Forthcoming, *European Economic Review*

A Perspective on Psychology and Economics

Matthew Rabin
Department of Economics
University of California–Berkeley

November 23, 2001

Abstract: This essay provides a perspective on the trend towards integrating psychology into economics. Some topics are discussed, and arguments are provided for why movement towards greater psychological realism in economics will improve mainstream economics.

Key words: Behavioral Economics, Psychology

JEL Classification: B49

This paper is based on the Marshall Lecture of the same title presented at the European Economic Association Meetings, Lausanne, Switzerland, September 1, 2001, but with the underwear joke removed. I thank Lorenz Goette, Steve Pischke, and audience members at that lecture for helpful comments, and I thank the MacArthur, National Science, and Russell Sage Foundations for financial support.

Contact information: Matthew Rabin, Department of Economics, 549 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, E-mail: rabin@econ.berkeley.edu, Web Page: http://emlab.berkeley.edu/users/rabin/index.html

I. Introduction

Over the years, researchers such as Danny Kahneman, Amos Tversky, Dick Thaler, and more recently Colin Camerer and George Loewenstein, have criticized some of the tenets of mainstream economics as psychologically unrealistic. Others, such as Tom Schelling and George Akerlof, have simultaneously been innovators in mainstream, rational-choice economics, while also proposing alternatives where they felt they were needed. And prominent economists such as Ken Arrow, Peter Diamond, Dan McFadden, and Robert Solow have done only relatively little research (compared to their total output) in the area, but have advocated the broadening of economics.[1]

This agitation for greater psychological realism is now yielding results. Commonly labeled under the rubric "behavioral economics," efforts to capture psychologically more realistic notions of human nature into economics have expanded enormously in the last decade. While there is still a lot of controversy, behavioral economics is on the verge of "going mainstream", especially in top departments in the U.S. The number of recent hirings, tenurings, conferences, etc., based on behavioral-economic research reflects its growing acceptance. The theme chosen by EEA President Jean Tirole for the three keynote addresses for the 2001 Meetings and the fact that the AEA awarded the John Bates Clark Medal this year to a (second-rate, failed) theorist specializing in behavioral economics indicate that the approach has been accepted as a promising development at the highest levels of the profession.

More importantly, behavioral economics has begun to insinuate itself into work-a-day economics. Researchers such as David Laibson in macroeconomics and Ernst Fehr in labour economics have established themselves within mainstream economic fields. In several of the top U.S. economics departments, graduate students are being offered field courses in behavioral economics, and students in such departments are writing dissertations in the area.

---

[1] Of course, many researchers over the years have argued for or pursued the agenda of importing insights from psychology, sociology, and elsewhere into economics; this brief synopsis does little justice to the many such researchers.

This recent explosion of interest raises the worry that it is just a fad. Indeed, prominent skeptics have predicted that interest in the area will peter out as researchers realize that this latest craze offers little value, and we certainly do witness such fads in economics. Unsurprisingly, I'm inclined to believe it is not a fad. As with all innovations and improvements, surely many are over-optimistic about the progress these innovations will bring. But the underlying premise of this movement is far too compelling to consider it transitory: *Ceteris paribus*, the more realistic our assumptions about economic actors, the better our economics. Hence, economists should aspire to making our assumptions about humans as psychologically realistic as possible. The idea that economists should incorporate behavioral evidence from psychology and elsewhere that indicate systematic and important departures from our discipline's habitual assumptions is so fundamentally and manifestly good economics, that I am confident this line of research will have long-term influence in economics.[2]

As an indication of the long-term influence this research program is likely to have, research has recently been evolving to what I'd call "second-wave behavioral economics"—which moves beyond pointing out problems with current economic assumptions, and even beyond articulating alternatives, and on to the task of systematically and formally exploring the alternatives with much the same sensibility and mostly the same methods that economists are familiar with. David Laibson addresses mainstream macro issues with mainstream tools, but adds an additional, psychologically-motivated parameter. Ernst Fehr addresses important core issues in labour economics but without *a prior* assuming 100% self interest. Theorists such as myself use mostly the standard tools of microeconomics in exploring the implications of these alternative assumptions. All said, this second wave of research continues to employ mainstream

---

[2] Perhaps a good analogy is the advent of game theory in economics. Indeed, while game theory is now a required core topic of every major U.S. Economics Department, I was told in 1985 by more than one respected and thoughtful economist that it was a passing fad. Like game theory, psychological economics clearly expands the range of phenomena economists can successfully study, and does so in what clearly is the spirit of economics. Like game theory, it is based not on a proposed paradigm shift in the basic approach of our field, but rather is a natural broadening of the field of economics. And as I discuss below, like game theory, psychological economics is destined to be absorbed within economics, not exist as an alternative approach.

economic methods, construed broadly. But this research shows that addressing standard economic questions with standard economic methods need not be based solely on the particular set of assumptions—such as 100% self-interest, 100% rationality, 100% self control, and many ancillary assumptions—typically made in economic models but not supported by behavioral evidence.

This research program is not only built on the premise that mainstream economic *methods* are great, but so too are most mainstream economic *assumptions*. It does not abandon the correct insights of neoclassical economics, but supplements these insights with the insights to be had from realistic new assumptions. For instance, rational analysis predicts that people care about the future, and hence save, and are more likely to save the longer their planned retirement. But psychologically-inspired models that allow the possibility of less-than-100% self-control *also* make the above predictions *and* allow us to investigate the possibility that people under-save, and over-borrow, and more nuanced and important predictions such as simultaneous high savings on illiquid assets and low savings on liquid assets. Rational analysis predicts that employees more likely to quit the lower their real wages and the higher the wages available elsewhere. But psychologically-inspired models that allow the possibility of some money illusion and loss aversion and fairness concerns *also* make the above predictions *and* allow us to investigate the possibility that people are more sensitive to recent cuts in nominal wages than can be explained purely in terms of concerns for relative real wages. Rational analysis predicts that the demand for addictive products is decreasing in current and expected future prices and that people more likely to consume substances they find enjoyable, and less likely to consumer substances with bad effects. Etc. But psychologically-inspired models that allow the possibility of less-than-100% time consistency and less-than-100% foresight *also* make the above predictions *and* allow us to investigate the possibility that people over-consume addictive substances.

This essay provides my own perspective on where such research integrating psychology into economics is and should be going. I will provide a few examples of some behavioral findings that I think are important to economics. But I will focus more heavily on arguing why integrating such findings into formal economics makes sense as a research program.

I choose such a focus with some hesitation. As a rule, it is bad to spend time on "methodological" and broad-stroke issues rather than the nitty gritty of the phenomena being studied. The goal of this research program is that it become "normal science", and, as such, the nitty gritty is the point. Papers and talks should (as with this year's Presidential and Schumpeter lectures) address the substance of this new research, not its methodological legitimacy. Indeed, a recurrent (tediously repetitive?) theme of this essay is that this research is not an alternative to the economic research program into which we were all socialized in graduate school, but the natural continuation of this research program. What most of us doing psychological economics spend most of our time on—and wish we could spend *all* of our time on—is not debates over methodology, but doing normal science. Because this approach is clearly gaining acceptance, essays like this should soon become anachronistic.

At this moment in the profession, however, there is still some residual resistance to expanding the scope of this type of research. The amount of time and intellectual energy—by journal editors, graduate advisors, and seminar audiences—devoted to articulating reasons why this research should not be done is still too high. Hence, I will also use this essay to engage some of the common reasons this research is resisted.

In Section II, I will briefly outline a framework for thinking about psychologically-motivated departures from classical economic assumptions, and then discuss a few notable topics where research has been most active. In Section III, I explore a variety of themes and perspectives on the way economists ought and ought not embrace greater psychological realism into economics.[3]

---

[3] In addition to short-changing the content of psychological economics, in this essay I make lots of assertions without citing any evidence, and I provide very few references. The following sources contain some of the relevant details and further citations. Richard Thaler's Anomalies columns from *Journal of Economic Perspectives*, the early ones of which are collected in Thaler (1994), provides a well-written array of some of the most important topics in behavioral economics. Rabin (1998), my own survey article in the *Journal of Economic Literature*, provides another brief survey of the material. Camerer (1995) provides an excellent detailed article on individual decision making. Throughout the '90s and more recently, The *Quarterly Journal of Economics* has published many articles in this area (see especially the May 1997 issue in memory of Amos Tversky). *Choices, Values, and Frames*, edited by Kahneman and Tversky (2000), provides a fabulous mixture of some classical, recent, and specially commissioned papers. It is to be

II. More Psychological Realism

A Framework for Modifying Unrealistic Assumptions

There are many assumptions that economists often make about human nature that behavioral and psychological research suggests are often importantly wrong. These include the assumptions that people

are Bayesian information processors;
have well-defined and stable preferences;
maximize their expected utility;
apply exponential discounting weighting current and future well-being;
are self-interested, narrowly defined;
have preferences over final outcomes, not changes;
have only "instrumental"/functional taste for beliefs and information.

Some of the above assumptions have always been subject to doubt, others are treated as core axioms. And some assumptions are not treated as core in principle, but pervasively maintained in all actual economic analyses. Whether we label particular assumptions "classical" is not very interesting. But it is useful to treat typical assumptions as a frame of reference for thinking about what we can learn about from psychological and behavioral research and what directions economists ought consider exploring in expanding our conception of human nature.

