Psychology and Economics

Matthew Rabin

Journal of Economic Literature, Vol. 36, No. 1 (Mar., 1998), 11-46.

Stable URL:
http://links.jstor.org/sici?sici=0022-0515%28199803%2936%3A1%3C11%3APAE%3E2.0.CO%3B2-B

Journal of Economic Literature is currently published by American Economic Association.

Journal of Economic Literature
Vol. XXXVI (March 1998), pp. 11–46

Psychology and Economics

MATTHEW RABIN
University of California at Berkeley

For comments on this essay or discussion on related topics, I thank Henry Aaron, George Akerlof, James Andreoni, Steven Blatt, Gary Charness, Eddie Dekel, Peter Diamond, Jon Elster, Erik Eyster, Ernst Fehr, David I. Levine, George Loewenstein, Rob MacCoun, James Montgomery, Vai-Lam Mui, Ted O'Donoghue, Drazen Prelec, Ilya Segal, Eldar Shafir, Gene Smolensky, Joel Sobel, Amos Tversky, four anonymous referees, 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 to the National Science, Russell Sage, Alfred P. Sloan, and MacArthur Foundations for financial support on research related to this essay. A more verbose and citation-heavy version of this essay, covering a few additional topics, is available as University of California at Berkeley Department of Economics Working Paper No. 97–251.

1. *Introduction*

BECAUSE PSYCHOLOGY systematically explores human judgment, behavior, and well-being, it can teach us important facts about how humans differ from the way they are traditionally described by economists. In this essay I discuss a selection of psychological findings relevant to economics.

Economics has conventionally assumed that each individual has stable and coherent preferences, and that she rationally maximizes those preferences. Given a set of options and probabilistic beliefs, a person is assumed to maximize the expected value of a utility function, $U(x)$. Psychological research suggests various modifications to this conception of human choice. Section 2 provides examples of what psychological research can teach us about making $U(x)$ more realistic than under standard economic assumptions. I begin by discussing research which suggests that a person's preferences are often determined by changes in outcomes relative to her reference level, and not merely by absolute levels of outcomes. In particular, relative to their status quo (or other reference points), people dislike losses significantly more than they like gains. I then discuss how people depart from pure self-interest (as narrowly defined) to pursue "other-regarding" goals such as fairness, reciprocal altruism, and revenge.

By focusing on evidence consistent with rational choice, Section 2 reviews evidence that requires relatively small modifications of the familiar economic framework. Section 3 reviews research on biases in judgment under uncertainty; because these biases lead people to make errors when attempting to maximize $U(x)$, this research poses a more radical challenge to the economics model. I discuss two biases in detail—how we infer too much from too little evidence and how we misread evi-

dence as confirming previously held hypotheses—and briefly explain a few more. I also discuss some of the psychological evidence on whether and when experience and learning 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 judgment when attempting to maximize $U(x)$, it is sometimes misleading to conceptualize people as attempting to maximize a coherent, stable, and accurately perceived $U(x)$. I begin by reviewing evidence that people have difficulties 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 conclude Section 4 with a discussion of self-control problems and other phenomena that arise because people have a short-run propensity to pursue immediate gratification that is inconsistent with their long-run preferences.

The customary disclaimer of review essays—that they do not pretend to be exhaustive—applies even more here than usual: The topics covered are only a small fraction of economically relevant psychology. Some omitted topics of obvious economic relevance are non-expected utility; status, envy, and social comparisons; conformity and herd behavior; self-serving biases and motivated cognition; 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); and a mass of research on learning from cognitive and developmental psychology. [1]

Two other limitations of scope are conspicuous. First, I emphasize what psychologists and experimental economists have learned about people, rather than *how* they have learned it. Consequently, the focus of this essay is not at all on experimental methods per se.[2] Second, my primary emphasis is on reviewing what psychology tells us about modifying our general assumptions about individual behavior, 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 leave the exploration of these implications to other forums.

Mainstream economics employs a powerful combination of methods: methodological individualism, mathematical formalization of assumptions, logical analysis of the consequences of those assumptions, and sophisticated empirical field testing. I believe these methods are tremendously useful, and

[1] 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) meant to capture cognitive limits of economic actors, but which do not invoke research on the specific patterns of errors that human beings make. For an excellent review of the role of bounded rationality in economics, see John 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 John Kagel and Alvin Roth (1995). While most of the findings I present in this essay come from psychologists, I 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.

an underlying premise of this essay is that we should strive to understand psychological findings in light of these methods.[3] These methods raise problems for doing full justice to behavioral reality: Because of the high premium economics places on the logic and precision of arguments and the quantification of evidence, attending to all facets of human nature is neither feasible nor desirable. The realization that many details of human behavior must be ignored, however, should not license institutionalized complacency about the behavioral validity of our assumptions; "tractability" and "parsimony" should be guiding principles in our efforts to make our research more realistic, not pretexts for avoiding this task. As it now stands, some important psychological findings seem tractable and parsimonious enough that we should begin the process of integrating them into economics. Incorporating other findings will take longer. But even in those 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. While I conclude in Section 5 by briefly discussing the general endeavor of incorporating psychological findings into economics, I have omitted from this essay many of the classical meta-arguments about whether it is "possible" that behavioral research identifies important departures from economists' habitual assumptions about human behavior. Of course it is possible, and in fact it is true. In this essay, I review what research shows some of those departures are.

2. *Preferences*

This section discusses how psychological findings suggest we modify the utility functions economists employ.

A. *Reference Levels, Adaptation, and Losses*

Overwhelming evidence shows that humans are often more sensitive to how their current situation differs from some reference level than to the absolute characteristics of the situation (Harry Helson 1964). For instance, the same temperature that feels cold when we are adapted to hot temperatures may appear hot when we are adapted to cold temperatures. Understanding that people are often more sensitive to changes than to absolute levels suggests that we ought incorporate into utility analysis such factors as habitual levels of consumption. Instead of utility at time 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 a utility function of the form $u_t(r_t;c_t)$, utility should be written in a more general form, $u_t(r_t;c_t)$. While some economists have over the years incorporated reference dependence into their economic analysis, it is fair to say that the ways and degrees to which reference points influence behavior have not fully been appreciated by economists.[4] Researchers have identified a pervasive feature of reference dependence: In a wide variety of domains, people are significantly more averse to losses than

[3] Indeed, I have organized the essay to reflect this premise. At times, difficulties in organization belie deeper difficulties of how to fit this research into the economics framework. For instance, while Section 2 discusses approaches to modeling loss aversion fully within the rational-choice framework, material in Section 4 suggests loss aversion isn't always best conceived of as compatible with utility maximization.

[4] There have been some earlier examples of economists considering reference dependence (James Duesenberry 1949; Richard Easterlin 1974; and Harl Ryder and Geoffrey Heal 1973).

they are attracted to same-sized gains (Kahneman, Jack Knetsch, and Thaler 1990). One realm where such *loss aversion* plays out is in preferences over wealth levels. Tversky and Kahneman (1991) suggest that in the domain of money (and in others where the sizes of losses and gains can be measured), people value modest losses roughly twice as much as equal-sized gains. 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. But loss aversion says that the value function abruptly changes slope at the reference level, so that people dislike even small-scale risk. For instance, most people prefer their status quo to a 50/50 bet of losing $10 or gaining $11. The standard concave-utility-function explanation for risk aversion is simply not a plausible explanation of such risk attitudes. Rajnish Mehra and Edward Prescott (1985) and Larry Epstein and Stanley Zin (1990) have, for instance, observed that the expected-utility framework cannot simultaneously explain both the small-scale and large-scale risk attitudes implied by macro data, and Rabin (1997) provides a "calibration theorem" that indeed shows that *no* concave utility function can simultaneously explain plausible small-scale and large-scale risk attitudes. A reference-based kink in the utility function is required to explain such risk attitudes within the expected-utility framework.[5]

Loss aversion is related to the striking *endowment effect* identified by Thaler (1980): Once a person comes to possess a good, she immediately values it more than before she possessed it. Kahneman, Knetsch, and Thaler (1990) nicely illustrate this phenomenon. They randomly gave mugs worth about $5 each to one group of students. Minimal selling prices were elicited from those given the mugs (with an incentive-compatible procedure that ensured honest reports). Minimal "prices"—sums of money such that they would choose that sum rather than the mug—were elicited from another group of subjects not given mugs. These two groups, "sellers" and "choosers," faced precisely the same choice between money and mugs, but their reference points differed: Those who were randomly given mugs treated the mugs as part of their reference levels or endowments, and considered leaving without a mug to be a loss, whereas individuals not given mugs considered leaving without a mug as remaining at their reference point. In one experiment, the median value placed on the mug was $3.50 by choosers but $7.00 by sellers. Such results have been replicated repeatedly by many researchers in many contexts.

As established by Knetsch and John Sinden (1984), William Samuelson and Richard Zeckhauser (1988), and Knetsch (1989), a comparable phenomenon—the *status quo bias*—holds in multiple-good choice problems. Here, loss aversion implies that individuals tend to prefer the status quo to changes that involve losses of some goods, even when these losses are offset by gains of other goods. Knetsch and Sinden (1984) and Knetsch (1989), for instance, randomly gave students either candy bars or 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. Ninety percent of *both* mug-owners and candy-owners chose not to trade. Because the goods were allocated randomly and transaction costs were

[5] Uzi Segal and Avia Spivak (1990), and others, however, develop a model of such first-order risk aversion outside the expected-utility framework.

minimal, the different behavior for the two groups of subjects presumably reflected preferences that were induced by the initial allocation. Knetsch (1989) and Tversky and Kahneman (1991) show that such preferences can be usefully captured by utility functions defined over reference levels as well as consumption levels.

