51(1): 111-25.

Shafir, E. (1993). "Choosing Versus Rejecting: Why Some Options Are Both Better and Worse Than Others." Memory and Cognition 21(4): 546-556.

Shafir, E., P. Diamond and A. Tversky (1996). Money Illusion, Princeton University.

Shafir, E. B., D. N. Osherson and E. E. Smith (1993). "The Advantage Model: A Comparative Theory of Evaluation and Choice under Risk." Organizational Behavior and Human Decision Processes 55: 325-378.

Shafir, E., I. Simonson and A. Tversky (1993). "Reason-based choice. Special Issue: Reasoning and decision making." Cognition 49(1-2): 11-36.

Shafir, E. and A. Tversky (1992). "Thinking through uncertainty: Nonconsequential reasoning and choice." Cognitive Psychology 24(4): 449-474.

Shalev (1994). An Axiomatization of Loss Aversion in a Multi-period Model, Tel Aviv University Israel Institute of Business.

Shalev, J. (1995). Loss Aversion and Bargaining, IIBR.

Shalev, J. (1995). Loss Aversion in a Multi-Period Model, Tel Aviv University.

Shea, J. (1995). "Union Contracts and the Life-Cycle/Permanent-Income Hypothesis." American Economic Review 85: 186-200.

Shefrin, H. and M. Statman (1993). "Ethics, Fairness and Efficiency in Financial Markets." Financial Analysts Journal 49(6 (Nov/Dec)): 21-29.

Shefrin, H. and R. Thaler (1988). "The Behavioral Life-Cycle Hypothesis." Economic Inquiry 26(October): 609-643.

Shefrin, H. M. and R. H. Thaler (1992). Mental accounting, saving, and self-control. Choice over time. G. Loewenstein and J. Elster, Russell Sage Foundation: 287-329.

Sherman, R. (1974). "The Psychological Difference Between Ambiguity and Risk." Quarterly Journal of Economics 88(1): 166-69.

Simon, H. A. (1961). Administrative Behavior. Second Edition. New York: Macmillan. Original publication: 1947.

Simonson, I. (1989). "Choice based on reasons: The case of attraction and compromise effects." Journal of Consumer Research 16(2): 158-174.

Simonson, I., Z. Carmon and S. O Curry (1994). "Experimental evidence on the negative effect of product features and sales promotions on brand choice." Marketing Science 13(1): 23-40.

Simonson, I., S. M. Nowlis and Y. Simonson (1993). "The effect of irrelevant preference arguments on consumer choice." Journal of Consumer Psychology 2(3): 287-306.

Simonson, I. and A. Tversky (1992). "Choice in context: Tradeoff contrast and extremeness aversion." Journal of Marketing Research 29(3): 281-295.

Simonson, I. and A. Tversky (1992). "Choice in context: Tradeoff contrast and extremeness aversion." Journal of Marketing Research 29(3): 281-295.

Skolnick, A. S. (1986). The Psychology of Human Development, Harcourt Brace Jovanovich.

Slovic, P. (1975). "Choice between equally valued alternatives." Journal of Experimental Psychology: Human Perception and Performance 1: 280-287.

Slovic, P. and S. Lichtenstein (1971). "Comparison of Bayesian and regression approaches to the study of information processing in judgment." Organizational Behavior & Human Performance 6(6): 649-744.

Smith, A. (1759). The Theory of Moral Sentiments. Reprinted, Indianapolis: Liberty Classics.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations. Reprinted R. H. Campbell and A. S. Skinner (eds.). Oxford: Calrendon Press.

Smith, V. L. (1969). "Measuring nonmonetary utilities in uncertain choices: The Ellsberg urn." Quarterly Journal of Economics 83: 324-29.

Smith, V. and J. M. Walker (1993). "Monetary Rewards and Decision Cost in Experimental Economics." Economic Inquiry 31(2): 245-261.

Sniezek, J. A. and F. S. Switzer (1989). The over-underconfidence paradox: High Pi's but poor unlucky me. Paper presented at the Judgment and Decision Making annual meeting in Atlanta, Georgia.