The goal of psychological economics is to investigate behaviorally grounded departures from these assumptions that seem economically relevant. For a more concrete frame of reference, consider the following formulation of the classical economic model of individual choice, where uncertainty is integrated as probabilistic states of the world, with a utility function that may depend on these states of the world, and the assumption that the person maximizes expected value:

---

hoped that soon there will appear other sources. A collection *Readings in Behavioral Economics*, edited by Colin Camerer, George Loewenstein, and myself is planned for next year. And the book *Psychology and Economics* I am supposed to have written three years ago shall—the gods of procrastination be willing—appear in the next three years.

$$\text{Max}_{x \in X} \sum_{s \in S} \pi(s) U(x|s),$$

where X is choice set, S is state space, $\pi(s)$ are the person's subjective beliefs updated by Bayes' Rule, and U are stable, well-defined preferences[4] From this characterization of the "classical" model, I like to categorize psych phenomena for economists into three categories:

I. New assumptions about preferences—what does U(x|s) really look like?
II. Heuristics and biases in judgment—how do people really form beliefs p(s)?
III. Lack of "stable utility maximization"—do people really $\text{Max}_{x \in X} \sum_{s \in S} p(s) U(x|s)$?

This organization reflects the goal of investigating psychological phenomena as they bear on economics, and hence to extract from this research formulations of alternative formal assumptions that strive for order, parsimony, tractability, and that focus especially on research that is most relevant to and most usable by economists. These goals involve trying to be as clear and orderly as possible in identifying exactly what departures are necessary, meaning that organizing these departures ought identify as precisely as possible where and how classical economic assumptions go awry. But organizing these departures this (or any other) way is somewhat "Procrustean" because some of the distinctions I am making are quite contrived. As always, this tension between clarity/conceptual tightness *vs.* trueness to the behavioral and psychological reality is a core problem in economic modeling.

The first category of departures is to identify ways to make U(x|s) more realistic, while maintaining the assumptions that beliefs $\pi(s)$ are formed rationally and that people fully rationally maximize $\sum_{s \in S} \pi(s) U(x|s)$. Some such departures—most notably, departures from pure self-interest—are often vociferously resisted in practical and applied economic research. But in principle—by focusing on evidence consistent with

[4] The formulation above assumes even more basic assumptions—that people formulate beliefs even when no "objective" probabilities are available, and that these beliefs are correctly updated according to the laws of probability. Economic models almost always include additional strong assumptions I won't discuss below, such as 'rational expectations" and common priors.

rational choice—such modifications are the least radical class of departure from economics, and permits us to continue to use many of the standard tools such as revealed-preference theory. Examples include the assumption that people have *reference-based utility*—they care a lot about changes (in wealth, consumption, ownership, health, etc.), not solely absolute levels. It also includes *non-expected utility*—that preferences are not linear in π, as in the formula $\sum_{s\in S}\pi(s)U(x|s)$, but rather maximize a more general form $U(x,\pi)$. People care differently about uncertainties reflecting subjective "uncertainty" and those uncertainties reflecting objective "risk". Finally, a topic that has received a lot of attention in recent years is *social preferences*—that people aren't 100% self-interested, but care about payoffs of others in a variety of ways.

The second category of departures are ways that, rather than forming beliefs $\pi(s)$ through proper Bayesian reasoning, people form potentially distorted beliefs $p(s)$ about the world. Research on judgment under uncertainty identifies heuristics and biases in forming probablilistic beliefs. This allows us to still assume that people maximize *perceived* expected utility $\sum_{s\in S}p(s)U(x|s)$.[5] While assuming $p(s) \neq \pi(s)$ raises considerable problems, because such modifications allow us still to assume a classical economic notion of motivation and behavior; with such modificatons, economic actors may be confused about he consequences of their actions, but tey are still trying to maximize their preferences.

The third category of assumption modification is to consider psychological findings that suggest that there may not be stable, well-defined, time-invariant, and "hedonically correct" preferences $U(x|s)$ such that behavior is best described by assuming that people maximize $\sum_{s\in S}p(s)U(x|s)$. Examples here include exploring the ways that people *mispredict* or *misremember* their own utility—there are identifiable patterns in how people misperceive their own future taste (e.g., they under-estimate how much those tastes will change), and even in how they evaluate their experienced well-being from past episodes (e.g., they tend to under-emphasize duration of the episode). There is also considerable evidence of *framing* and *context effects:* A lot of decisions are so sensitive

[5] Or, if we wish to consider a generalized non-expected-utility formulation, $U(x,p)$.

to the framing or context of the choice set that it is difficult to associate these decisions as coming from framing- or context-free preferences on those choice sets.

I now proceed by giving (a little) more detailed discussion of three more types of departures: The ways that people caring about change rather than final states, how we care about others' well-being rather than solely ourselves, and how we care disproportionately about our well-being rather than about our future well-being.

Caring About Changes

A core feature of humans is that we are highly attuned to changes in our circumstances, not merely the absolute levels. We can feel colder—even in the same attire—if it is 50° F in the summer than if it is 45° F in the winter.

This fact about human nature carries over to preferences. For instance, our sense of well-being from our total consumption is not solely a function of its level, but also on how that level compares to what we are used to. And how we feel about not having an item depends not just on intrinsic taste for that item, but on whether or not we owned that item moments ago. And the related phenomenon of hedonic adaptation is a primary fact about human nature: Even for major life events, once a new steady state is reached, we tend over time to return to previous hedonic level. So the event of *becoming* wealthy, not just *being* wealthy, can often be a major source of satisfaction, and once we get used to new standard of living we may, day to day, be roughly as happy as when we were poor.

While the identification and measurement of how we feel about changes is an active area of research, one core aspect of our reference-based preferences is known to be crucial: *Loss aversion*. The sensation of loss relative to status quo and other reference points looms very large relative to gains. This has been identified and emphasized in a great deal of experimental work. It is seen in the evaluation of losses and gains in money, and hence attitudes towards financial risk. And it is seen in the evaluation of loss and gains of consumer items, as revealed in the "endowment effect"—the fact that people who have randomly been given virtually any object will instantly value the object more than those who have not been endowed with the object.

Reference-dependence can be thought of simply in terms of a new assumption about preferences. Letting c be a vector of the levels of wealth or consumption of goods or activities, and r be a vector of reference levels in the same dimensions, incorporating reference dependence into utility theory involves merely a switch from the function $U(c)$ to $U(c,r)$. In this sense, it is among the least radical changes one can make to economics. In fact, acknowledging—and even modeling—reference-dependence has a long history within economics.[6]

There are, however, two ways in which economics has not fully come to terms with the crucial role of reference levels in determining preferences. First, economists still haven't recognized how pervasive and fundamental is the role of changes. Assuming people care about changes to the status quo should not be treated as merely an afterthought or exotic exception to the rule that people care about absolute levels, nor even an agenda item introduced whenever there is an identified anomaly that the classical model cannot handle. Rather, it is often a crucial factor in assessing the behavior and welfare of individuals. Economists ought to develop a language and approach of treating preferences over changes as a fundamental component of preferences, and empirical methods ought be developed to do so. For instance, I believe economists should build on early attempts by some researchers to consider the welfare effects of individual income and national growth with central attention to the possibility that increases consumption may not bring lasting increases in satisfaction to the individuals involved.

The second step in more fully coming to terms reference effects is more problematic: Attitudes towards losses and other changes do not appear to be interpretable fully in utility-maximization terms. People in fact probably over-react to changes, especially losses, for a variety of reasons. Two are of special note, and have already become a focus of attention by behavioral researchers. First, some of the behavioral reaction to losses and gains seems attributable to a specific type of misprediction of

---

[6] Unfortunately, much of this has been done without assuming loss aversion or another behaviorally identified feature of preferences—diminishing sensitivity, the tendency of people to put less weight on marginal changes for changes that are further away from the reference point.

preferences: People exaggerate how long sensations of gains and losses will last.[7] We think that the pain of losing some object or some change in environment or health or the joy of increasing wealth or the elation of a new-found love will last longer than it will. By exaggerating the persistence of the sensation of loss and gain, we tend to over-react to changes. The second main reason we tend to over-react to changes is because we isolate particular experiences and decisions from each other. Losing $20 in a bet, losing a watch we just bought, or losing $10,000 in the stock market in a week all feel bad, but tend to feel worse because we too rarely think in broader, long-term perspective, where these losses will almost surely be wiped out in the longer term by other gains.

The way that people isolate separate instances of monetary gains and losses relates to a major problem in economics. Perhaps *the* most often used assumption in economics is that "risk aversion" derives from diminishing marginal utility of wealth within the expected-utility model: $U''(w) < 0$. This assumption is not just made as a simplifying assumption. It is *used:* It complicates models relative to the assumption of risk neutrality, and is used because it changes the results.

Over the years many economists have pointed out that the standard way of conceiving of risk aversion over money is not plausible in most instances in which it is applied, and, often misleading. Daniel Kahneman has called this "Bernoulli's Error": Two centuries ago, Daniel Bernoulli showed that you can explain risk aversion by assuming a concave utility-of-wealth function, and motivated this assumption with the correct argument that we have diminishing marginal utility for wealth: Money is less valuable to us if we are wealthy than if we are poor. Economists have used this argument ever since. Within the classical framework, the *only* reason to dislike financial risk is because of the change in marginal utility associated with fluctuations in lifetime wealth.