In addition to loss aversion, another important reference-level effect is *diminishing sensitivity*: The marginal effects in perceived well-being are greater for changes close to one's reference level than for changes further away. As Kahneman and Tversky (1979) note, diminishing sensitivity is a pervasive pattern of human perception, where our perceptions are a concave function of the magnitudes of change. For instance, we are more likely to discriminate between 3° and 6° changes in room temperature than between 23° and 26° changes. This applies to both increases and decreases in temperature. In the context of preferences over uncertain monetary outcomes, diminishing sensitivity implies that the slope of a person's utility function over wealth becomes flatter 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. Kahneman and Tversky (1979) found that 70 percent of subjects report that they would prefer a 3/4 chance of losing nothing and 1/4 chance of losing $6,000 to a 2/4 chance of losing nothing and 1/4 chance each of losing $4,000 or $2,000. Because the preferred lottery here is a mean-preserving spread of the less-preferred lottery, the responses of 70 percent of the subjects are inconsistent with the standard concavity assumption.[6]

In order to study the effects of reference points in a dynamic utility-maximization framework, we need to take into account how people feel about the effects their current choices have on their future reference points. To maximize their long-run utilities when reference points matter, people must determine two things beyond how they feel about departures from reference points: How current behavior affects future reference points and how they feel about changes in their reference points. Economists Ryder and Heal (1973) model the process by which reference points change with the formula $r_t \equiv \alpha c_{t-1} + (1-\alpha) r_{t-1}$, 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, though there seems to be little evidence on this topic.[7] Evidence is similarly sparse about how people's preferences depend on changes in reference points.[8] With-

[6] Because this "risk-loving" tendency is in conflict with the diminishing marginal utility of income, however, it has often been conjectured that risk aversion may reappear for large losses that might push a person to extremely low consumption levels.

[7] Bowman, Minehart, and Rabin (1997) combine the Ryder and Heal approach of rational-expectations, reference-dependent utilities with a utility function that incorporates loss aversion and di- minishing sensitivity. Duesenberry (1949) implicitly posited a reference function closer to $r_t \equiv \text{Max}\ \{c_\tau : \tau < t\}$—that is, a person's reference level was her highest past consumption level.

[8] For exceptions, see Loewenstein and Nachum Sicherman (1991), Robert Frank (1985 ch. 15, 1989), and Frank and Robert Hutchens (1993), who have identified a tendency for people to prefer income profiles that are steady or increasing over time to same-sized decreasing profiles, strongly indicating that people prefer not to become accustomed to levels of consumption they know they cannot maintain. For consideration of

out assumptions about these relationships, there will be a relatively small set of circumstances where loss aversion and diminishing sensitivity can be integrated into models of dynamic utility maximization.

There have been some initial attempts to study loss aversion, the endowment effect, and the status quo bias in economic contexts. Raymond Hartman, Michael Doane, and Chi-Keung Woo (1991) find empirical evidence for the existence of a status quo bias in consumer demand for electricity; Bowman, Minehart, and Rabin (1997) replicate evidence by John 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.

B. *Social Preferences and Fair Allocations*

It is common for undergraduates to encounter the following quote from *The Wealth of Nations* (Adam Smith 1776, pp. 26–27) 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 probably no other two-word description of human motives comes close to "self-interest" in explaining economic behavior. Yet pure self-interest is far from a complete description of human motivation, and realism suggests that economists should move away from the presumption that people are *solely* self-interested. Robyn 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.

Examples of economic behavior induced by social goals are donations to public television stations, voluntary reductions of water-use during droughts, and conservation of energy to help solve the energy crisis (Kenneth Train, Daniel McFadden, and Andrew Goett 1987). One context in which fairness has been studied is monopoly pricing (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

---

some other factors involved in the preference for increasing wage profiles, see Loewenstein and Sicherman (1991, pp. 76–82).

of industrial relations (Stacy Adams 1963; Akerlof 1982; and Akerlof and Janet Yellen 1990).

Experimental research makes clear that preferences depart from pure self-interest in non-trivial ways: Subjects contribute to public goods more than can be explained by pure self-interest; they often share money when they could readily grab it for themselves; and they often sacrifice money to retaliate against unfair treatment. Debates remain, as some researchers have emphasized the possibility of producing laboratory environments (e.g., well-organized, private-information double-auction spot markets) that induce behavior that is closer to purely self-interested than is behavior in other settings. But, disentangling debates over the nature of preferences from strong ancillary assumptions about which institutions and environments matter in the real world, there does not appear to be debate among behavioral researchers about whether underlying preferences depart non-trivially from pure self-interest.

One form of social motivations on which economists have focused is *altruism*, in the sense that people 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 $U_1(x) \equiv (1-r)\cdot\Pi_1(x) + r\cdot\Pi_2(x)$,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; by assuming $r > 0$, we can investigate when and how concern for others affects behavior and welfare.

Simple altruism may parsimoniously capture important phenomena in many contexts. But there is a mass of experimental evidence that indicates it is often a very wrong description of social preferences.[9] To get a sense for how social preferences differ from simple altruism, consider what can be called "behavioral distributive justice": How do people choose to divide resources?[10] 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 $U_1 \equiv (1-r)\cdot\Pi_1 + r\cdot W_1(\Pi_1,\Pi_2)$, where W_1 is person 1's view of the proper allocation, and Π_1 is (as above) her self-interested payoff, and $0 \le r \le 1$ measures the weight this person puts on self-interest versus proper allocation. 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 worth of money or other goods on the ground. How would the average person, acting as a third

[9] See Dennis 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 "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 Max $U_1(x) \equiv (1-r)\cdot\Pi_1(x) + r\Pi_2(x)$, we must still resolve whether Person 1's experienced well-being is described by $U_1(x)$ or $\Pi_1(x)$.

[10] 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.

party, decide to split the surplus between the two?[11] The simple-altruism perspective would suggest solutions such as giving the surplus to the person who is poorer—or for non-money goods—the person who values the goods more. But research shows that many people in many contexts do not find the "maximal-benefits" criterion attractive. One simple alternative norm is prevalent: Resources should be split 50–50. Except in extreme cases, often we ignore issues of relative usefulness and 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 often seem 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. Because it takes a greater allocation to increase the utility of a person who values a resource less, the maximin criterion typically implies that *more* than half the resources are allocated to the person who values those resources *less*. Consider the following hypothetical situation that Maya Yaari and Menahem Bar-Hillel (1984, p. 8) posed to 163 subjects:

> 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 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. (Such an allocation would seem to accord somewhat to 50–50 norms.) The maximin allocation, however, 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." 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):

[11] I assume the surplus 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 the massive psychological literature on "equity theory" shows that people feel that those who have put more effort into creating resources have more claim on those resources (Ellen Berscheid, David Boye, and Elaine Walster 1968).

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). Because the maximin solution is subtle, the results strongly suggest subjects thought about the problem, and did not merely choose some simple focal point.[12]

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 milligrams of Vitamin F from both fruits, rather than 50 milligrams. 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 Smith than they are to Jones. Positing that, "Sooner or later . . . [the maximin criterion] runs the risk of becoming morally unsound," Yaari and Bar-Hillel push the limits of the maximin criterion by telling subjects that Smith derives only 9.1 milligrams of Vitamin F from both fruits, so that grapefruits are 11 times more valuable to Jones than Smith. While a far greater number of respondents now seemed willing to tolerate unequal welfare gains, *still* only 12 percent of respondents felt that all 12 grapefruits should be given to Jones. Moreover, only 18 percent thought that Jones should get more than half the grapefruits while 38 percent thought Smith should get more than half.

While Yaari and Bar-Hillel 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, her choice will depend both on what she feels is a just allocation and on how much she values a just outcome relative to her self-interest. Andreoni and John Miller (1996) consider just such a situation.[13] They asked each subject to unilaterally allocate money between herself and an anonymous second party. Subjects were given various "exchange rates" for allocating money; for instance, some subjects were told that for every $1 they sacrificed, the other party would get $3. Over half the subjects behaved in a way significantly inconsistent with pure self-interest. Among those whose behavior was least self-interested, two thirds chose approximately the maximin allocation rule, and one third chose approximately the dollar-maximizing allocation. These results

[12] The very subtlety of the maximin solution, however, points to a concern with these data that Yaari and Bar-Hillel themselves draw attention to: The questions here were posed in writing at the end of a college entrance exam, which may have induced subjects to treat the questions as problems to solve, rather than as opinions they should express. Informal surveys by Yaari and Bar-Hillel in very different contexts, and experimental results reported below suggest that the problem-solving context does not fully explain the patterns they found in this study.

[13] 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.

suggest that the maximin rule resonates with many people even when allocating money, and even when self-interest is at stake. Allowing proportions of a "pie" to have different monetary values to different parties has also been a theme in the experimental bargaining research (Roth and J. Keith Murnighan 1982; and Kagel, Chung Kim, and Donald Moser 1996). In these settings, disentangling self-interested bargaining strategies from preferences for just allocations is quite difficult, but results seem comparable to Andreoni and Miller's.

In moving from abstract, context-free allocation problems to everyday economic fairness judgments, things become significantly more complicated. First, as elsewhere, reference levels are crucial. Thaler (1985) 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. Relatedly, people's general perceptions of fair behavior may adjust over time. Kahneman, Knetsch, and Thaler (1986a, 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." Robert Franciosi et al. (1995) experimentally support this hypothesis by testing reactions to unfair price increases in a laboratory posted-offers market; they show that the role of fairness considerations in price-determination diminishes with repetition, suggesting that in competitive spot markets people may eventually come to believe that the prevailing market price is fair. Other experiments find virtually no change in either behavior or perceptions of fairness over time (Fehr and Armin Falk 1996). In any event, because adjustments of fairness judgments are not immediate, fairness considerations may help explain the sort of medium-run wage and price stickiness studied by macroeconomists, and evidence that market outcomes are likely to be self-interested exists only for competitive spot markets. See, for example, Fehr, Erich Kirchler, and Andreas Weichbold (1994) for an experimental study of labor markets where behavior never converges to the self-interested outcome.