Soll, J. (1996). "Determinants of Overconfidence and Miscalibration: The Roles of Random Error and Ecological Structure." Organizational Behavior and Human Decision Processes 65(2): 117-137.

Stangor, C. (1988). "Stereotype accessibility and information processing." Personality & Social Psychology Bulletin 14(4): 694-708.

Stangor, C. and D. N. Ruble (1989). "Strength of expectancies and memory for social information: What we remember depends on how much we know." Journal of Experimental Social Psychology 25(1): 18-35.

Steiner, D. D. and J. S. Rain (1989). "Immediate and delayed primacy and recency effects in performance evaluation." Journal of Applied Psychology 74(1): 136-142.

Stevenson, R. L. (1884). A Humble Remonstrance. Henry James and Robert Louis Stevenson: A Record of Friendship and Criticism. J. A. Smith, Rupert, Hart-Davis: 86-100.

Stigler, G. J. and J. K. Kindahl (1970). The Behavior of Industry Prices. New York: National Bureau of Economic Research.

Strohmer, D. C. and A. L. Chiodo (1984). "Counselor hypothesis testing strategies: The role of initial impressions and self-schema." Journal of Counseling Psychology 31(4): 511-520.

Strohmer, D. C. and L. J. Newman (1983). "Counselor hypothesis-testing strategies." Journal of Counseling Psychology 30(4): 557-565.

Strohmer, D. C. and V. A. Shivy (1994). "Bias in counselor hypothesis testing: Testing the robustness of counselor confirmatory bias." Journal of Counseling & Development 73(2): 191-197.

Strohmer, D. C., V. A. Shivy and A. L. Chiodo (1990). "Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy." Journal of Counseling Psychology 37(4): 465-472.

Strotz, R. H. (1955). "Myopia and Inconsistency in Dynamic Utility Maximization." Review of Economic Studies 23(3): 165-180.

Stuckless, N. and R. Goranson (1992). "The Vengeance Scale: Development of a measure of attitudes toward revenge." Journal of Social Behavior & Personality 7(1): 25-42.

Summers, L. H. (1993). Does the stock market rationally reflect fundamental values? Advances in behavioral finances. R. H. Thaler, Russell Sage Foundation: 153-166.

Svenson, O. (1981). "Are We Less Skillful than our Fellow Drivers Are?" Acta Psychologica 47: 143-148.

Thaler, R. (1980). "Toward a Positive Theory of Consumer Choice." Journal of Economic Behavior and Organization 1: 39-60.

Thaler, R. H. (1981). "Some empirical evidence of dynamic inconsistency." Economic Letters 81: 201-7.

Thaler, R. (1985). "Mental Accounting and Consumer Choice." Marketing Science 4(3): 199-214.

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

Thaler, R. H. (1988). "Anomalies: The Ultimatum Game." Journal of Economic Perspectives 2(4 (Fall 1988)): 195-206.

Thaler, R. H. (1994). "Psychology and Savings Policies." American Economic Review 84(2): 186-192.

Thaler, R. H. and H. M. Shefrin (1981). "An Economic Theory of Self-Control." Journal of Political Economy 89(2): 392-406.

Thompson, L. (1990). "Negotiation behavior and outcomes: Empirical evidence and theoretical issues." Psychological Bulletin 108(3): 515-532.

Tomassini, L. A., I. Solomon, M. B. Romney and J. L. Krogstad (1982). "Calibration of Auditors' Probabilistic Judgments: Some Empirical Evidence." Organizational Behavior & Human Performance 30(3): 391-406.

Train, K. E., D. L. McFadden and A. A. Goett (1987). "Consumer Attitudes and Voluntary Rate Schedules for Public Utilities." Review of Economics and Statistics 64: 383-391.

Tversky, A. and C. Fox (1995). "Weighting Risk and Uncertainty." Psychological Review 102(2): 269-283.

Tversky, A. and T. Gilovich (1989). "The Cold Facts about the "Hot Hand" in Basketball." Chance 2(1): 16-21.

Tversky, A. and T. Gilovich (1989). "The Hot Hand: Statistical Reality or Cognitive Illusion." Chance 2(4): 31-34.