But this is a wildly miscalibrated explanation for why we dislike risks on the scale of $10, $100, $1,000, or even $10,000. If (say) you dislike a 50/50 Lose $100/Gain $110 gamble, it is *not* because of the change in marginal value of consumption due to $100 decrease or $110 increase in your lifetime wealth. This is simply way too big a change in marginal utility for way too small a change in wealth. This has been long understood by

[7] This seems to derive from a more general form of hedonic misprediction, whereby

many, but I and others have recently crystallized the problems by showing how general the problem is: There really doesn't exist a non-insane utility function within the expected-utility, diminishing-marginal-utility-of-wealth framework such that you will turn down $100/$110 bets over a broad range of initial wealth levels. *Any* such utility function also predicts outrageously wild risk aversion over larger stakes.[8] The classical expected-utility framework predicts essentially risk neutrality over non-huge stakes. And this is counter-factual: We *do* observe that people are risk averse over non-huge stakes. Hence, people's aversion towards all sorts of economic risks—leading to such tastes as a desire for extended warranties on consumer items, or our aversion to large deductibles on insurance—are simply not enlightened by the standard expected-utility framework.

Hence, we know that the standard explanation for risk attitudes is largely wrong. Our attitudes towards risk are instead primarily by attitudes towards change in wealth levels. Your current life time wealth is a complicated stochastic creature with some huge mean and variance. And your reaction to a loss of $100 isn't the difference in your anticipated lifetime expected utility between your existing complicated stochastic distribution of lifetime wealth and your new distribution of lifetime wealth corresponding to the shift $100 to the left of this big complicated distribution. Rather, what is salient to you is your sensation of losing $100.

While many particular results and insights gathered under the auspices of the expected-utility framework presumably carry over to better models of risk, many implications of the standard model—such as predictions of what types of insurance consumers do and don't buy, and how economic actors combine risks—are importantly wrong and misleading.

---

people under-predict adaptation.

[8] See Rabin (2000) and Rabin and Thaler (2001) for expanded and more precise statements of the incompatibility of the expected-utility framework and risk aversion over non-huge stakes.

Caring About Others

Economic actors are mostly self-interested. But as many economists have recognized over the years, self-interest, as narrowly defined in virtually all economic models, isn't all of human motivation. Moreover, the departures from the standard self-interest assumption are potentially important for economics, for issues like understanding the short-run reaction to market price changes, political economy, and especially labour-market institutions.

A simple hypothesis for how people care about others' well-being is natural for economists, and has the longest history in economics: Altruism—positive concern for others as well as yourself. Altruism can be either "general" or "targeted"; you may care about all others' well-being, or maybe selected others' (friends, family) well-being. Most often, *ceteris paribus*, the more a sacrifice helps somebody the more likely you are to be willing to make this sacrifice. This is as predicted by simple altruistic preferences that assume people weight others' utility positively in their own utility function. In this sense, assuming simple altruism provides insight into departures from self-interest.

But such simple altruism is not adequate for understanding many behaviors. Two other aspects of social preferences show up prominently in psychological and recent experimental-economic evidence. First, people care about the fairness and equity of the distribution of resources, beyond ways that it increases total direct well-being. Second, people care about intentions and motives, and want to reciprocate the good or bad behavior of others.

The literature identifying the nature of social preferences is among the most active areas of research in experimental economics. Let me quickly illustrate with some examples.[9] All of these decisions involve decisions as to how much money (either pennies in Berkeley, California, or pesetas in Barcelona, Spain) to allocate two anonymous parties. The first example involves Party C choosing between two different allocations for two other anonymous parties, A and B:

C chooses between (A,B) allocations: ($7.50,$3.75) vs. ($4.00,$4.00)
Approximate findings: 50% 50%

A natural interpretation of these findings (consistent with other experimental evidence) is that C may want to help these parties, but cares about both social efficiency and "equality"—producing a sort of "Rawlsian" desire to help the worse off. Those who care relatively more about social efficiency choose the higher total-surplus outcome ($7.50,$3.75), while those caring more about helping the worse off choose ($4.00,$4.00).

Now let us consider the same situation, except that B—one of the two interested parties—is making the choice. She may choose differently than does the disinterested Party C because of self-interest, or because she would be envious if she comes out behind, or for other reasons. The findings are as follows:

B chooses between (A,B) allocations: ($7.50,$3.75) vs. ($4.00,$4.00)
Approximate findings: 40% 60%

B does indeed seem to have similar preferences as neutral party C, though is a bit less willing to choose the allocation that is good for A and bad for herself. This difference (which in these cases and by replication is small but statistically significant) may be because B is self-interested, or because she is envious of coming out behind A.

The previous two examples help illustrate how parties might assess the attractiveness of different allocations in what might be termed a "reciprocity-free" context. That is, one party is making a decision that affects one or more other parties who have not themselves behaved nobly or ignobly. To see how reciprocation of the behavior of others might affect choice, now suppose that B makes the same choice as in the previous example, but chooses *after A has created this choice by rejecting ($5.50,$5.50)*. A's decision to forego an allocation of ($5.50,$5.50) in favor of trying to get B to choose ($7.50,$3.75) is clearly selfish and unfair behavior, since it involves a miniscule increase in total surplus while leading to an unequal allocation. The findings are as follows:

---

[9] The games and findings are from Charness and Rabin (forthcoming), but they are similar to many of the findings in this literature.

Following a choice by A to forego the allocation ($5.50,$5.50) to give B this choice, B chooses between (A,B) allocations: ($7.50,$3.75) vs. ($4.00,$4.00)
Approximate findings: 10% 90%

We see that B is much less likely to want to sacrifice to give the good allocation to A following this obnoxious choice by A.

Note that B's choice in the previous two examples is identical in terms of outcomes. And yet here, and in many related examples, players in games behave systematically differently as a function of previous behavior by other players. This shows that people care not just about outcomes, but also how they arrived at those outcomes. The fact that preferences cannot be defined solely over outcomes can be reconciled with preference theory, but requires an expansion of the notion of what enters the utility function. But the extra complications appear necessary to do justice in economic models to such issues as employee and citizen concerns for procedural justice, and the complications are crucial for understanding the nature of retaliation and reciprocal altruism.

Self Interest and Economics

Among experimentalists—and others paying attention to the evidence—the debate over whether there are systematic, non-negligible departures from self-interest is over. And because departures from self-interest is largely compatible with the utility-maximization framework, there has been a recent explosion of research measuring and modeling non-self-interested preferences. But to those of us who have been observing the struggle to start actively researching non-self-interested behavior, resistance by economists (including, until quite recently, many experimentalists) has been frustrating and surprising. A remarkable amount of energy had been devoted to giving self-interested explanations for laboratory behavior that seems to be a departure from self-interest.

It may be instructive to examine the resistance to the evidence. Some of these explanations are understandable (worries that experimentalists haven't guaranteed that subjects are sheltered from the shadow of reputational or repeated-game concerns that

might make sacrifice in a particular session money-maximizing in the long run) to bizarre alternative explanations (that it is bounded rationality that causes subjects to repeatedly but mistakenly choose fair outcomes in favor of selfish outcomes in simple binary choices).

To many of us, seeing an experimental subject sacrificing (say) $8 to punish an unfair ($92,$8) offer in the ultimatum game looks straightforwardly like a preference for the allocation ($0,$0) over ($92,$8) when motivated by retaliation. There is simply nothing perplexing about somebody sacrificing $8 to punish a jerk who wants to split $100 $92/$8 rather than $50/$50 (or at least $60/$40). Seemingly, people rejecting unfair offers in the ultimatum game are consciously choosing, among two possible outcomes, the one they prefer. But observing experimental economists resist this interpretation, I've been tempted to propose that a famous maxim of economics ought be modified:

De gustibus non est disputandum … exceptum if non-selfishum.

To translate this from the original Latin, "Preferences are not to be questioned … unless they aren't selfish."[10] The point is that some economists have become so enamored of selfishness as the sole human motivation worthy of note that even the most basic presumption of economics—that people are behaving according to their preferences, especially in simple choices—has been abandoned. While much behavior is boundedly rational—this is one of the key lessons of psychological economics—it is jarring to see economists resist a preference interpretation to easily interpretable, simple choices.

An analogy helps illustrate the ironic nature of many economists' responses to evidence of taste for fairness and retaliation. And it helps highlight the role of familiarity and habit in how economists interpret evidence—a theme I shall return to later in the essay. Consider again a subject who rejects an offer by another anonymous subject of splitting $100 by ($92,$8) after realizing that this other subject *could have* proposed to split the $100 by ($50,$50). Confronted by subjects choosing ($0,$0) over ($92,$8),

many economists ten or fifteen years ago were perplexed—and tried to explain away—how it is that the person would sacrifice $8 and get "nothing" in return. Maybe it wouldn't survive repetition, maybe the person was confused, maybe (despite the seemingly anonymous setting) it was sophisticated reputation-building. Arguments abounded, variously clever or contrived, for why this seemingly straightforward choice to part with $8 wasn't what it appeared to be—a willingness to pay $8 to take revenge.