Finally, in attempting to capture behavioral findings with models of social preferences, it is important to note that people seem to implicitly (but pervasively) consider equitable sharing over *changes* in total endowments, not total endowments themselves. Preferences defined over final wealth states cannot plausibly explain rules such as 50/50 sharing or the maximin criterion. With plausible assumptions about *initial* endowments entering any moderate size division-of-the-pie situation, *any* social welfare function defined with respect to overall consumption levels will almost always yield all-or-nothing allocations. Apparently, people generally have a one-pie-at-a-time conception of fair-division problems. 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. Nevertheless, any attempt to capture behavioral norms of fairness and dis- tributional justice with formal models of social preference must confront the "piecemeal" nature of these norms.

C. *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. Rather, social preferences over other people's consumption depend on the behavior, motivations, and intentions of those other people. The same people who are altruistic toward deserving people are often indifferent to the plight of undeserving people, and motivated to hurt those whom they believe to have misbehaved. 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*. 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. First note that, because the marginal social value of water is greater the less water there is, there are diminishing social benefits of conservation: 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 would cause you to intensify your conservation efforts. This prediction is inconsistent with intuition and empirical evidence: People are more inclined to conserve water if they think other people are conserving, *not* if they think others are splurging. People reciprocate the lack of public spiritedness in others—they don't counteract it.

Evidence in support of reciprocal altruism comes from experimental studies of the voluntary provision of public goods. Dawes and Thaler (1988) conclude that, for most experiments of one-shot public-good decisions in which the individually optimal self-interested contribution is close to zero percent, the contribution rate varies between 40 percent and 60 percent of the socially optimal level.[14] Many of these experiments hint that contributions toward public goods are not the result of simple altruism, though the evidence for reciprocal altruism is varied, and often indirect. For instance, Rachel Croson (1995) finds a strong positive correlation between subjects' contribution levels to a public good and their beliefs about how much others were contributing. Further indirect evidence is that pre-decision communication greatly enhances cooperation (David Sally 1995). One reason why communication enhances contributions may be that reciprocal altruism essentially turns public-goods situations into "coordination games," where high contributions are efficient equilibria, and low contributions are inefficient equilibria. As in general coordination games, therefore, pre-game communication can help players coordinate on the efficient equilibria.

Indeed, the common emphasis when describing the prisoner's dilemma on

[14] 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 Claudia Keser (1996) for evidence that generous contributions are not merely an artifact of experimental design.

the inability of the two captured prisoners to communicate with each other indicates that we implicitly believe that the prisoner's dilemma also really amounts to a coordination game. If defecting were truly a dominant strategy, pre-game communication would not matter. More direct evidence of reciprocal altruism, in the context of the prisoner's dilemma, comes from Shafir and Tversky (1992). When subjects were told that their anonymous partner in a prisoners' dilemma had cooperated, 16 percent also cooperated; when subjects were told that their partner did not cooperate, only 3 percent 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 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 want to punish a firm for being unfair.

A crucial feature of the psychology of reciprocity is that people determine their dispositions toward others according to motives attributed to these others, not solely according to actions taken. When motivated by reciprocal altruism, for instance, 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, Richard Goranson and Leonard Berkowitz (1966, p. 229) conducted an experiment in which confederates posing as subjects were in a position to help real subjects fill out some worksheets. One third of the subjects were told that the confederate had voluntarily offered to help; one third were told that the experimenter had *instructed* the confederate to help; and one third were told that the confederate *might* be willing to help, but the confederate was instructed to refuse to help. When the subjects were later given an opportunity to assist the confederates, they reciprocated earlier help, but did so significantly more when it was voluntary than when it was involuntary.

Volition is also central to the propensity to retaliate against negative actions. Sally Blount (1995) asked subjects about their willingness to accept take-it-or-leave-it offers made by anonymous other parties on how to split $10.[15] 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. Another group was told that *a third party* (also an anonymous student) was to determine the offer made. In this variant, the person 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 by the subject's decision. A final group of subjects were told that the offer would be generated *randomly* by a computer-simulated roulette wheel. In one study, the average minimal acceptable offers for those groups were $2.91, $2.08, and $1.20. That is, people did reject very low of-

[15] The "ultimatum game" of the sort studied by Blount was first developed by Werner Güth, Rolf Schmittberger, and Bernd Schwarze (1982). For reviews of the (massive) literature developed since, see Thaler (1988), Güth and Reinhard Tietz (1990), and Camerer and Thaler (1995).

fers even if computer or third-party generated, but were less keen to reject offers which were not the result of volition by the person who would be hurt by the rejections.

The importance of intentions goes even further, however, than consideration of whether a person's actions are voluntary. Suppose, for instance, you are eating lunch with parents and a 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 uncoordinated attempt to *prevent* a spill, they are probably less angry.[16]

Such examples indicate that interpreting other peoples' motives depend on what we believe *their beliefs* about the consequences of their actions are. Another example of the importance of beliefs is, if you think somebody has been generous to you solely to get a bigger favor from you in the future, then you do not view his generosity to be as pure as if he had expected no reciprocity from you. For example, Arnold Kahn and Thomas Tice (1973) found that subjects' reactions to others' statements of intentions depended on whether they thought those making statements knew that their intentions would be made known to the subjects.[17]

The role of reciprocity and volition appears in some important economic contexts. Akerlof (1982) posits that firms and workers can be thought of as engaging in "gift exchange," a view of social exchange emphasized in sociology and especially anthropology. 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. Fehr, Georg Kirchsteiger, and Arno Riedl (1993) tested this hypothesis in laboratory models of labor markets. Subjects were assigned roles as "firms" or "workers." Firms offered a wage—involving a real monetary transfer from firm to worker—and workers responded by choosing an "effort" level, where this effort was monetarily costly to workers. The results were that most workers chose effort levels higher than their money-maximizing levels. Moreover, while low wages induced little or no effort by workers, workers rewarded firms for setting high wages by providing high effort.

What is the source of high effort lev-

[16] I have confirmed this hypothesis in a field study conducted in North London.

[17] Frank (1994, p. 21) tells the following story that colorfully summarizes the importance of intentions: 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 plays off the obvious silliness of presuming that the second boy's satisfaction with events only depends on the resulting allocation. Yet this comical presumption is the basis of most attempts by economists to model "social preferences" defined solely over outcomes. To formalize the role of intentions in fairness judgments, Rabin (1993) adopts the framework developed by John Geanakoplos, David Pearce, and Ennio Stacchetti (1989), who modify conventional game theory by allowing payoffs to depend on players' beliefs as well as their actions. By positing that my beliefs about your beliefs are arguments in my utility function, we can model my beliefs about your motives as directly influencing my fairness judgments.

els by workers in response to high wages by firms? While workers may simply be choosing to share some of their additional wealth from higher wages with the firm, they may also be reciprocating the volitional generosity of firms. Charness (1996) conducts experiments that helps us differentiate these hypotheses. In Fehr, Kirchler, and Weichbold (1994), it is clear to the worker-subjects that the firms choose wages of their own volition. Charness (1996) replicates this condition, but also conducts variants of the experiment where 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. Results indicated 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.

3. *Biases in Judgment*

Economists traditionally have 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. Tversky and Kahneman (1974, p. 1124) help conceptualize observed departures from perfect rationality by noting that people rely on "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. As the quote clearly suggests, the research described here does not at all suggest economists should abandon the assumption that people are intelligent and purposive in their decision making. Rather the research explores how people depart from *perfect* rationality, positing biases that represent specific and systematic ways that judgment departs from perfect rationality. For the remainder of this section, I describe some of this research, presenting two biases at length and more quickly outlining several others. I conclude by discussing some evidence for when people do and don't learn to correct biases.

A. *The Law of Small Numbers*

According to a bias called "the law of small numbers" (Tversky and Kahneman 1971), people exaggerate how closely a small sample will resemble the parent population from which the sample is drawn.[18] 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. Be-

[18] Kahneman and Tversky relate the law of small numbers to people's tendency to under-use base rates. Tversky and Kahneman (1974) provide evidence for the *representativeness heuristic*. Bayes's 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.

cause we expect close to the same probability distribution of types in small groups as in large groups, for example, we tend to view it as comparably likely that at least 80 percent of 20 coin flips will come up heads than that at least 80 percent of 5 coin flips will come up heads; in fact, the probabilities are about 1 percent and 19 percent, respectively. 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 percent said that they thought the number of days would be about the same. Only 22 percent of subjects correctly answered 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.[19]

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 gambler's fallacy leads people to over-infer the probability distribution from short sequences. Because we underestimate the frequency of a mediocre financial analyst making lucky guesses three times in a row, 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. 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

[19] While people believe in the law of small numbers, they apparently don't believe in the law of large numbers: We *underestimate* the resemblance that large samples will have to the overall population. Kahneman and Tversky (1982a), 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 percent. To overstate it a bit, people seem to have a universal probability distribution over sample means that is insensitive to the sample size.

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. 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. Thomas Gilovich, Robert Vallone, and Tversky (1985) and Tversky and Gilovich (1989a, 1989b) have argued that the almost universally accepted phenomenon of the hot hand is non-existent in basketball. The exaggerated belief in hot hands seems partly explained by the misperception that purely random streaks are too long to be purely random.

B. *Belief Perseverance and Confirmatory Bias*

A range of research suggests that once forming strong hypotheses, people are often too inattentive to new information contradicting their hypotheses. Once you become convinced that one investment strategy is more lucrative than another, you may not sufficiently attend to evidence suggesting the strategy is flawed. A particularly elegant demonstration of such "anchoring" is found in Jerome Bruner and Mary 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 and final degree of focus were identical for all subjects. 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. David Perkins (1981) argues that such experiments provide support for 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.

This form of anchoring does not necessarily imply that people *misinterpret* additional evidence, only that they ignore additional evidence. Psychological evidence reveals a stronger and more provocative phenomenon: People tend to *misread* evidence as *additional* support for initial hypotheses.[20] If a teacher initially believes that one student is smarter than another, she has the propensity to confirm that hypothesis when interpreting later performance.