Tversky, A. and D. Griffin (1991). Endowment and contrast in judgments of well-being., Pergamon Press, Inc, Oxford, England.

Tversky, A. and D. Kahneman (1973a). "Availability: A heuristic for judging frequency and probability." Cognitive Psychology 5(2): 207-232.

Tversky, A. and D. Kahneman (1973b). Judgment under uncertainty: Heuristics and biases., Eugene, Ore.: Oregon Research Inst., Vol. 13(1), 1973.

Tversky, A. and D. Kahneman (1974). "Judgment under uncertainty: Heuristics and biases." Science 185(4157): 1124-1131.

Tversky, A. and D. Kahneman (1981). "The framing of decisions and the psychology of choice." Science 211(4481): 453-458.

Tversky, A. and D. Kahneman (1982a). Judgments of and by Representativeness. Judgment under Uncertainty: Heuristics and Biases. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press: 84-98, Chap. 6.

Tversky, A. and D. Kahneman (1982b). Evidential Impact of Base Rates. Judgment under Uncertainty: Heuristics and Biases. D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press: 153-160, Chap. 10.

Tversky, A. and D. Kahneman (1983). "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probabilistic Judgment." Psychological Review 90(4): 293-315.

Tversky, A. and D. Kahneman (1986). "Rational Choice and the Framing of Decisions." Journal of Business 59: S251-S278.

Tversky, A. and D. Kahneman (1991). "Loss Aversion in Riskless Choice: A Reference-Dependent Model." Quarterly Journal of Economics(November): 1039-1061.

Tversky, A. and D. Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." Journal of Risk and Uncertainty 5(November): 297-323.

Tversky, A., S. Sattath and P. Slovic (1988). "Contingent weighting in judgment and choice." Psychological Review 95(3): 371-384.

Tversky, A. and E. Shafir (1992a). "The disjunction effect in choice under uncertainty." Psychological Science 3(5): 305-309.

Tversky, A. and E. Shafir (1992b). "Choice under conflict: The dynamics of deferred decision." Psychological Science 3(6): 358-361.

Tversky, A. and I. Simonson (1993). "Context-dependent Preferences." Management Science forthcoming(39): 1179-189.

Tversky, A. and R. Thaler (1990). "Anomalies: Preference Reversals." Journal of Economic Perspectives 4(2): 201-211.

Tversky, A. and P. Wakker (1995). "Risk Attitudes and Decision Weights." Econometrica 63(6): 1255-1280.

Tyler, T. (1988). "What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures." Law and Society Review 22(1): 103-135.

Tyson, T. (1990). "Believing That Everyone Else Is Less Ethical: Implications for Work Behavior and Ethics Instruction." Journal of Business Ethics 9(9): 715-721.

Vallone, R. P., D. W. Griffin, S. Lin and L. Ross (1990). "Overconfident prediction of future actions and outcomes by self and others." Journal of Personality & Social Psychology 58(4): 582-592.

van de Kragt, A. J. C., R. M. Dawes and J. M. Orbell "Are People Who Cooperate 'Rational Altruists'?" .

van de Kragt, A., J. Orbell and R. Dawes (1983). "The minimal contributing set as a solution to public goods problems." American Political Science Review 77: 112-122.

van Lenthe, J. (1993). "ELI: An interactive elicitation technique for subjective probability distributions." Organizational Behavior & Human Decision Processes 55(3): 379-413.

van Praag, B. M. S. (1971). "The Individual Welfare Function of Income in Belgium: An Empirical Investigation."
Varey, C. A. and D. Kahneman (1992). "Experiences extended across time: Evaluation of moments and episodes." Journal of Behavioral Decision Making 5(3): 169-185.

Varey, C. A. and D. Kahneman (1992). "Experiences extended across time: Evaluation of moments and episodes." Journal of Behavioral Decision Making 5(3): 169-185.

Varian, H. (1974). "Equity, Envy, and Efficiency." Journal of Economic Theory 9: 63-91.

Wason, P. C. (1960). "On the failure to eliminate hypotheses in a conceptual task." Quarterly Journal of Experimental Psychology 12: 129-140.

Wason, P. C. (1968). "Reasoning about a rule." Quarterly Journal of Experimental Psychology 23: 273-281.