Now suppose, by contrast, that we observed somebody in the same situation instead *accept* the ($92,$8) offer, leave the room $8 wealthier and then … go to the local cinema, pay $8 to see the movie *The Road Retaliator*, in which a character (played by Sylvestor Gibsonegger) tracks down and kills some fiendish bad guy who killed his wife. No well-trained economist would look for explanations for why this person spending $8 to see a movie wasn't really expressing a taste for seeing the movie. There would be no fretting about the irrationality of spending $8, only to leave the cinema two hours later with "nothing" in return. Hence, there would be no contrived and clever alternative explanations for what this subject was really trying to achieve by parting with $8 to see this movie. Paying $8 to see the movie is his preference. No problem.

But think about this. The status quo of Economics ten or fifteen years ago was that paying $8 to see a revenge fantasy of a fictitious protagonist taking fictitious revenge on a fictitious bad guy who has fictitiously wronged him falls tightly under economists' *de-gustibus-non-est-disputandum* sensibility, whereas a subject spending $8 to take *real* revenge on a *real-life* bad guy who has wronged the subject *himself* needed explaining. Looked at by an outsider not wedded to the assumption of 100% self-interest (or not a fan of popular action movies), the different reaction to retaliatory behavior versus cinematic behavior would be entirely perplexing. To me it is not so much perplexing as it is indicative of the power of habitual thinking by participants in an academic discipline. I return later to other examples (using the movie analogy) of arguments economists use in resisting unfamiliar assumptions that contrast pointedly with those employed when considering the "normal science" of familiar assumptions.

---

[10] I hope my impressive mastery of Latin here and elsewhere in this essay dispels any

Caring About Now

People like to experience pleasant things soon and to delay unpleasant things until later. To capture this preference for gratification earlier rather than later, economists traditionally model such tastes by assuming that people discount streams of utility over time exponentially.

But the exponential form of discounting has a special property that has been shown repeatedly to be false. It is the *unique* functional form that generates *time-consistent* preferences, whereby the preference between any two intertemporal tradeoffs in momentary well-being—between, say, getting lesser satisfaction earlier versus a greater amount of satisfaction later—is the same no matter when asked. The behavioral evidence, by contrast, overwhelmingly and incontrovertibly shows that people exhibit *present-biased preferences*: A person discounts near-term incremental delays in well-being more severely than she discounts distant-future incremental delays. We are more averse to delaying today's gratification until tomorrow than we are averse to delaying the same gratification from 90 days to 91 days from now.

This difference in attitudes towards delay in gratification generates time inconsistency when considering potential dynamics of behavior. Consider, for instance, the following two choices of work patterns:

7 hours work April 1, relax April 2
or
relax April 1, 7.7 hours work April 2

Suppose that opportunity costs of time, the disutility of work, the productivity of work, etc., are all identical on April 1 versus April 2. The only intrinsic difference between the two days is when in the march of time they occur.

If asked to make the choice above on January 1, you will surely prefer the first to the second choice, since it involves less work in total. You are choosing between 7.0 hours of work 90 days from now and 7.7 hours 91 days from now. The choice is obvious: less work. But if asked on April 1, you might choose the second. This is

---

European prejudice that we Americans lack serious classical education.

because if asked on April 1, the choice is now between 7.0 hours of work *today* and 7.7 hours of work *tomorrow*. Many of us might have the discipline to do the work immediately, but some of us would instead put off the onerous work until the future.

To model intertemporal preferences formally, let U^τ be "intertemporal preferences" and let u_t be instantaneous utilities. Economists assume exponential discounting: $U^\tau \equiv \int_{t=\tau} e^{-r(t-\tau)} u_t$, where $r > 0$ is a parameter. The first alternative to exponential discounting proposed by psychologists and others trying to capture present-biased preferences was "hyperbolic" discounting: $U^\tau \equiv \int_{t=\tau} 1/((t-\tau)+k) u_t$, where $k > 0$ is a parameter. In part because the continuous-time hyperbolic discounting function is difficult to deal with, and in part because the specific functional form of hyperbolic discounting is neither literally correct nor very important, recently researchers, beginning with Laibson (1994), have been modeling present-biased preferences with the following discrete-time discounting function:

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

In the above, the parameters β and δ are less than 1, with δ very close to 1. The discrete-time exponential model corresponds to $\beta = 1$.

How are these preferences time-inconsistent? They would predict precisely the behavior discussed in the work example above, for instance, if we set $\beta = .8$ and $\delta \approx 1$, if we assumed the disutility of work is linear in hours worked. 7.7 hours of work 91 days from now generates 10% more perceived discounted disutility than does 7.0 hours of work 90 days from now, since both get discounted by .8. But 7.7 hours tomorrow generates perceived disutility of .8 x 7.7 ≈ 6.2, which is less than the perceived disutility of 7.0 today.

Common sense, millennia of folk wisdom, and hundreds of psychological experiments all support present-biased preferences. While much of the psychological evidence is weak by economic standards, all evidence that exists points to present-based preferences. As absurd as it sounds, it is probably true to say that exactly zero papers in all social and behavioral sciences have proposed a test of the basic exponential versus

hyperbolic discounting when the two make discernibly different predictions, and claimed exponential explains the generated data better.[11]

But besides the direct evidence on discounting, and the large number of indirect studies that demonstrate the type of time inconsistency of behavior that is implied by present-biased preferences, I think there are two non-standard ways of seeing the superiority of the present-biased model.

First, the truth of present-biased preferences is obvious once you take off your economists' hats and think like a human. Exponential discounting says you have the same urgency to bring forward gratification from 91 to 90 days from now as you do from tomorrow to today. This is false. We feel that today is different than tomorrow. We don't feel that 90 days from now are different than 91 days from now.

Second, despite economists thinking that exponential models could address the intuitive notion that we dislike delaying gratification, it is in fact entirely miscalibrated as an explanation of most cases we can think of people not resisting gratification. Indeed, we don't need to think directly about the 90 days vs. 91 days to see the inadequacy of exponential discounting, but only measure the discounting between today and tomorrow to demonstrate the inadequacy of the model.

Consider the work example from above. Suppose we observe somebody choosing to avoid 7 hours of work on April 1 when he knows this will generate 7.7 hours of work on April 2—and doesn't feel any difference in the opportunity costs, etc., between the two dates. If we had no evidence that the person wouldn't similarly put off work from

---

[11] It is worth noting—and revealing about the role that habitual thinking plays in research—that those who first developed exponential discounting in our profession never claimed it was good assumption. It was proposed as an unrealistic, psychologically false, convenient simplification, and it is fair to say that never in the history of economics has a researcher *claimed* to have established (much less actually have done so) that exponential discounting was a better fit for any data than hyperbolic discounting. But over the years, as economists used discounting in our models more and more, exponential discounting has become a second-nature background assumption. Exponential discounting has become an assumption and crucial axiom of economics without ever having been a hypothesis. Seen in this light, the total lack of evidence for exponential discounting is less surprising: Since the only people who believe in the assumption are those socialized to treat it as a maintained hypothesis rather than a testable hypothesis, nobody has tried to demonstrate it.

April 1 to April 2 when asked on January 1, you might think we have no reason to doubt the exponential model. Assuming an exponential parameter $\delta < 1$, it might be thought, provides a simple, parsimonious explanation for the delay.

But this is not right. Suppose [u(relax)-u(7.7 hours work)] ≥ 1.1 [u(relax)-u(7 hours work)]. That is, assume merely that there is non-decreasing disutility of work. Then if you try to explain preferring 7.7 hours of work tomorrow to 7 hours today with exponential discounting, then you must assume yearly discount factor of $\delta <$ .00000000000000002. This conclusion comes from the arithmetic truth that $.9^{365} \approx 2 * 10^{-17}$. Discounting by 10% from one day to the next means—if you assume, *as you must if you believe in exponential discounting*, that the discounting will be at the same rate for every day—that over a year you will discount at the rate of $.9^{365}$. Since this is a ridiculous—and behaviorally counterfactual—discount factor, we know that the observed discounting is not consistent with exponential discounting.

By other similarly easy arithmetic exercises—such as observing that $.99^{365} < .026$, $.99^{31} < .74$, and $.999^{365} < .7$—we see that even far less extreme a taste for immediate gratification than exhibited by the April-Fools procrastinator is inconsistent with exponential discounting. Saying you are an exponential discounter and care even 1% more about today's well-being than tomorrow says, for instance, that you care now 36% more about your well-being March 10, 2007 than on April 10, 2007 and 4000% more about your well-being March 10, 2007 than on March 10, 2008; saying you care 1/10 % more about today than tomorrow says you care twice as much about May 12, 2020 as May 12, 2022. In each case, the short-term discounting is plausible, but the long-term discounting implied by exponential discounting is not..

These calibration exercises are very much like the ones I discussed above for expected-utility theory: Just as expected-utility theory based on diminishing marginal utility is—unrecognized by economists who make exactly the opposite argument—a theory of risk neutrality in small-stakes gambles, so too exponential discounting is—unrecognized by economists who make exactly the opposite argument—a theory of complete short-term patience. Any degree of short-term impatience that shows up on the radar screen implies ridiculous long-term impatience—if you cram it in the exponential-discounting framework.

Hence, exponential discounting is a theory of virtually 100% short-term patience. By contrast, present-biased preferences readily and realistically accommodate simultaneous non-negligible short-term impatience with non-ridiculous long-term discounting. If your immediate discounting is different than your future discounting, you can be extremely patient in the very long run while very impatient in the short run.