Some evidence for confirmatory bias is a series of experiments demonstrating how providing the *same* ambiguous information to people who differ in their

[20] For a formal model of confirmatory bias, see Rabin and Schrag (1997).

initial beliefs on some topic can move their beliefs *further apart*. To illustrate such polarization, Charles Lord, Lee Ross, and Mark Lepper (1979, pp. 2102) 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 of 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 toward the death penalty, and how they had changed their beliefs regarding its deterrent efficacy. At confidence levels of $p < .01$ or stronger, Lord, Ross, and Lepper found that proponents of the death penalty became on average more in favor of the death penalty and believed more in its deterrent efficacy, while opponents became even *less* in favor of the death penalty and believed even *less* in its deterrent efficacy. Scott Plous (1991) replicates the Lord, Ross, Lepper results in the context of judgment about the safety of nuclear technology.[21]

John Darley and Paget Gross (1983) demonstrate a related and similarly striking form of polarization. Seventy undergraduates were asked to assess a nine-year-old girl's academic skills in several different academic areas. Before completing this task, the students received information about the girl and her family and viewed a video tape of the girl playing in a playground. One group of subjects was given a fact sheet that described the girl's parents as college graduates who held white-collar jobs; these students viewed a video of the girl playing in what appeared to be a well-to-do suburban neighborhood. The other group of subjects was given a fact sheet that described the girl's parents as high school graduates who held blue-collar jobs; these students viewed a video of the same girl playing in what appeared to be an impoverished inner city neighborhood. Without being supplied any more information, half of each group of subjects was then asked to evaluate the girl's reading level, measured in terms of equivalent grade level. There was a small difference in the two groups' estimates—those subjects who had viewed the "inner-city" video rated the girl's skill level at an average of 3.90 (i.e., 9/10 through 3rd grade) while those who had viewed the "suburban" video rated the girl's skill level at an av-

[21] 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. 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 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.

erage of 4.29. The remaining subjects in each group were shown a second video of the girl answering (with mixed success) a series of questions. Afterwards, they were asked to evaluate the girl's reading level. The inner-city video group rated the girl's skill level at an average of 3.71, significantly *below* the 3.90 estimate of the inner-city subjects who did not view the question-answer video. Meanwhile, the suburban video group rated the girl's skill level at an average of 4.67, significantly *above* the 4.29 estimate of the suburban subjects who did not view the second video. Even though the two groups viewed the *identical* question-and-answer video, the additional information further polarized their assessments of the girl's skill level. Darley and Gross (1983) interpret this result as evidence of confirmatory bias—subjects were influenced by the girl's background in their initial judgments, but their beliefs were evidently influenced even more strongly by the effect their initial hypotheses had on their interpretation of further evidence.

Certain types of evidence flows seem to be most conducive to confirmatory bias. Ambiguity of evidence is widely recognized to be an important mediating factor in both confirmatory bias and overconfidence (see, e.g., Gideon Keren 1987; and Dale Griffin and Tversky 1992). Keren (1988) notes the lack of confirmatory bias in visual perceptions, and concludes that confirmatory tendency depends on some degree of abstraction and the need for interpretation not present in simple visual tasks. Lord, Ross, and Lepper (1979, p. 2099) posit that when faced with complex and ambiguous evidence, we emphasize the strength and reliability of confirming evidence but the weaknesses and unreliability of disconfirming evidence. They also report an impression that may resound with those observing economists' reactions to behavioral evidence that might be damaging to habitual economics assumptions:

> 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.

The above passages hint at the role that selective scrutiny of evidence plays in confirmatory bias. Another form of "scrutiny-based" confirmatory bias is what I shall call *hypothesis-based filtering*. 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. Lord, Ross, and Lepper note a similar distinction in reflecting on the bias in their experiment: It is legitimate 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.

Even when each individual datum is

unambiguous, confirmatory bias can be generated when people must statistically assess correlations extended over time. Richard Nisbett and Ross (1980) argue that the inability to accurately perceive correlation is one of the most robust shortcomings in human reasoning, and people often imagine correlations between events when no such correlation exists.[22] Dennis Jennings, Teresa Amabile, and Ross (1982) argue that illusory correlation can play an important role in the confirmation of false 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.

C. *Other Biases*

I briefly outline a few more biases that might interest economists.[23] The first is *anchoring and adjustment.* Slovic and Sarah 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, pp. 1128) provide the following example:

> [S]ubjects 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. 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.*[24] Baruch Fischhoff (1975,

[22] Loren Chapman and Jean 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. Charles Stangor (1988) and David Hamilton and Terrence Rose (1980) also discuss the role of illusory correlation in the context of confirmatory-like phenomena. More generally, 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."

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

[24] For two excellent recent reviews of the hindsight bias, see Scott Hawkins and Reid Hastie (1990) and Jay Christensen-Szalanski and Cynthia 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

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, 3) military stalemate with a peace settlement, 4) military stalemate 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 each reported outcome, when compared to a control group not told any outcome, subjects' average ex post guesses of their hypothetical ex ante estimates were 15 percent higher than those of the control group. People don't sufficiently "subtract" information they currently have about an outcome in imagining what they would have thought without that information.[25]

A pervasive fact about human judgment is that people disproportionately *weight salient, memorable,* or *vivid* evidence even when they have better sources of information.[26] For instance, our assessment of a given city's crime rate 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 (1973) discuss, for example, how salience may distort clinicians' assessments of the relationship between severe depression and suicide. Incidents in which patients commit sui-

for the existence of the bias. They conclude that the bias is very real. (They also argue that the effects are "small.")

[25] 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 (Hawkins and Hastie 1990, p. 319).

[26] 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 Susan Fiske and Shelley Taylor (1991, chs. 5,7).

cide are much more likely to be remembered than are instances where patients do not commit suicide. This is likely to lead to an exaggerated assessment of the probability that depressed patients will commit suicide.

Finally, there is a mass of psychological research that finds people are prone toward 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).[27]

D. *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 virtually eliminate biases? These are reasonable conjectures, and such factors probably do on average moderate biases. But the conjectures do not appear to be nearly as valid as economists imagine. Kahneman and Tversky (1982a) and Tversky and Kahneman (1982), 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 very much does not support the strong versions of the experts-get-things-right and in-the-real-world-people-learn hypotheses.

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. In the context of overconfidence, for instance, Griffin and Tversky (1992) and Andrea Baumann, Raisa Deber, and Gail Thompson (1991) conclude that people who are aware of their own accuracy overall are overconfident on a case-by-case basis. 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: Kahneman and Tversky (1982b, p. 495) call such errors *errors of applications*, 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. 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.

As was demonstrated in the context

[27] See, e.g., Stuart Oskamp (1965), Jayashree Mahajan (1992), and Paul Paese and Maryellen 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 condition.

of confirmation bias discussed above, "learning" can even sometimes tend to *exacerbate* errors. 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 when predictability is low, experts are often more susceptible to overconfidence than are laypersons. Griffin and Tversky (1992, p. 430) provide illustrations:

> 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.

While a reasonable conjecture is that greater thoughtfulness and intelligently searching for patterns in the world would help douse biases, the quote specifically targets (economists take note . . .) as susceptible to overconfidence those "experts who have rich models of the system in question." Indeed, many authors have hypothesized the role of the reasoning process itself in exacerbating the confirmatory bias discussed above, which in turn leads to overconfidence (see, e.g., Craig Anderson, Lepper, and Ross 1980; Ross et al. 1977; and Timothy Wilson and Jonathan Schooler 1991). Wilson and Suzanne LaFleur (1995, pp. 23–24) conduct an experiment on the role of "reasoning" in strengthening confidence: Members of six sororities at the University of Virginia were asked at the beginning of the semester to predict their own future behaviors toward fellow sorority members. Each subject was asked to predict "yes" or "no" whether she would engage in each of six different behaviors—and to assess her confidence in her prediction. Randomly, some of the members were asked to list "why they might or might not perform each of the . . .behaviors . . . People were given a separate page on which to list their reasons for each behavior. They were told that the purpose of this task was to 'organize their thoughts,' that 'no one will actually read what you have written,' and that 'your reasons will be discarded.' They were urged to 'list as many reasons as you can think of, filling up this page if you can'." Other subjects were asked to make the same six predictions, but *not* asked to think of reasons for their behavior.

At the end of the semester, subjects were asked if they had actually performed each of the six activities. The results from this experiment showed that the act of reasoning increased subjects' overconfidence regarding their predictions of their own behavior. While those who reasoned about their predictions were roughly as confident as those who did not reason (reasoners predicted their own Yes/No predictions would be accurate 80 percent of the time, while the control group predicted their accuracy at 82 percent), the reasoners were in fact significantly *less* accurate than the control group in their predictions; reasoners' predictions were accurate 71 percent of the time, compared to 79 percent for the control group.

4. *Is "Maximizing Utility" the Right Model?*

The varied material of this section suggests that it may be wrong to con-

ceptualize some types of economic behavior in terms of an agent who maximizes a stable, coherent utility function. Indeed, some of the material here also calls into question some of the interpretations presented earlier. For instance, loss-averse behavior may often reflect a flawed rule of thumb employed because people misperceive their own long-run well-being, rather than a modified utility function as suggested in Section 2. By calling into question the utility-maximization interpretation of behavior given in Section 2, such material suggests greater difficulty in improving the behavioral realism of formal economics. On the other hand, the material at the end of the section on self-control problems provides a great opportunity to improve the realism of economics: Researchers have shown that a (relatively) simple multiple-self model of time-inconsistent discounting tractably modifies our familiar exponential model to yield a model that is manifestly more realistic behaviorally and surely has important economic consequences.

A. *Do We Know What Makes Us Happy?*

The research on biases reviewed in Section 3 indicates that people misjudge the probabilistic consequences of their decisions. But other research suggests that, even when they correctly perceive the *physical* consequences of their decisions, people systematically misperceive the well-being they derive from such outcomes. We often systematically mispredict our future experienced utility, even when those predictions rely only on accurate assessments of our past experienced utility (Kahneman 1994; and Kahneman, Peter Wakker, and Rakesh Sarin 1997). As Kahneman (1994, p. 21) puts it, "These considerations suggest an explicit distinction between two notions of utility. The *experienced utility* of an outcome is the measure of the hedonic experience of that outcome. . . . The *decision utility* of an outcome . . . is the weight assigned to that outcome in a decision." The realization that decision and experienced utility may be systematically different cuts to the core of our models of choice. It also cuts to the core of our methods of research, requiring us to formulate ways of inferring and eliciting preferences that go beyond a "revealed preference" method to attempt to infer people's hedonic experiences, through such methods as self reports of satisfaction and even psychological measurements.