Wakker, P. (1989). "Continuous Subjective Expected Utility with Non-Additive Probabilities." Journal of Mathematical Economics 18(1): 1-27.

Weber, M. and C. Camerer (1987). "Recent developments in modeling preferences under risk." OR Spektrum 9: 129-51.

Weber, M., G. Loomes, H. Keppe, G. Meyer-Delius (1994). "Framing Effects in Experimental Markets." Lehrstuehle fuer Finanzwirtschaft, Universitaet Mannheim.

Weg, E. and V. Smith (1993). "On the failure to induce meager offers in ultimatum games." Journal of economic psychology 14(N1): 17-32.

Weil, P. (1990). "Nonexpected Utility in Macroeconomics." Quarterly Journal of Economics 105(1): 29-42.

Weinstein, N. D. (1980). "Unrealistic Optimism About Future Life Events." Journal of Personality and Social Psychology 39(5): 806-820.

Weinstein, N. D. (1984). "Why it won't happen to me: Perceptions of risk factors and susceptibility." Health Psychology 3(5): 431-457.

Wertenbroch, K. (1993). "Consumer Self-Control: How Much Vice and Virtue Do We Buy?", Graduate School of Business, University of Chicago.

White, M. J., D. R. Brockett and B. G. Overstreet (1993). "Confirmatory bias in evaluating personality test information: Am I really that kind of person?" Journal of Counseling Psychology 40(1): 120-126.

Whitehead, G. I., S. H. Smith and T. L. McNabney (1987). "Primacy and recency effects on predictions of future performance as a function of the presentation of past performance." Journal of Social Behavior & Personality 2(4): 543-550.

Wilke, H. and J. T. Lanzetta (1970). "The obligation to help: The effects of amount of prior help on subsequent helping behavior." Journal of Experimental Social Psychology 6(4): 488-493.

Williamson, O. (1985). The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. New York: The Free Press.

Wilson, T. D. and S. J. LaFleur (1995). "Knowing what you'll do: Effects of analyzing reasons on self-prediction." Journal of Personality & Social Psychology 68(1): 21-35.

Wilson, T. D. and J. W. Schooler (1991). "Thinking too much: Introspection can reduce the quality of preferences and decisions." Journal of Personality & Social Psychology 60(2): 181-192.

Winman, A. and P. Juslin (1993). "Calibration of sensory and cognitive judgments: Two different accounts." Scandinavian Journal of Psychology 34(2): 135-148.

Wood, J. V. (1989). "Theory and Research Concerning Social Comparisons of Personal Attributes." Psychological Bulletin 106: 231-248.

Wright, A. A. (1994). "Primacy effects in animal memory and human nonverbal memory." Animal Learning & Behavior 22(2): 219-223.

Wyatt, D. F. and D. T. Campbell (1951). "On the Liability or Stereotype of Hypothesis." Journal of Abnormal and Social Psychology 46: 496-500.

Yaari, M. E. (1977). Consistent Utilization of an Exhaustible Resource, Or, How to Eat an Appetite-Arousing Cake, Center for Research in Mathematical Economics and Game Theory Research memorandum No. 26.

Yaari, M. E. and M. Bar-Hillel (1984). "On Dividing Justly." Social Choice and Welfare 1984: 1-24.

Yates, J. F. (1990). Judgment and decision making. Prentice-Hall, Inc; Englewood Cliffs, NJ, US.

Yates, J. F. and S. P. Curley (1986). "Contingency judgment: Primacy effects and attention decrement." Acta Psychologica 62(3): 293-302.

Yates, J. F. and L. G. Zukowski (1976). "Characterization of ambiguity in decision making." Behavioral Science 21(1): 19-25.

Young, H. P. (1993). "The Evolutions of Conventions." Econometrica 61: 57-84.

Zuckerman, M. (1979). "Attribution of Success and Failure Revisited, or: The motivational bias is alive and well in attribution theory." Journal of Personality 47(2): 245-287.

Zuckerman, M., C. R. Knee, H. S. Hodgins and K. Miyake (1995). "Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set." Journal of Personality & Social Psychology 68(1): 52-60.