If present-biased preferences are more behaviorally accurate than exponential discounting, is the realism of importance to economists? Yes. To name but a small subset of its applications, incorporating present-biased preferences into economics likely to help us better understand: savings behavior, credit-card debt, the nature of marketing and advertising consumer goods, procrastination at work and at home, organizational design (to fight procrastination), the self-help industry, welfare participation rates, job search by the unemployed, and why people live poor and die prematurely from smoking, alcoholism, overweight, gambling, illicit drug use, unsafe sex, and other risky activities. This relatively tractable modification of economic theory unambiguously increases realism and seems to have many economic implications; this largely explains its rapid recent growth, and explains why it will and should continue to replace (when the distinction matters) exponential even more rapidly in coming years.

III. Themes and Perspectives

The examples in Section II are manifestly not an exhaustive list of even the most important departures from classical assumptions that have been identified by psychologists or have been investigated by behavioral economists. Rather than providing more examples, I turn now to a consideration of some of the more general issues in why and how economists should start to embrace this research.

Psychological Economics and Mainstream Economics

How can economists embrace all that is good in economics while dedicating ourselves to more realistic conception of human nature as it pertains to economic situations? The methods of economic analysis—methodological individualism, mathematical formalization of assumptions, logical analysis of the consequences of those assumptions, and sophisticated experimental and field-empirical testing—have many virtues. But these methods create a necessary evil: We must use highly simplified and stylized models of human cognition, preferences, and behavior that, in every instance, omit a tremendous amount of psychological reality. To formulate precise and testable hypotheses, ignoring some facet of human nature is unavoidable.

Psychology, by contrast, does (and *should*) dig deeper into the details of human nature, and isn't (and *shouldn't* be) as obsessed with the mathematical precision, generality, and empirical implementability of its findings. With these tradeoffs in mind, the different "scientific preferences" between the disciplines of psychology and economics can be conceived of in terms of the indifference curves of Figure 1.

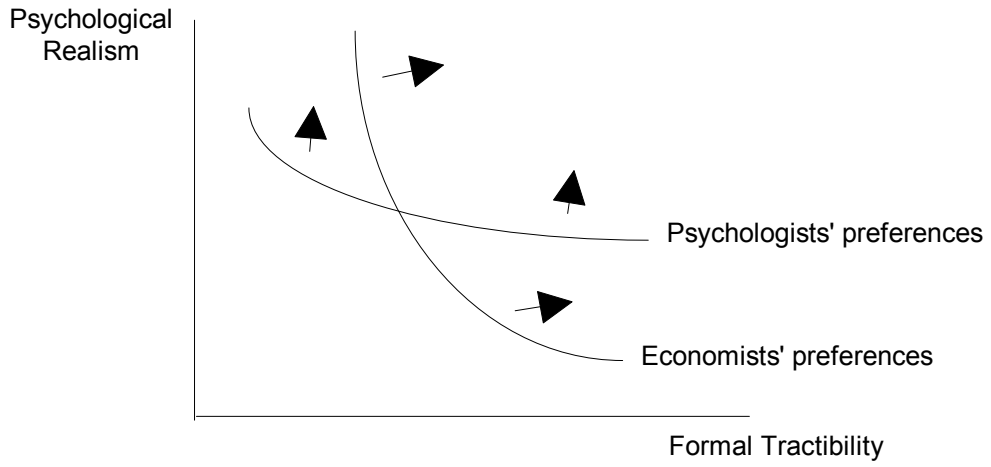

Figure 1

While Figure 1 does reflect a big part of the reason economists employ less psychologically-realistic assumptions than do psychologists, I think there are at least two

ways that it fully explains neither the difference between the disciplines nor the resistance to psychological economics.

First, the realization that many details of human behavior must be ignored does not justify blanket complacency about the behavioral validity of our assumptions. It is plainly and patently bad social science to say we don't care how realistic our assumptions are. In the dimensions of Figure 1, economists' preferences should be steeper than psychologists'—but not vertical. When the truth is complicated, then it is complicated; and when these complications need to be attended to do insightful research, then they need to be attended to.[12]

The second reason the above preferences don't seem to be the real issue is apparent when you think about it: Economists can't *really* claim (with a straight collective face) to be very "complexity-averse." Look at our journals. Look at our emphasis in publication and job promotion on technical prowess at manipulating complicated models or data sets. Look at the latest game-theoretic solution concepts or the latest life-cycle savings models. Economists do not shy away from complicated models nearly as much as some claim when embroiled in the midst of abstract methodological debates. It is odd on the one hand to be told during such debates that economists must forego behavioral realism for the sake of keeping our models simple—when in the other hand we are holding a copy of *Econometrica.*

Indeed, the disconnect between the professed urgency to keep things simple and the actuality of a very complicated models is most frustrating for behavioral economists when unparsimonious and intractable hypotheses have been proposed merely because they use "standard" assumptions. To return to the earlier example, early attempts to explain behavior in experiments such as the ultimatum game as not really being departures from self-interest would—if actually followed through on in developing economic applications—generate more complicated models than the types of social-preference models currently being developed.[13] Similarly, I think it is clear that ten years

---

[12] As Albert Einstein put it, "Make your theory as simple as possible—but no simpler."

[13] I am not claiming that the models of social preferences being developed to explain rejections in the ultimatum game, for instance, will be as simple as the unmodified self-interested model that predicts no rejections. I am claiming that if economists actually

from now we will have reasonably tractable present-biased models of savings patterns that are more general, more accurate, and *simpler* than recent rational-choice models that try to explain savings behavior that cannot be accommodated by the simpler rational-choice models. There are many other examples where behavioral hypotheses will end up providing simpler, more tractable, and more useful (less *post hoc*) explanations than existing models. Especially when realizing that economists will become more agile over time in working with psychologically-inspired models as we familiarize ourselves with them, it is empirically false that these psychological models will be significantly more complicated than the classical models that have less explanatory power.

All said, my impression is that the real difference in taste between psychological economics (as I envision it) and a lot of classical economics is better represented by the indifference curves of Figure 2, where parsimony, tractability, generality, etc., are held fixed.

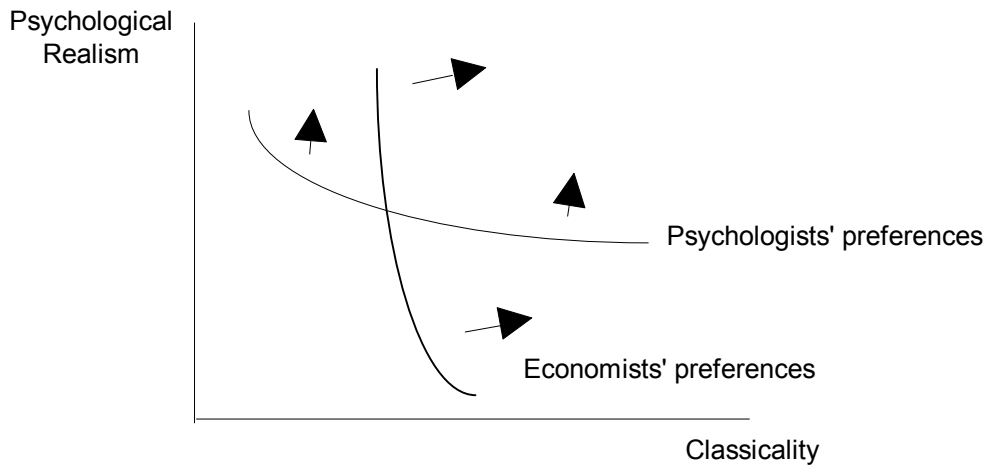

Figure 2

---

formalized some of the verbal explanations of reputations models and signaling theories that predict actual behavior such as ultimatum-game rejections, then *these* formalized models would likely be more complicated than the social-preferences models being developed.

Cautions and Worries

As the amount of research integrating psychology into economics expands, we will of course see more research in this arena that is bad, and most of the potential flaws that might arise parallel those that can be found in any research program. But there are two types of problems to which psychological economics is particularly susceptible.

First, the growing acceptance of behavioral economics has made some economists overly receptive to alternative assumptions that are no better grounded in psychological reality than the classical assumptions they are replacing. We must never lose sight of the goal of making economic assumptions more realistic. The goal is not merely to replace existing assumptions with random different ones. It is to replace the ones that are importantly imperfect with new ones that better capture realistic and economically important aspects of human nature. This means both recognizing how right so many economics assumptions are, and being discriminating about the psychological soundness of the new evidence. The standard for alternatives ought not be just whether there exists a "psychological" explanation, but does the evidence (and, where appropriate, intuition) indicate that it is the *right* explanation?