How do people misperceive their utilities? One pattern is that we tend to underestimate how quickly and how fully we will adjust to changes, not foreseeing that our reference points will change. In a classic study, Philip Brickman, Dan Coates, and Ronnie Janoff-Bulman (1978) interviewed both lottery winners (with average winnings of about $479,545) and a control group; they found virtually no difference in rated happiness of lottery non-winners and winners. While such interview evidence is inconclusive, the researchers controlled for alternative explanations (such as selection bias or biased presentation by interviewers). Two effects seemed to explain why lottery winners would be less happy than the winners had presumably anticipated. First, mundane experiences become less satisfying by contrast to the "peak" experience of winning the lottery. Second, we become habituated to our circumstances: Along the lines of the material presented in Section 2, eventually the main carriers of utility become not the absolute levels of consumption, but departures from our (new) reference level.[28]

[28] Brickman, Coates, and Janoff-Bulman (1978) also found more equivocal evidence for these effects by interviewing 29 paraplegics and quadri-

People do not anticipate the degree of such adaptation, and hence exaggerate expected changes in utility caused by changes in their lives.[29] 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. And, 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.

Yet loss aversion often seems to be a judgmental bias: In decisions with significant long-run consequences, people should put less weight than they do on their initial experience of losses. Indeed, some researchers invoke loss aversion more as an irrational rule of thumb than as a rational utility function. Shlomo 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 et al. (1997) 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. In some more extreme examples of loss aversion it is hard to believe that the "transition utility" can rationally rank high relative to long-term utility. For instance, Thaler (1980) compared subjects' willingness to pay for a cure for a disease that leads to a quick and painless death with probability .001 versus the minimum price you would accept to voluntarily subject yourself to the same disease. Subjects often required an order of magnitude more money to expose themselves to the disease than they would pay for a cure. People charge heavy premiums for losses relative to their status quo, even when it is hard to imagine that any experienced "transition utility" is significant relative to long-term utility consequences—here, whether or not you live beyond a week.

Another example of how people misperceive utility consequences of their choices is Richard Herrnstein and Prelec's (1992b) theory of *melioration*. Based on a mass of evidence gathered from 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 a current choice has on the utilities of later choices. Say you eat at one of two restaurants 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:

---

plegics about their happiness before their accidents, their current happiness, and their expected future happiness. 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. However, accident victims put more emphasis on, and took 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.

[29] In the simple model of reference-point adjustment discussed in Section 2, this can be translated as saying that people systematically underestimate the parameter α.

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, to take into account internalities, but the evidence suggests that we take too little account of the global utility effects of individual choices. Herrnstein and Prelec (1992b, pp. 257–58) argue that even when people *do* seem aware of the shape of their global utility functions, they may not properly maximize those preferences because they take an overly "piecemeal" approach.[30]

A major way people predict utility they will derive from future experiences is to recollect utility from comparable past experiences. While we might presume that people accurately recollect their utility from familiar experiences, research on the endowment effect hints that this presumption may not be accurate: If we systematically misperceive the long-run consequences of giving up minor consumer items such as mugs, we may not have learned to assess correctly the utility consequences of even 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, by collecting periodic hedonic reports by subjects of their current well-being. In evaluating the overall utility from such an extended episode, one must formulate criteria for adding up flows of experienced well-being. Kahneman (1994) posits that an uncontroversial 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. Kahneman (1994) argues that experiments suggest biases in how people's own retrospective evaluations of episodes compare to their experienced well-being. First, in evaluating past episodes, people tend to remember the extremes of pain and pleasure more than the average. Second, when an "episode" is well-defined (e.g., a vacation), people tend to put too much weight on the end of the episode (e.g., the last night of the vacation) in assessing their overall experience of the episode. Finally, we tend to neglect the duration of an episode. In assessing the dissatisfaction of an extremely unpleasant medical procedure (colonoscopy), for instance, patients seem to all but neglect the duration of the procedure—which ranged from 4 to 69 minutes. Of course, one must carefully consider the pain and pleasure associated with an episode before and after the actual episode; anticipation and recollection of pain, for instance, are clearly important

[30] 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 Peter Zych (1994) for an experimental exploration of this distinction.

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. Such an interpretation of most of the experimental evidence, however, seems tenuous.

The fact that we don't always correctly predict experienced utility is obviously important for welfare implications of choice, and it prescribes caution in reliance on revealed-preference-based welfare economics. But there may be important behavioral implications of a related phenomenon whereby people misperceive their future *behavior*. Loewenstein and Daniel Adler (1995) performed an experiment based on the endowment-effect experiments of Kahneman, Knetsch, and Thaler (1991) discussed in Section 2. 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 their minimal selling prices were elicited. 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, and behaved significantly differently than they had predicted about themselves *moments* earlier. (Such a procedure underestimates the true degree of misperception, because people don't like to contradict recently expressed predictions of their own behavior. Indeed, subjects who had made no prediction averaged a selling price of \$5.62.)

C. *Elicitation Effects*

People often lack stable preferences that are robust to different ways of eliciting those preferences.[31] The most prominent set of research that points to such an interpretation of choice behavior concerns *framing effects*: Two logically equivalent (but not *transparently* equivalent) statements of a problem lead decision makers to choose different options. An important and predictable influence of framing on choice relates 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, pp. S254–55) give the following example of framing effects, taken from a study of medical decisions by Barbara McNeil et al. (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,

[31] 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., Diamond and Jerry Hausman (1994).

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, but similar framing effects were found in choices over lotteries with small monetary stakes, and Tversky and Kahneman (1986) cite some important real-world examples of framing effects. For instance, people react differently to firms charging different prices for different services (or the same service at different times) depending on whether the lower price is called a discount or the higher price is called a surcharge. Similarly, Thomas Schelling (1981) noticed huge differences in his students' attitudes toward tax deductions for children depending on how the deductions were framed. Money illusion provides perhaps the best example of the importance of framing effects for economics. Kahneman, Knetsch, and Thaler (1986a) 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. A nominal wage increase of 5 percent in a period of 12 percent inflation offends people's sense of fairness less than a 7 percent 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 (Shafir, Diamond, and Tversky (1997).

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 to some extent are a topic for 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.

Related phenomena even more strongly call into doubt the view that choices reflect stable, well-defined preferences. *Preference reversals* have been studied widely by economists and psychologists over the years: When confronted with certain pairs of gambles with roughly the same expected value, people often choose one of the pair over the other, while *pricing* the other more highly. To use an example from Tversky and Thaler (1990), consider an H bet that with 8/9 chance yields \$4 and with 1/9 chance yields \$0, and an L bet with a 1/9 chance to win \$40 and 8/9 chance of \$0. Most subjects choose the H bet over the L bet when asked to choose between the two. But when asked to state the lowest price at which they would be willing to sell each gamble, most subjects put a higher price on the L bet. More generally, people choose bets with a high chance of winning small amounts, but put a higher price on bets with a low chance of winning big amounts; economic theory predicts these two different elicitation proce-

dures should yield the same preferences.

Itamar Simonson and Tversky (1992) provide examples of *context effects*, where the addition of a new option to a menu 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. (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 any information revealed by the choice sets.) As another example, Simonson and Tversky (1992) ran an experiment that illustrates that elicited subjects' preference for an elegant Cross pen versus receiving $6. While only 36 percent of subjects choosing only between the Cross pen and the $6 chose the Cross pen, 46 percent of subjects who were also given the choice of a less attractive pen chose the Cross pen. In both these examples, the addition of an option that compared unfavorably (as more expensive or lower quality) to an existing option enhanced the perceived attractiveness of the existing option.

While people are often unaware that the menu of choices influences their decisions, Simonson and Tversky note that 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. Indeed, such findings suggest an alternative to the utility-maximization framework that may help explain framing effects, preference reversals, and context effects: People may make choices in part by asking themselves whether they have a "reason" to choose one option over another (Shafir, Simonson, and Tversky 1993).

D. *Time-Variant Preferences*

People have a taste for immediate gratification. We procrastinate on tasks such as mowing the lawn that involve immediate costs and delayed rewards and do soon things such as seeing a movie that involve immediate rewards and delayed costs. 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.

Casual observation, introspection, and psychological research all suggest that the assumption of time-consistency is importantly wrong. Our short term tendency to pursue immediate gratification is inconsistent with our long term preferences. 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. Kris 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." 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.[32]

Hence, a person's preferences today over her future delays in rewards are different than her future preferences over those same delays, so that preferences are *not* time consistent. Formal models of such time-*variant* preferences have been developed.[33] Edmund Phelps and Robert Pollak (1968) capture the taste for immediate gratification with a simple 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, where both β and δ lie between 0 and 1:

For all t,

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

The parameter δ determines how "time-consistently patient" a person is, just as in exponential discounting. If $\beta = 1$, then these preferences are simply exponential discounting. But for $\beta < 1$, these preferences capture in a parsimonious way the type of time-inconsistent preferences so widely observed. To see how these preferences capture the preference for immediate gratification, suppose that you had a choice between doing ten hours of an unpleasant task on April 14, versus spending eleven hours to complete the same task on April 15. Assume that your instantaneous disutility from doing work is simply the number of hours of work— $u_t(10) = -10$ and $u_t(11) = -11$ for all t. Suppose that $\delta = 1$, but that $\beta = .8$ for a one-day delay: You are willing to suffer a given loss in utility tomorrow for a gain in utility today that is 80 percent as large.

Suppose that April 14 has arrived and you are considering whether or not to work. You can experience a disutility of −10 by working today, or experience a discounted utility of $.8 \cdot (-11) = -8.8$ by delaying the work until tomorrow. You will, therefore, delay work. Contrast this with what your decision would be if, instead of choosing when to work on April 14, you are told by your boss that you must decide on February 1. Because from February 1 you discount *both* dates by β, you will choose to work 10 hours on April 14 rather than 11 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 February 1 is different than your choice on April 14. 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 14. This example seems well-calibrated: On April 14, most of us are apt to put off the work until April 15, even if it means a little more work. Absent a substantive difference between the two dates, virtually no one would choose the delay if asked on February 1.

To examine dynamic choice given time-variant preferences given these preferences, 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

[32] These numbers are calculated from the data presented by Kirby and Herrnstein (1995, p. 85–86). Other psychological research showing preferences are not time-consistent includes Shin-Ho Chung and Herrnstein (1967), George Ainslie (1991), Ainslie and Herrnstein (1981), Thaler (1981), and Loewenstein and Prelec (1992).

[33] For economics papers on time-inconsistent discounting, see, e.g., Robert Strotz (1955), Steven Goldman (1979, 1980), Schelling (1978), Thaler and Hersh Shefrin (1981), David Laibson (1994, 1997), and O'Donoghue and Rabin (1997a, 1997b).

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 that are stable over time—is a radical departure from the utility-maximization framework. But because this conceptualization of intertemporal choice uses a familiar tool—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 may scheme to manipulate 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 14; you commit to do the task on April 15; or you *wait* until April 14 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 14 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 that you won't overeat 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.

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. You may have expectations about your propensity to misbehave, or you may naïvely believe that your preferences in the future will match your current preferences. If today you prefer not to overeat tomorrow, you may naïvely believe that you will feel the same way when facing an enticing bowl of ice cream tomorrow. If on February 1 you prefer less work on April 14 to more work on April 15, you may believe you'll feel the same way in April.

Strotz (1955) labels people who are fully aware of their future self-control problems as *sophisticated*, and people who are fully unaware that they will have a self-control problem as *naïve*. While some degree of sophistication is implied by the existence of some of the self-commitment devices illustrated above, 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 earlier, 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 prefer-

ences *won't* accord with your current preferences. For example, people may repeatedly not have the "will power" to forego tempting foods or quit smoking while predicting that *tomorrow* they will have this will power. While behavioral evidence that calibrates the degree of sophistication seems sparse, Loewenstein (1996, pp. 281–82) reaches the conclusion that people may be naïve indirectly from psychological findings such as the evidence of people mispredicting changes in utility.

Whether they are sophisticated or naïve, people's time-inconsistent propensity for immediate gratification is important in a variety of economic realms. As investigated by several researchers (see, e.g., Thaler and Shefrin 1981; and Laibson 1997), such preferences may be important to savings behavior because the benefits of current consumption are immediate, whereas the increased future consumption that saving allows is delayed. Self-control problems are also clearly important in the demand for addictive goods and fatty foods. Similarly, the role of self-control in purchasing decisions is well known among marketing experts (Stephen Hoch and Loewenstein 1991). Naughty goods are sold in small packages because people tend to avoid large packages of such goods to prevent overconsumption.

5. *Conclusion*

Over the years, economists have proffered many reasons for downplaying the relevance of behavioral research challenging our habitual assumptions. Claims abound that evidence inconsistent with our traditional model of human behavior can be neglected because the evidence derives from observations of people insufficiently motivated to behave themselves according to economic assumptions, or because it fails to bear sufficiently great burdens of proof, or because the implied behavior is unlikely to matter in the types of (market) settings that economists care about. Because behavioral research is so often assessed in light of such arguments, it is common when presenting psychological findings to discuss broad methodological objections and attempt to rebut them.

I refrain from doing so. It is my strong impression that many of the arguments invoked against the reality or relevance of behavioral research derive from unfamiliarity with the details of this research. Hence, my hope and guess is that as economists become more familiar with this research, such arguments will dissipate. And as the aggressive uncuriosity shown in the past toward behavioral research continues to diminish, we can look forward to focusing entirely on its substance.

While none of the broad-stroke arguments for inattention to psychological research are compelling, obviously not all psychological research will be both confirmed by field data and proven to be of great economic importance. Indeed, abandoning the view that hypotheses departing from rationality, self-interest, or other habitual assumptions are methodologically illicit can free us to evaluate these hypotheses with the same rigorous standards that our discipline, at its best, applies elsewhere. We can confront plausible hypotheses about human behavior with both healthy skepticism and genuine curiosity, empirically test their validity, and carefully draw out their economic implications. And, as we apply these rigorous standards, we can then begin to treat claims that (say) investors irrationally infer too much from short-term performance, or that employees feel resentful when mistreated, as presump-

tively plausible and presumptively relevant hypotheses worth keeping in mind in our economic analysis.

REFERENCES

ADAMS, J. STACY. "Wage Inequities, Productivity and Work Quality," *Ind. Relat.*, Oct. 1963, *3*, pp. 9–16.

AINSLIE, GEORGE W. "Derivation of 'Rational' Economic Behavior from Hyperbolic Discount Curves," *Amer. Econ. Rev.*, May 1991, *81*(2), pp. 334–40.

AINSLIE, GEORGE W. AND HERRNSTEIN, RICHARD J. "Preference Reversal and Delayed Reinforcement," *Animal Learning & Behavior*, Nov. 1981, *9*(4), pp. 476–82.

AKERLOF, GEORGE A. "Labor Contracts as Partial Gift Exchange," *Quart. J. Econ.*, Nov. 1982, *97*(4), pp. 543–69.

———. "Gift Exchange and Efficiency-Wage Theory: Four Views," *Amer. Econ. Rev.*, May 1984, *74*(2), pp. 79–83.

AKERLOF, GEORGE A. AND YELLEN, JANET L. "The Fair Wage-Effort Hypothesis and Unemployment," *Quart. J. Econ.*, May 1990, *105*(2), pp. 255–83.

ANDERSON, CRAIG A.; LEPPER, MARK R. AND ROSS, LEE. "Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information," *J. Personality & Social Psychology*, Dec. 1980, *39*(1–6), pp. 1037–49.

ANDREONI, JAMES. "Cooperation in Public-Goods Experiments: Kindness or Confusion?" *Amer. Econ. Rev.*, Sept. 1995, *85*(4), pp. 891–904.

ANDREONI, JAMES AND MILLER, JOHN H. "Giving According to GARP: An Experimental Study of Rationality and Altruism." Working Paper No. 9601. Social Systems Research Institute, U. Wisconsin at Madison, 1996.

BAUMANN, ANDREA O.; DEBER, RAISA B. AND THOMPSON, GAIL G. "Overconfidence Among Physicians and Nurses: The 'Micro-Certainty, Macro-Uncertainty' Phenomenon," *Social Science & Medicine*, 1991, *32*(2), pp. 167–74.

BENARTZI, SHLOMO AND THALER, RICHARD H. "Myopic Loss Aversion and the Equity Premium Puzzle," *Quart. J. Econ.*, Feb. 1995, *110*(1), pp. 73–92.

BERSCHEID, ELLEN; BOYE, DAVID AND WALSTER, ELAINE. "Retaliation as a Means of Restoring Equity," *J. Personality & Social Psychology*, 1968, *10*(4), pp. 370–76.

BLOUNT, SALLY. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior & Human Decision Processes*, Aug. 1995, *63*(2), pp. 131–44.

BOWMAN, DAVID; MINEHART, DEBORAH AND RABIN, MATTHEW. "Loss Aversion in a Consumption-Savings Model." Mimeo, U. California at Berkeley, 1997.

BRICKMAN, PHILIP; COATES, DAN AND JANOFF-BULMAN, RONNIE. "Lottery Winners and Accident Victims: Is Happiness Relative?" *J. Personality & Social Psychology*, Aug. 1978, *36*(8), pp. 917–27.

BRUNER, JEROME S. AND POTTER, MARY C. "Interference in Visual Recognition," *Science*, Apr. 24, 1964, *144*(3617), pp. 424–25.

CAMERER, COLIN F. "Individual Decision Making," in *Handbook of experimental economics*. Eds.: JOHN KAGEL AND ALVIN E. ROTH. Princeton: Princeton U. Press, 1995, pp. 587–703.

CAMERER, COLIN ET AL. "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quart. J. Econ.*, May 1997, *112*(2), pp. 407–41.

CAMERER, COLIN F. AND THALER, RICHARD H. "Ultimatums, Dictators and Manners," *J. Econ. Perspectives*, Spring 1995, *9*(2), pp. 209–19.

CHAPMAN, LOREN J. AND CHAPMAN, JEAN P. "Genesis of Popular But Erroneous Psychodiagnostic Observations," *J. Abnormal Psychology*, 1967, *72*(3), pp. 193–204.

———. "Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs," *J. Abnormal Psychology*, 1969, *74*(3), pp. 271–80.

———. "Test Results Are What You Think They Are," *Psychology Today*, Nov. 1971, *5*(6), pp. 18–22.

CHARNESS, GARY. "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation." Mimeo, U. California at Berkeley, 1996.

CHRISTENSEN-SZALANSKI, JAY J. AND WILLHAM, CYNTHIA. "The Hindsight Bias: A Meta-Analysis," *Organizational Behavior & Human Decision Processes*, Feb. 1991, *48*(1), pp. 147–68.

CHUNG, SHIN-HO AND HERRNSTEIN, RICHARD J. "Choice and Delay of Reinforcement," *J. Experimental Analysis of Behavior*, 1967, *10*(1), pp. 67–74.

CONLISK, JOHN. "Why Bounded Rationality?" *J. Econ. Lit.*, June 1996, *34*(2), pp. 669–700.

CROSON, RACHEL T. A. "Expectations in Voluntary Contribution Mechanisms." Working Paper No. 95–03–02, Department of Operations and Information Management, The Wharton School, U. Penn., 1995.

DARLEY, JOHN M. AND GROSS, PAGET H. "A Hypothesis-Confirming Bias in Labeling Effects," *J. Personality & Social Psychology*, Jan. 1983, *44*(1), pp. 20–33.

DAWES, ROBYN M. AND THALER, RICHARD H. "Anomalies: Cooperation," *J. Econ. Perspectives*, Summer 1988, *2*(3), pp. 187–97.