My second worry about the development of psychological economics is a bit more speculative, and based on impressions I have about not the research and modeling of psychology, but rather the psychology of research and modeling. I argued above that there is no inherent negative correlation between psychological realism on the one hand, and taste for tractability, formalism, parsimony, and simplification on the other hand. But, historically, and currently, there *is* a correlation between those who like clear and simple mathematical models and those who care about and study behavioral realism. The psychology of modeling and research may help explain why one discipline (psychology) studies people in detail without precise simplified models while another (economics) pervasively employs precise models without seriously attending to the behavioral reality of the underlying assumptions. To motivate this, consider a depressing quote from another realm:

> The history of the Victorian Age will never be written: we know too much about it. For ignorance is the first requisite of the historian—ignorance, which simplifies and clarifies, which selects and omits, with a placid perfection unattainable by the highest art.
>
> —Lytton Strachey, preface to Eminent Victorians

Strachey here seemed to imply the discipline of history—that seeks to capture the general patterns, and the spirit of an age—is crippled if we know enough details about an era to realize that no simple story really gets things quite right. Similarly, I get the impression that knowing the messy reality of human psychology hampers our willingness to model general patterns with simple, stylized models. We find it harder to construct simple models if we don't first convince ourselves that people really *are* as simple as our model says. This is unfortunate, and may in part explain the bifurcation between psychology and economics back when the two disciplines separated. And I have a "separation anxiety" about current research trends: Those that employ precise formal models are keen to avoid studying the behavioral evidence so as to convince themselves that we have our models just right; and those that are keen to improve the realism of our assumptions fret so much about getting the details right that they do not tolerate usefully stylized models.[14]

I may be naïve about our ability to perform the "highest art" in coping with messy reality, but I hope our discipline doesn't re-bifurcate into behavioralists and modelers. Just as knowing that no simple generalization about a time and place is literally true shouldn't prevent historians from being willing to attempt usefully general statements, so too learning enough psychology to know that no simple generalization about some facet

---

[14] For example, recall that present-biased preferences better describes the way people discount than does exponential discounting. To my knowledge, the simple, unadorned exponential discounted-utility model explains no data set as well as the simple, unadorned present-biased-preferences discounted-utility model when the two make substantially different predictions. But there is also some convincing research showing that *neither* discounted-utility model can really capture all behavior. Since those who are aware of the greater behavioral accuracy of the present-biased model are more likely to be aware of the imperfections in the present-biased model, the bifurcation I fear has started to appear—some researchers made aware of the more extended evidence have used the not-one-hundred-percent-correct feature of the hyperbolic-discounting function as an argument against it. Worse, and rather incoherently, this has in turn been picked up by some economists as a justification for sticking with the exponential model.

of human behavior is literally true shouldn't prevent us from attempting to develop stylized, tractable models that aid us in economic insight.

Some Common Objections

I now discuss some of the types of explanations I've seen and heard for why economists should resist greater psychological realism. I'll begin with some seemingly methodological arguments. One class of arguments derives from the following understandable plea:

"We Can't Consider All Alternatives."

There are an infinite number of theories that are consistent with every empirical finding. Hence, we need some sort of discipline to not have a new theory for every new experimental paper. Economists worry that, if we allow new assumptions, then researchers could come along and assume *anything*.

In this sense, the reaction to psychological economics is similar to the reactions by economists to other innovations. As game theory, information economics, and especially transactions-costs economics rose to supplement Walrasian economics, it was often complained (and sometimes with merit) that "you can explain anything" with these models, and hence embracing this broadening of the discipline of economics will turn it into an undisciplined non-discipline, with no restrictions whatsoever on our assumptions.

As with the resistance to the previous challenges to the status quo, this worry about psychological economics has some merit. But even more so than with the previous resistance, it seems to me mostly misguided. In the case of psychological economics, the biggest problem with this complaint is blatant: As noted above, the whole point of this agenda is *not* to come up with random new undisciplined hypotheses. While it is true that psychological economics will lend itself to a healthier tolerance for *ad hoc* assumptions—sometimes the complexity and context-specificity of humans merit a tolerance for a wider range of context-specific modifications to assumptions—observing the actual content of recent behavioral economics does not lend itself to interpreting

proposed new assumptions as random. A more common complaint about some behavioral economists (myself included) by those who've heard us frequently is the opposite one—that we are tediously repeating the same themes over and over again. Principles such as loss aversion, diminishing sensitivity, and self-control problems are ones that we are using over and over again to explain a wide range of phenomena, not *post hoc* inventions to explain particular behavioral anomalies.

Where does the perception that psychological economists are making random, "ad hoc" assumptions come from? I think it comes largely from economists' unawareness of the psychological and experimental evidence. The assumptions being proposed are not the ones economists have been trained in, and as a discipline this makes them seem like they are coming from nowhere. I think many economists tacitly use the fallacious rule of thumb:

"Non-Varian hoc, ergo ad hoc"

Translated from the Latin, this means: "That assumption was not in our graduate microeconomics text; therefore it is some random assumption that you're making up." This reaction is wrong-headed on many accounts. First, psychologists *did* learn many of these principles from *their* graduate texts. Second, many behavioral hypotheses (e.g. we get angry, we have self-control problems) we knew *before* graduate school—and *unlearned* in graduate school. (Or at least trained ourselves not to think about when we have our economists' hats on.) Finally, especially with the advent of experimental economics, even if economists are unconvinced by extant psychological evidence, our reaction ought not be the presumption that the evidence is wrong. We ought to test it. In summary, whether or not economists are at this moment familiar with the new assumptions being proposed, they are *not* coming out of thin air. They are being proposed because they seem to be behaviorally true.

Sometimes the resistance to new assumptions seems not so much from economists seriously challenging the validity of these alternative assumptions, but rather from a perceived methodological mandate not to even debate the validity of alternatives. Some economists seem to come close to the belief in maintaining classical assumptions against just-as-simple and more realistic alternatives with the methodological claim that a

discipline's current assumptions automatically deserve a sort of epistemological pride of place. This amounts to a sort of prescriptive or normative Kuhnianism:

"Thomas Kuhn says we shouldn't 'think outside the box' …
until the box is wholly shattered"

Thomas Kuhn's (1970) familiar insights into the difficulty communities of scientists have in abandoning set paradigms were not (as I interpret it) prescriptive—they were descriptive. It is not good science to declare we *shouldn't* mess incrementally with a paradigm, nor to insist that incremental critiques and improvements ought be ignored until we replace the current paradigm in one fell swoop. As we find pieces of the classical model that are wrong, then insofar as we can recognize how to replace them, we ought replace them.

A related methodological encumbrance to the progress of psychological economics is one of the most powerful mechanisms across disciplines to maintaining current hypotheses (and paradigms) against the preponderance of the evidence that these hypotheses are probably false: Placing the burden of proof on hypotheses outside the paradigm. In both formal statistical terms and in more subtle ways, it is clear that many of the instances where economists have argued the current models are adequate, they do not mean that the current model seems to fit *better* than proposed alternatives, but merely that the classical model isn't yet provably false. Maintaining the classical model as the null hypothesis that must be disproved lends itself to maintaining the model even when the accumulated evidence is strongly against it. It seems to me plainly appropriate in scientific terms to stop treating the classical assumptions as the maintained hypotheses in our analysis, and start treating them as special cases, corresponding to particular parameter values of a generalized model, and then investigating what are the best fits for those parameter values.

On particular variant of the view that we ought shun psychological improvements to our models until a superior alternative is proven is that most mischievous of clichés:

"If it ain't broke, don't fix it"

The standard model is successful in explaining a lot, it is sometimes argued, so why make a fuss? Besides a general heuristic of avoiding complacency-enhancing clichés, this attitude ought be avoided because it misconstrues the nature of psychological alternatives to current assumptions. It is illogical to doubt a behavioral hypothesis by showing that standard hypothesis being challenged isn't 100% wrong. The right question is whether standard assumptions are less than 100% right, and whether the shortfalls are sufficiently identifiable, sufficiently systematic, and sufficiently important that economists should study them. Many mostly right assumptions are importantly wrong or incomplete, and after decades of figuring out all the ways our assumptions are mostly right, it is more productive to start asking how they are importantly wrong.

The if-it-ain't-broke criterion is frustratingly beside the point to those of us who believe that classical economic assumptions are a wonderful foundation upon which to build. People are largely self-interested. If we are allowed only one hyphenated adjective describing human motivation, "self-interested" would be my choice. But we are not *completely* self-interested, and the departures appear not to be economically negligible. People have some self-control and significant propensity to pursue long-run desires over immediate gratification. But we are not *completely* self-controlled, and the departures appear not to be economically negligible. Much of human behavior is usefully conceived of in terms of rational maximization of coherent preferences. But our tastes are not *completely* well-defined, stable, and coherent, and the departures appear not to be economically negligible. In all these realms, economics is not "broke" in the sense of being useless, but it should still be fixed.

Besides "methodological" arguments for dismissal of the agenda of psychological economics, there are also substantive ones. The one that I've heard most often and in the most variants—and that perplexes me most—is the assertion that we needn't worry about unfamiliar new assumptions, because they won't survive in markets. In its generic form, it can be stated as follows:

"Markets will wipe [any unfamiliar psychological phenomenon] out"

While I have occasionally seen specific variants of this argument in print, most often it is made orally and on the fly, in the context of rebutting. More often than not, "wipe-out

arguments" are logically wrong. They are not bad psychology, but bad economics. In our discipline ruled by careful, formal arguments, the frequency of oral, loose, off-the-cuff seminar-audience arguments about markets negating departures from our habitual assumptions is striking.

There are realms, such as low-transactions-costs asset markets, where markets might greatly diminish the implications of some psychological phenomena. While an active debate exists as to interpreting how influential irrationalities are in even the most competitive asset markets, I want to here state three reasons why even if highly competitive asset markets *do* wipe out the influence of psychological phenomena, we should still not ignore the psychological phenomena.