DIAMOND, PETER A. AND HAUSMAN, JERRY A. "Contingent Valuation: Is Some Number Better than No Number?" *J. Econ. Perspectives*, Fall 1994, *8*(4), pp. 45–64.

DUESENBERRY, JAMES S. *Income, saving and the theory of consumer behavior*. Cambridge: Harvard U. Press, 1949.

EASTERLIN, RICHARD A. "Does Economic Growth Improve the Human Lot? Some Em-

pirical Evidence," in *Nations and households in economic growth: Essays in honor of Moses Abramovitz.* Eds.: PAUL A. DAVID AND MELVIN W. REDER. New York: Academic Press, 1974, pp. 89–125.

EPSTEIN, LARRY G. AND ZIN, STANLEY E. "'First-Order' Risk Aversion and the Equity Premium Puzzle," *J. Monet. Econ.*, Dec. 1990, *26*(3), pp. 387–407.

FEHR, ERNST AND FALK, ARMIN. "Wage Rigidities in a Competitive Incomplete Contract Market: An Experimental Investigation." Mimeo, Institute for Empirical Research in Economics, U. Zurich, 1996.

FEHR, ERNST; KIRCHLER, ERICH AND WEICHBOLD, ANDREAS. "When Social Forces Remove the Impact of Competition: An Experimental Investigation." Mimeo, Department of Economics, U. Technology, Vienna, 1994.

FEHR, ERNST; KIRCHSTEIGER, GEORG AND RIEDL, ARNO. "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quart. J. Econ.*, May 1993, *108*(2), pp. 437–59.

FEHR, ERNST AND ZYCH, PETER. "The Power of Temptation: Irrationally Myopic Excess Consumption in an Addiction Experiment." Mimeo, Institute for Empirical Research in Economics, U. Zurich, 1994.

FISCHHOFF, BARUCH. "Hindsight Is Not Equal To Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty," *J. Experimental Psychology: Human Perception & Performance*, Aug. 1975, *104*(1), pp. 288–99.

FISCHHOFF, BARUCH; SLOVIC, PAUL AND LICHTENSTEIN, SARAH. "Knowing with Certainty: The Appropriateness of Extreme Confidence," *J. Experimental Psychology: Human Perception & Performance*, Nov. 1977, 3(4), pp. 552–64.

FISKE, SUSAN T. AND TAYLOR, SHELLEY E. *Social cognition.* New York: McGraw-Hill, 1991.

FRANCIOSI, ROBERT ET AL. "Fairness: Effect on Temporary and Equilibrium Prices in Posted-Offer Markets," *Econ. J.*, July 1995, *105*(431), pp. 938–50.

FRANK, ROBERT H. *Choosing the right pond: Human behavior and the quest for status.* Oxford: Oxford U. Press, 1985.

———. "Frames of Reference and the Quality of Life," *Amer. Econ. Rev.*, May 1989, *79*(2), pp. 80–85.

———. "Local Status, Fairness, and Wage Compression Revisited." Mimeo, Department of Economics, Cornell U., 1994.

FRANK, ROBERT H. AND HUTCHENS, ROBERT M. "Wages, Seniority, and the Demand for Rising Consumption Profiles," *J. Econ. Behav. Org.*, Aug. 1993, *21*(3), pp. 251–76.

GEANAKOPLOS, JOHN; PEARCE, DAVID AND STACCHETTI, ENNIO. "Psychological Games and Sequential Rationality," *Games Econ. Behav.*, Mar. 1989, *1*(1), pp. 60–79.

GILOVICH, THOMAS; VALLONE, ROBERT AND TVERSKY, AMOS. "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, July 1985, *17*(3), pp. 295–314.

GOLDMAN, STEVEN M. "Intertemporally Inconsistent Preferences and the Rate of Consumption," *Econometrica*, May 1979, *47*(3), pp. 621–26.

———. "Consistent Plans," *Rev. Econ. Stud.*, Apr. 1980, *47*(3), pp. 533–37.

GORANSON, RICHARD E. AND BERKOWITZ, LEONARD. "Reciprocity and Responsibility Reactions to Prior Help," *J. Personality & Social Psychology*, 1966, 3(2), pp. 227–32.

GRIFFIN, DALE AND TVERSKY, AMOS. "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, July 1992, *24*(3), pp. 411–35.

GÜTH, WERNER; SCHMITTBERGER, ROLF AND SCHWARZE, BERND. "An Experimental Analysis of Ultimatum Bargaining," *J. Econ. Behav. Organ.*, Dec. 1982, 3(4), pp. 367–88.

GÜTH, WERNER AND TIETZ, REINHARD. "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results," *J. Econ. Psych.*, Sept. 1990, *11*(3), pp. 417–49.

HAMILTON, DAVID L. AND ROSE, TERRENCE L. "Illusory Correlation and the Maintenance of Stereotypic Beliefs," *J. Personality & Social Psychology*, Nov. 1980, *39*(5), pp. 832–45.

HARLESS, DALE W. AND CAMERER, COLIN F. "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, Nov. 1994, *62*(6), pp. 1251–89.

HARTMAN, RAYMOND S.; DOANE, MICHAEL AND WOO, CHI-KEUNG. "Consumer Rationality and the Status Quo," *Quart. J. Econ.*, Feb. 1991, *106*(1), pp. 141–62.

HAWKINS, SCOTT A. AND HASTIE, REID. "Hindsight: Biased Judgments of Past Events After the Outcomes Are Known," *Psychology Bulletin*, May 1990, *107*(3), pp. 311–27.

HELSON, HARRY. *Adaptation level theory: An experimental and systematic approach to behavior.* New York: Harper & Row, 1964.

HERRNSTEIN, RICHARD J. AND PRELEC, DRAZEN. "Melioration," in. *Choice over time.* Eds.: GEORGE LOEWENSTEIN AND JOHN ELSTER. New York: Russell Sage Foundation, 1992, pp. 235–63.

HOCH, STEPHEN J. AND LOEWENSTEIN, GEORGE F. "Time-Inconsistent Preferences and Consumer Self-Control," *J. Cons. Res.*, Mar. 1991, *17*(4), pp. 492–507.

JENNINGS, DENNIS, L.; AMABILE, TERESA M. AND ROSS, LEE. "Informal Covariation Assessment: Data-based versus Theory-based Judgments," in DANIEL KAHNEMAN, PAUL SLOVIC, AND AMOS TVERSKY, eds. 1982, pp. 211–30.

KAGEL, JOHN H.; KIM, CHUNG AND MOSER, DONALD. "'Fairness' in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs," *Games Econ. Behav.*, Mar. 1996, *13*(1), pp. 100–10.

KAHN, ARNOLD AND TICE, THOMAS E. "Returning a Favor and Retaliating Harm: The Effects

of Stated Intentions and Actual Behavior," *J. Experimental Social Psychology*, Jan. 1973, *9*(1), pp. 43–56.

KAHNEMAN, DANIEL. "New Challenges to the Rationality Assumption." *J. Inst. Theoretical Econ.*, Mar. 1994, *150*(1), pp. 18–36.

KAHNEMAN, DANIEL; KNETSCH, JACK L. AND THALER, RICHARD H. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *Amer. Econ. Rev.*, Sept. 1986a, *76*(4), pp. 728–41.

———. "Fairness and the Assumptions of Economics," *J. Bus.*, Oct. 1986b, *59*(4), pp. S285–300.

———. "Experimental Tests of the Endowment Effect and the Coase Theorem," *J. Polit. Econ.*, Dec. 1990, *98*(6), pp. 1325–48.

———. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *J. Econ. Perspectives*, Winter 1991, *5*(1), pp. 193–206.

KAHNEMAN, DANIEL; SLOVIC, PAUL AND TVERSKY, AMOS, eds. *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge U. Press, 1982.

KAHNEMAN, DANIEL AND TVERSKY, AMOS. "On the Psychology of Prediction," *Psychology Rev.*, July 1973, *80*(4), pp. 237–51.

———. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Mar. 1979, *47*(2), pp. 263–91.

———. "Subjective Probability: A Judgment of Representativeness," 1982a in DANIEL KAHNEMAN, PAUL SLOVIC, AND AMOS TVERSKY, eds. 1982, pp. 32–47.

———. "On the Study of Statistical Intuitions," 1982b in DANIEL KAHNEMAN, PAUL SLOVIC, AND AMOS TVERSKY, eds. 1982, pp. 493–508.

KAHNEMAN, DANIEL; WAKKER, PETER P. AND SARIN, RAKESH. "Back to Bentham? Explorations of Experienced Utility," *Quart. J. Econ.*, May 1997, *112*(2), pp. 375–406.

KEREN, GIDEON. "Facing Uncertainty in the Game of Bridge: A Calibration Study," *Organizational Behavior & Human Decision Processes*, Feb. 1987, *39*(1), pp. 98–114.

———. "On the Ability of Monitoring Non-Veridical Perceptions and Uncertain Knowledge: Some Calibration Studies," *Acta Psychologica*, May 1988, *67*(2), pp. 95–119.

KESER, CLAUDIA. "Voluntary Contributions to a Public Good When Partial Contribution Is a Dominant Strategy," *Econ. Letters*, Mar. 1996, *50*(3), pp. 359–66.

KIRBY, KRIS N. AND HERRNSTEIN, RICHARD J. "Preference Reversals Due to Myopic Discounting of Delayed Reward," *Psychology Science*, Mar. 1995, *6*(2), pp. 83–89.

KNETSCH, JACK. "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *Amer. Econ. Rev.*, Dec. 1989, *79*(5), pp. 1277–84.

KNETSCH, JACK L. AND SINDEN, JOHN A. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," *Quart. J. Econ.*, Aug. 1984, *99*(3), pp. 507–21.

KREBS, DENNIS L. "Altruism—An Examination of the Concept and a Review of the Literature," *Psychology Bulletin*, 1970, *73*(4), pp. 258–302.