The first is the very definition of "wipe out". Financial economists often care only about approximate market prices, and explicitly rely on arbitrage arguments to explain how markets will be "efficient" through the efforts of even a few more rational or better informed traders. But the existence of rational traders equilibrating prices by arbitraging against irrational agents means that there are distributional and efficiency consequences. For those of us who care quite a lot not just about the trading price of assets, but also whether some investors are failing to maximize their lifetime utility—retiring poorer or sending their children to less expensive colleges than they could—the market efficiency definition used by financial economists is inappropriately limited. It is not the same one used in economics more generally, which concerns not solely the price generated, but—far more importantly—the allocation achieved. Insofar as asset markets influence the economy not solely through how the prices of stocks and bonds affect company investment designs, economists should care about far more than this one aspect of asset markets. Debates over market efficiency in this context have been hijacked by the narrow and specific aspect of efficiency employed by only one field within our discipline.

Second, not all economic behavior is mediated by frictionless, Walrasian markets. It has been a *long time* in most economics departments, journals, etc., since we have

assumed perfectly competitive markets as the only institution of interest.[15] My point is not that studying psychological phenomena in the context of competitive markets is unimportant. Because markets are such a major institution, and may magnify or mitigate the effects of certain psychological phenomena, it is important to investigate the market implications of these phenomena. But perfect competition is manifestly not the sole environment of interest.[16]

This leads to my final point about the inappropriateness of ignoring phenomena merely because they don't manifest themselves in competitive markets. If the effects of some aspect of human nature are "wiped out" by realistic markets, this is very important fact—and arguably *intensifies* rather than diminishes the importance for economics of studying that aspect of human nature. One of the main things economists teach the world—arguably, *the* main thing—is about how markets *compare* to other allocation mechanisms. A central theme of every Economics 1 course is the putative efficiency of markets in comparison to distortionary taxes, natural or unnatural monopolies, price regulations, etc. Hence, it is wrong to say that we are unconcerned with some psychological feature of market participants just because markets wipe out the implications of the feature. To compare market outcomes to other outcomes, this is precisely the type of thing we do care about. If, for instance, markets destroy fair behavior that might manifest itself in non-market settings, then we should (when articulating our welfare theorems) compare unfair market behavior and outcomes to potentially fairer ones in non-market settings. Or if markets somehow eliminate

---

[15] Historically, the focus on highly competitive market environments has been especially pronounced among experimental economists. It was jarring for me as I came of graduate school (at MIT) twelve years ago and started following experimental economics to see the very narrow notion of economic institutions (typically, highly anonymous double auctions) studied by the experimental economists. This focus on perfect competition, right or wrong, simply didn't match the focus of research and teaching at MIT and much the rest of economics. Fortunately, this narrow focus of experimental economics has decayed magnificently in recent years.

[16] Interestingly, even economists who do research on non-Walrasian institutions seem prone to apply the Walrasian setting as the test of whether a phenomenon is of interest. It seems that, when engaged in a challenge to our familiar way of doing things, we tend to invoke a archetype of an economic situation that does not correspond to the actual economic situations we study typically in our normal-science, workaday research.

cognitive errors that we'd see in non-competitive environments, this would be a major, under-explored efficiency feature of markets.[17] To many of us, comparative institutional analysis is the core goal of economics. To abandon this mode of analysis when we consider departures from classical assumptions is to abandon a big piece of the classical goals of economics.

There are many other objections given to introducing psychological phenomena into economic analysis. In lieu of a more complete list of such objections and my rebuttals, and to elucidate some of the above arguments, a shall instead relate a none-too-subtle parable that continues from the movie example earlier in the essay, and that continues with that theme of how differently economists react to unfamiliar assumptions than to classical ones.

At the Movies: Economics from Another Planet

I earlier illustrated an untenable difference in economists' reaction—and standards of proof—in accepting the evidence for the unfamiliar assumption that people have a taste for retaliating against unfair treatment to such paragons of acceptable preferences as enjoying action movies. This example can be stretched further.

Suppose that in another galaxy there is a planet—Planet Nonhollywood—where the actual economy developed exactly has it has here on Earth, and that the economics profession evolved *almost* exactly the same as here. There is just one difference between the economic professions on the two planets: On Planet Nonhollywood, economists had traditionally not studied preferences over "entertainment" items—things that weren't physically consumed, but merely "psychologically consumed". I don't want to be

---

[17] The argument to ignore more realistic assumptions simply because classical ones do fine in Walrasian markets is analogous to a decision to stop assuming that firms maximize profits rather than to (more simply) assume that they set price equal to marginal cost. The price-equals-marginal-cost hypothesis does fine in competitive environments, so, it might be asked, why worry about departures from that assumption? The answers are many. To take one example, price-equals-marginal-cost would be an

judgmental about the coherence of our colleagues across the universe; for whatever reason, they had developed what they thought in their own minds was a sensible distinction. Roughly corresponding to what we would label entertainment items, they had a strong sense that things that could not be eaten, touched, pushed, etc., and especially that were temporary, were simply not a direct part of anybody's utility function.

By contrast to the economists, psychologists on Planet Non-Hollywood talked all the time about such phenomena, but economists dismissed them as wooly-headed and unscientific. Because the economy is identical to that of the Earth's, many people (including economists) would go see movies and entertain themselves in other ways. And a few economists started to suggest that maybe people did intrinsically enjoy "non-tangible consumption," noting all the money in the economy devoted to the enjoyment of others' company, entertainment industry, the role of ambience in restaurants, the fact that people travel to see beautiful buildings and paintings even if they couldn't touch them. But most economists had simply managed to ignore these phenomena, or come up with alternative explanations using (what to them were) acceptable, "classical" assumptions.

Now imagine you travel to Planet Nonhollywood to give seminars on the movie industry, arguing that people intrinsically value movies. The following are some of the responses you might encounter.

"But there are alternative 'standard' explanations!"

Your evidence was very weak, and acceptable standard explanations abound for why people would pay to see movies even if they didn't intrinsically enjoy them. When so many standard explanations exist, why introduce a crazy new assumption? For instance, movie-goers could be going just for the food; and indeed, cinemas make more money from the sale of food than from the movies themselves. The large amounts of food consumed—and the convenient, efficient seating by which to eat it—support this explanation. Indeed, all sorts of predictions of the standard model are borne out, making

---

awful assumption to maintain in our research and in our teaching when discussing the effects of monopoly.

the new-fangled wanting-to-see-the-movie interpretation redundant. For instance, the better and cheaper the food, the more likely people are to go.[18]

Also, you didn't adequately demonstrate that people weren't going to the movies intending to make money rather than spending it—hoping to find money underneath the seats in front of them. When you admitted not having the data to rebut this hypothesis, but said you found it implausible, you are dismissed as unscientific, and treated to an anecdote of an audience member once having found twenty Nonhollywood dollars on his seat.

Or, you were told, movie-going could be a signal of wealth by wasting money: There is lots of evidence that people (especially males trying attract mates) like to pay for movies. The fact that when wealth-signaling is most likely—as on courting rituals—people are most likely to pay for the movies strongly supports this explanation. Audience members in fact provided anecdotes of all the times they went to movies they didn't want to see merely to be on a date—*and* they paid for both tickets, hence doubly signaling their wealth. When you asked them why their date wanted to go, you got told that the move (and the food) were free to the date, all supporting the standard interpretation of no intrinsic taste for movies.

Other audience members argued:

"But the alleged 'preference' is 'unstable'"

It was often pointed out, and backed up by research, that this alleged preference for seeing movies is highly sensitive, and therefore not a real preference. While it is true that some people like going to the movie, it varies a great deal. It depends on mood, time of day, etc. Indeed, while behavioral researchers claim to have evidence of people willing

[18] And when you unwisely pointed out the huge demand for TVs, where the food-buying explanation seemed more tenuous, you were readily rebutted by being told that people merely watched TV for for the information gleaned from cereal and detergent commercials, and that the companies paying shows to air commercials were simply signaling their quality by being willing to pay for the wasted minutes of non-commercial watching, and that it was an equilibrium for viewers to sit through shows they didn't want to see because they knew they would be rewarded by informative commercials.

to pay for movies, a great deal of experimental evidence by economic experimentalists show that this taste goes away under only slightly different conditions.

Moreover, when the experiment was done properly—in the way economic experimentalists understood how to do experiments—the taste for movies nearly completely went away. Evidence from well-run economic experiments shows that this alleged taste for movies is highly ephemeral.

"***But evidence shows people learn they don't like movies …"***

While a few psychologists have argued that they have evidence that people seem to like movies, these experiments are run under novel conditions, and don't allow learning. Indeed, the standard in psychology experiments was to only ask people to see a movie *once*. Hence, you were told, we do not learn whether this behavior represents a robust preference. But experiments showed that, while a person might pay $8 to see the movie once, maybe twice, if you keep asking him for $8 to see the movie, eventually stops paying. Clearly he learns he doesn't want to see the movie! Once play "converges" to "equilibrium" behavior by subjects, we see no genuine preference for movies.[19]

Indeed, an audience member pointed out that this provides further support for the money-under-the-seats interpretation of movie going: People start going to the movies, then when they don't find money under the seats in front of them, they learn it involves losing money rather than gaining money, so they stop going.

Another audience member pointed out that this movie-seeing couldn't last, since it is clearly irrational:

"But this behavior is 'non-consequentialist', and hence irrational!"

The notion that people might care directly for the movie they see is so much "psychobabble." Movie-goers walk out of the theatre with no more than they walked in, $8 poorer. To pay $8 for a two-hour process that puts nothing in a person's hand or

[19] Audience members admitted that subjects still paid to see Johnny Depp movies (he is as popular there as here) after twenty rounds, but noted that the trend was downward, so that surely if the experiment were conducted for more rounds, they would have eventually learned that they don't like watching even Johnny Depp.

stomach is, you are told, irrational. Some researchers had started to toy with such "non-consequentialist" preferences, and conceded it was an interesting possibility, but realized such preferences couldn't be rational—getting nothing for your $8 can't really be rational.

Indeed, if there were people who went around giving $8 for nothing in return, they would quickly be driven from the market, so that there behavior would not matter:

"But those behaving like this will be driven from the market!"

An audience member assured you that somebody willing to pay $8 for a movie could be "Dutch-booked": If people paid $8 just to sit in front of a screen, then somebody could make money off of them![20] When you respond that, yes, somebody could and *is* making money off of those willing to pay the $8, another audience member assures you that if people were really willing to pay $8 for nothing in return, they would in short order be bilked of all their money by an arbitrageur. When you shyly suggest that a consumer's willingness sometimes to give some of his money to see a movie doesn't mean he'll pay infinite amounts to anybody who offers movies, or suggest it might be costly to provide these movies, you get scoffed at for being ad hoc, changing your story, and being very loose about what preferences you were proposing.

Somebody else also points out that, because people are irrationally paying $8 for nothing, such movie-goers will be driven from the market by those who are as happy not seeing the movies. These others will have all this money that movie-goers won't have, and hence have greater survivorship in the goods market. Moreover, because the spending needs of non-movie-goers are lower, they will undercut movie-goers in the labour market, accepting the same jobs at lower wages, so that would-be movie-goers will be left unemployed and not be able to go to see movies anyhow!

That line of argument makes no sense to you, and as you are trying to articulate a rebuttal to such a far-fetched story, you get hit with the seminar-stopping argument you always fear:

---

[20] Nobody on Planet Nonhollywood had any more idea as to where the term "Dutch book" comes from than we do on Planet Earth.

"But none of the evidence is for 'real stakes'"

Somebody raises his hands, and while acknowledging that maybe your experimental results are robust, you haven't demonstrated willingness to pay for movies for stakes that really matter. "Sure," an audience member starts, "they will go see the movie for eight dollars—but would they for eight *hundred* dollars?"

You know you are beaten. You hem and haw and explain you would love to get more funds from the Nonhollywood Science Foundation to write bigger-stakes experiments, and note that a colleague has run similar experiments on a poorer nearby planet, and while the willingness to pay for movies is smaller, it is still non-negligible. You can't convince the audience you still care about movie-going even though it involves non-huge stakes. You are not going to convince this audience.[21]

---

[21] But, happily, that times are changing, and you find friendlier audiences willing to acknowledge a taste for entertainment. You begin your seminar at some such place, and look forward to discussing how you measure and model this taste for entertainment and its economic implications when you get asked … ***"How would this evolve?"*** An audience member points out that, while your evidence is interesting and convincing, at first blush it would seem that those without the taste for wasting money on movies would surely have better survival likelihood. So the *really* interesting question is how these preferences would survive evolution. You say politely (or not so politely) why you don't care how it evolved, but felt sure it was there. You have an argument back and forth, before continuing with your seminar to a patently bored audience, who are busy asking themselves the more interesting question of how this new-fangled motive you are discussing could have evolved.

After this seminar, out of curiosity you study the recent seminars on Planet Nonhollywood. You estimate that at about 85% of seminars about "non-tangible consumption" (like movies) and 2% on seminars on "real" consumption, (like taxi rides), speakers are asked how the behavior would evolve. And, in reading, you notice that 50% of the papers written on movie-going and 0% written on taxi rides are about how it would be evolutionarily possible for people to want to consume the product. None of the burgeoning research on the evolutionary roots of entertainment-preferences seem equipped to update anybody's beliefs about whether people actually go to the movies, but it was argued that it was important to understand these evolutionary roots of the phenomenon. One reason given was that it would help persuade other economists that it was possible to want to go see movies, so they would stop looking for other explanations. On that score, you felt you needed to make your peace with the emphasis on evolution as it was helping you get economists to accept the reality of this preference (which you thought they should accept merely because it was real).

Back On Earth

Is this tale of arguments against the people-enjoy-movies hypothesis a fair parable of what kinds of arguments we see here on Earth against the sorts of psychological assumptions I have discussed?

It would of course be unfair to say that all economists who have resisted psychological assumptions have made arguments analogous to those above. In fact, few of these arguments are written down (though some are) because they would look transparently silly if they were made about what we (on Earth) consider standard preferences. The intended moral of my tale is, of course, precisely that many of the arguments used against unfamiliar assumptions are awful economics and don't hold up to scrutiny. They would be embarrassing to authors when made in print, and would in any event not be accepted by editors.

But these types of arguments were the types of arguments frequently made at seminars ten years ago, and occasionally made today, against unfamiliar but psychologically sound assumptions. In fact, many of the examples parallel economists' reactions to those researchers trying to argue that people have intrinsic taste for fairness and other non-self-interested preferences. These are the types of arguments I used to hear all the time (but far less frequently now).

All manner of self-interested explanations for rejecting unfair offers in the ultimatum game were offered. To those of us who see nothing mysterious about rejecting such offers, the pursuit of money-maximizing interpretations is as strange as a money-maximizing interpretation of movie-going. The fascination many economists have had with the "instability" of the taste for fairness—and the predisposition to argue that the experimental manipulation that produces the least concern for fairness yields the "true" stable preferences—has struck some of us as wholly disproportionate in comparison to the lack of focus on the "instability" of classical assumptions about preferences. Showing that there exist conditions where a preference for fairness "goes away" isn't a demonstration that these preferences don't exist. And the tendency of experimental

economists to assume that all temporal changes in behavior represent "learning" would be recognized as transparently misguided if applied to standard assumptions. It is silly to label the diminishing marginal utility of seeing the same movie repeatedly as a person learning he doesn't like the movie. But experimental economists studying the important topic of learning, or wedded to a methodology of repetition in running experiments, are ignoring the possibility that people with a taste for revenge and similar non-self-interested tastes might similarly exhibit satiation.[22]

And the consequentialist arguments that those who reject money in the ultimatum game get "nothing" in return similarly derives from habitual thinking what constitutes "something" and what constitutes "nothing". And the related arguments that paying money for nothing but the satisfaction of revenge is a money-losing behavior and therefore unable to survive in the market or in evolution seem highly confused. Every time a consumer buys anything he loses money. And every time he buys something without the highest caloric and nutritional survival payoff we have an evolutionary mystery on our hands for those that want to solve an evolutionary mystery. For those interested in economic outcomes, it will be more sensible and more adaptive to assume that people are willing to spend money on whatever they are willing to spend money on.

And at the lion's share of seminars on non-self-interested behavior five years ago the speaker would be cross-examined about the low stakes in the experiments being discussed. It is perfectly reasonable to care a lot about what the shape of the demand curve for revenge looks like. (And the answer as far as the evidence so far suggests is that it is not very steep—many people are willing to pay quite a lot for justice.) But the mere fact that the taste for revenge and fairness is finite, and diminishes when it is more costly to purchase, makes it like every single other taste economists study, not something

---

[22] This mis-identification is very understandable both because learning *is* such an important phenomenon and because existing theories of social preferences don't provide predictions about how behavior might change over time. But the tendency to equate temporal changes in behavior with learning economic experiments is clearly something economists wouldn't be inclined to do in other contexts. And especially in experiments that are cognitively and strategically simple, interpreting a trend towards self-interested behavior as "learning" by inherently self-interested players, rather than satiation in pursuit of social goals, or reference-dependent adjustment of preferences, strikes me as a bad default presumption.

to be dismissed. And the presumption that economists don't consider even $8 transactions to be "real stakes" is plainly false, as the movie-going example illustrates.

Final Thoughts

Above I have proposed that economists have until recently resisted new assumptions about non-self-interested preferences based on economically flawed arguments. Similar flawed arguments are used to resist other modifications that will improve the psychological realism of economics. As economists are starting to realize that meta-arguments for dismissing these modifications are not helpful to economics, and to realize that adding greater psychological realism will improve it rather than undermine economics, such defensive arguments are decreasing. Happily, the trend is towards integrating apparently true and apparently relevant new psychological assumptions into economic analysis.

References

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

Charness, G. and M. Rabin, forthcoming, Understanding social preferences with simple tests, Quarterly Journal of Economics.

Kahneman, D. and A. Tversky, 2000, eds., Choices, Values, and Frames (New York: Russell Sage Foundation; Cambridge, UK: Cambridge University Press).

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

Laibson, D., 1994, Essays in hyperbolic discounting, dissertation, Economics (MIT).

Rabin, M., March 1998, Psychology and economics, Journal of Economic Literature (XXXVI), 11-46.

Rabin, M., September 2000, Risk aversion and expected-utility theory: a calibration theorem, Econometrica 68(5), 1281-1292.

Rabin, M. and R. Thaler, Winter 2001, Risk aversion, Journal of Economic Perspectives 15(1), 219-232.

Thaler, R., 1992, The winner's curse: paradoxes and anomalies of economic life (New York: Free Press).