———. "Prosocial Behavior, Equity, and Justice," in *Equity and justice in social behavior.* Eds.:JERALD GREENBERG AND RONALD L. COHEN. New York: Academic Press, 1982, pp. 261–308.

LAIBSON, DAVID. "Essays in Hyperbolic Discounting." Mimeo, M.I.T., 1994.

———. "Golden Eggs and Hyperbolic Discounting," *Quart. J. Econ.*, May 1997, *112*(2), pp. 443–78.

LOEWENSTEIN, GEORGE. "Out of Control: Visceral Influences on Behavior," *Organizational Behavior & Human Decision Processes*, Mar. 1996, *65*(3), pp. 272–92.

LOEWENSTEIN, GEORGE AND ADLER, DANIEL. "A Bias in the Prediction of Tastes," *Econ. J.*, July 1995, *105*(431), pp. 929–37.

LOEWENSTEIN, GEORGE AND PRELEC, DRAZEN. "Anomalies in Intertemporal Choice: Evidence and an Interpretation," *Quart. J. Econ.*, May 1992, *107*(2), pp. 573–97.

LOEWENSTEIN, GEORGE AND SICHERMAN, NACHUM. "Do Workers Prefer Increasing Wage Profiles?" *J. Lab. Econ.*, Jan. 1991, *9*(1), pp. 67–84.

LORD, CHARLES. G.; ROSS, LEE AND LEPPER, MARK R. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *J. Personality & Social Psychology*, Nov. 1979, *37*(11), pp. 2098–2109.

MAHAJAN, JAYASHREE. "The Overconfidence Effect in Marketing Management Predictions," *J. Marketing Res.*, Aug. 1992, *29*(3), pp. 329–42.

MCNEIL, BARBARA J. ET AL. "On the Elicitation of Preferences For Alternative Therapies," *New England J. Medicine*, May 27, 1982, *306*, pp. 1259–62.

MEHRA, RAJNISH AND PRESCOTT, EDWARD C. "The Equity Premium: A Puzzle," *J. Monet. Econ.*, Mar. 1985, *15*(2), pp. 145–61.

NISBETT, RICHARD E. AND ROSS, LEE. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall, 1980.

O'DONOGHUE, TED AND RABIN, MATTHEW. "Doing It Now or Later." Working Paper No. 97–253. U. California at Berkeley, 1997a.

———. "Incentives for Procrastinators." Discussion Paper No. 1181. Northwestern U., CMSEMS, 1997b.

OSKAMP, STUART. "Overconfidence in Case-Study Judgments." *J. Consulting Psychology*, 1965, *29*(3), pp. 261–65.

PAESE, PAUL W. AND KINNALY, MARYELLEN. "Peer Input and Revised Judgment: Exploring the Effects of (Un)Biased Confidence," *J. Applied Social Psychology*, Dec. 1993, *23*(23), pp. 1989–2011.

PERKINS, DAVID N. *The mind's best work*. Cambridge: Harvard U. Press, 1981.

PHELPS, EDMUND S. AND POLLAK, ROBERT A. "On Second-Best National Saving and Game-Equilibrium Growth," *Rev. Econ. Stud.*, Apr. 1968, *35*(2), pp. 185–99.

PLOUS, SCOTT. "Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer?" *J. Applied Social Psychology*, July 1991, *21*(13), pp. 1058–82.

RABIN, MATTHEW. "Incorporating Fairness into Game Theory and Economics," *Amer. Econ. Rev.*, Dec. 1993, *83*(5), pp. 1281–1302.

———. "Risk Aversion and Diminishing Marginal Utility: A Calibration Theorem." Mimeo, U. California at Berkeley, 1997.

RABIN, MATTHEW AND SCHRAG, JOEL. "First Impressions Matter: A Model of Confirmatory Bias." Working Paper No. 97–250. U. California at Berkeley, 1997.

ROSS, LEE ET AL. "Social Explanation and Social Expectation: Effects of Real and Hypothetical Explanations on Subjective Likelihood," *J. Personality & Social Psychology*, Nov. 1977, *35*(11), pp. 817–29.

ROTH, ALVIN E. AND MURNIGHAN, J. KEITH. "The Role of Information in Bargaining: An Experimental Study," *Econometrica*, Sept. 1982, *50*(5), pp. 1123–42.

RYDER, HARL E., JR. AND HEAL, GEOFFREY M. "Optimal Growth with Intertemporally Dependent Preferences," *Rev. Econ. Stud.*, Jan. 1973, *40*(1), pp. 1–33.

SALLY, DAVID. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society*, Jan. 1995, *7*(1), pp. 58–92.

SAMUELSON, WILLIAM AND ZECKHAUSER, RICHARD. "Status Quo Bias in Decision Making," *J. Risk Uncertainty*, Mar. 1988, *1*(1), pp. 7–59.

SCHELLING, THOMAS C. "Egonomics, or the Art of Self-Management," *Amer. Econ. Rev.*, May 1978, *68*(2), pp. 290–94.

———. "Economic Reasoning and the Ethics of Policy," *Public Interest*, Spring 1981, *63*, pp. 37–61.

SEGAL, UZI AND SPIVAK, AVIA. "First Order versus Second Order Risk Aversion," *J. Econ. Theory*, June 1990, *51*(1), pp. 111–25.

SHAFIR, ELDAR; DIAMOND, PETER AND TVERSKY, AMOS. "Money Illusion," *Quart. J. Econ.*, May 1997, *112*(2), pp. 341–74.

SHAFIR, ELDAR; SIMONSON, ITAMAR AND TVERSKY, AMOS. "Reason-Based Choice," *Cognition*, Special Issue: Reasoning and Decision Making. Oct.-Nov. 1993, *49*(1–2), pp. 11–36.

SHAFIR, ELDAR AND TVERSKY, AMOS. "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice," *Cognitive Psychology*, Oct. 1992, *24*(4), pp. 449–74.

SHEA, JOHN. "Union Contracts and the Life-Cycle/Permanent-Income Hypothesis," *Amer. Econ. Rev.*, Mar. 1995a, *85*(1), pp. 186–200.

———. "Myopia, Liquidity Constraints, and Aggregate Consumption: A Simple Test," *J. Money, Credit, Banking*, Aug. 1995b, *27*(3), pp. 798–805.

SIMONSON, ITAMAR AND TVERSKY, AMOS. "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *J. Marketing Res.*, Aug. 1992, *29*(3), pp. 281–95.

SLOVIC, PAUL AND LICHTENSTEIN, SARAH. "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior & Human Performance*, Nov. 1971, *6*(6), pp. 649–744.

SMITH, ADAM. *An inquiry into the nature and causes of the wealth of nations*. Reprinted ROY H. CAMPBELL AND ANDREW S. SKINNER, eds. Oxford: Clarendon Press, [1776] 1976.

STANGOR, CHARLES. "Stereotype Accessibility and Information Processing," *Personality & Social Psychology Bulletin*, Dec. 1988, *14*(4), pp. 694–708.

STROTZ, ROBERT H. "Myopia and Inconsistency in Dynamic Utility Maximization," *Rev. Econ. Stud.*, 1955, *23*(3), pp. 165–80.

THALER, RICHARD H. "Toward a Positive Theory of Consumer Choice," *J. Econ. Behav. Organ.*, Mar. 1980, *1*(1), pp. 39–60.

———. "Some Empirical Evidence of Dynamic Inconsistency," *Econ. Letters*, 1981, *81*, pp. 201–07.

———. "Mental Accounting and Consumer Choice," *Marketing Sci.*, Summer 1985, *4*(3), pp. 199–214.

——— "Anomalies: The Ultimatum Game," *J. Econ. Perspectives*, Fall 1988, *2*(4), pp. 195–206.

THALER, RICHARD H. AND SHEFRIN, HERSH M. "An Economic Theory of Self-Control," *J. Polit. Econ.*, Apr. 1981, *89*(2), pp. 392–406.

TRAIN, KENNETH E.; MCFADDEN, DANIEL L. AND GOETT, ANDREW A. "Consumer Attitudes and Voluntary Rate Schedules for Public Utilities," *Rev. Econ. Statist.*, Aug. 1987, *69*(3), pp. 383–91.

TVERSKY, AMOS AND GILOVICH, THOMAS. "The Cold Facts about the 'Hot Hand' in Basketball," *Chance*, 1989a, *2*(1), pp. 16–21.

———. "The Hot Hand: Statistical Reality or Cognitive Illusion," *Chance*, 1989b, *2*(4), pp. 31–34.

TVERSKY, AMOS AND KAHNEMAN, DANIEL. "Belief in the Law of Small Numbers," *Psychology Bulletin*, Aug. 1971, *76*(2), pp. 105–10.

———. "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, Sept. 1973, *5*(2), pp. 207–32.

———. "Judgment Under Uncertainty: Heuristics and Biases," *Science*, Sept. 1974, *185*(4157), pp. 1124–31.

———. "Judgments of and by Representativeness," in DANIEL KAHNEMAN, PAUL SLOVIC, AND AMOS TVERSKY, eds. 1982, pp. 84–98.

———. "Rational Choice and the Framing of Decisions," *J. Bus.*, Oct. 1986, *59*(4, Part 2), pp. S251–78.

———. "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quart. J. Econ.*, Nov. 1991, *106*(4), pp. 1039–61.

TVERSKY, AMOS AND THALER, RICHARD H. "Anomalies: Preference Reversals," *J. Econ. Perspectives*, Spring 1990, *4*(2), pp. 201–11.

WILSON, TIMOTHY D. AND LAFLEUR, SUZANNE J. "Knowing What You'll Do: Effects of Analyzing Reasons on Self-Prediction," *J. Personality & Social Psychology*, Jan. 1995, *68*(1), pp. 21–35.

WILSON, TIMOTHY D. AND SCHOOLER, JONATHAN W. "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions," *J. Personality & Social Psychology*, Feb. 1991, *60*(2), pp. 181–92.

YAARI, MAYA E. AND BAR-HILLEL, MENAHEM. "On Dividing Justly," *Social Choice Welfare*, 1984, *1*(1), pp. 1–24.

YATES, J. FRANK. *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall, 1990.