Journal of Economic Perspectives—Volume 17, Number 3—Summer 2003—Pages 159–176

In Honor of Matthew Rabin: Winner of the John Bates Clark Medal

Colin Camerer and Richard H. Thaler

Although there is some evidence that Matthew Rabin existed before 1990, we had the pleasure of discovering him for ourselves when, in the early 1990s, he sent each of us a copy of his manuscript "Incorporating Fairness into Game Theory and Economics" [2]. Matthew was, at this time, an assistant professor in Berkeley's economics department, having recently finished his graduate training at MIT. The paper was remarkable in many ways, and it induced us both to call around and ask: "Who is this guy Rabin?" Now, just a decade later, we find ourselves writing an article in honor of his winning the John Bates Clark award. So, who is this guy?

Fairness

In explaining who Matthew Rabin is, and why he deserved the Clark award, we will start with that remarkable fairness paper. Economists and game theorists have long used two standard assumptions in modeling behavior: rationality and self-interest. These working assumptions persisted in spite of growing experimental evidence that both rationality and self-interest are "bounded." In game theoretic contexts, evidence that people care how much others get was abundant: In a standard one-trial prisoner's dilemma game, where the rational selfish choice is to defect, roughly half the players cooperate. Or consider the ultimatum game, in

■ *Colin Camerer is Rea A. and Lela G. Axline Professor of Business Economics, California Institute of Technology, Pasadena, California. Richard H. Thaler is Robert P. Gwinn Professor of Behavioral Science and Economics, Graduate School of Business, University of Chicago, Chicago, Illinois. Their e-mail addresses are ⟨camerer@hss.caltech.edu⟩ and ⟨thaler@gsb.uchicago.edu⟩, respectively.*

Matthew Rabin

which one player (the proposer) makes an offer to divide some pie between herself and a responder, and the other player, the responder, can either accept, in which case the division is made as proposed by the proposer, or reject, in which case both players get nothing. Selfish rational responders will accept any positive offer, but offers of less than 20 percent of the pie are typically rejected, even when the pie is substantial—say, a "pie" of $400 for Americans or several days' wages in less-developed countries (Camerer, 2003, chapter 2). Many economists had reacted to this anomalous evidence by either trying to make it go away (say, by giving experimental subjects more opportunities to learn or by raising the stakes) or by trying to rationalize it (typically, by arguing that subjects are mistakenly applying a rational repeated game strategy in a one-shot game). Neither of these reactions generated much new insight.

Matthew took a different tack that was unusual at the time, taking the experimental evidence at face value and using it to infer a utility function incorporating fairness. Suppose that responders in the ultimatum game reject offers of 5 percent of the pie because they think that a 95–5 division is "unfair," and they are willing to forego 5 percent in order to punish the offending proposer. Suppose that subjects in the prisoner's dilemma game cooperate because they realize that if everyone cooperates, they all are all better off. Rabin set out to incorporate such common-sense notions into a standard game-theoretic model.

Rabin went about this task with what has now become his standard modus operandi. He began by critically reading everything relevant, which in this case means dozens of experimental papers by psychologists and economists. (He soon had compiled a massive bibliography of papers on psychology and economics and induced our friend George Loewenstein to describe him as a human "Pac-Man,"

Exhibit 1
Papers by Matthew Rabin Cited in this Essay

1. "Reneging and Renegotiation," Berkeley Department of Economics Working Paper No. 91-163, April 1991.
2. "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, pp. 1281–302, December 1993.
3. "A Model of Pre-Game Communication," *Journal of Economic Theory*, 63, pp. 370–91, August 1994.
4. "Cognitive Dissonance and Social Change," *Journal of Economic Behavior and Organization*, 23, pp. 177–94, 1994.
5. "Moral Preferences, Moral Constraints, and Self-Serving Biases," Berkeley Department of Economics Working Paper No. 95-241, August 1995.
6. "Daniel Kahneman and Amos Tversky," in *American Economists of the Late Twentieth Century*, Warren Samuels, ed., Cheltenham, U.K.: Edward Elgar Publishing Ltd., pp. 111–37, 1996.
7. "Deviations, Dynamics, and Equilibrium Refinements" (with Joel Sobel), *Journal of Economic Theory*, 68:1, pp. 1–25, January 1996.
8. "Cheap Talk" (with Joseph Farrell), *Journal of Economic Perspectives*, 10:3, pp. 103–18, Summer 1996.
9. "Psychology and Economics," *Journal of Economic Literature*, 36, pp. 11–46, March 1998.
10. "Choice Bracketing" (with Daniel Read and George Loewenstein), *Journal of Risk and Uncertainty*, 19:1-3, pp. 171–97, December 1999.
11. "Doing It Now or Later" (with Ted O'Donoghue), *American Economic Review*, 89:1, pp. 103–24, March 1999.
12. "First Impressions Matter: A Model of Confirmatory Bias" (with Joel Schrag), *Quarterly Journal of Economics*, 114:1, pp. 37–82, February 1999.
13. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68:5, pp. 1281–292, September 2000.
14. "Risk Aversion" (with Richard Thaler), *Journal of Economic Perspectives*, 15:1, pp. 219–32, Winter 2001.
15. "Choice and Procrastination" (with Ted O'Donoghue), *Quarterly Journal of Economics*, pp. 121–60, February 2001.
16. "Projection Bias in Predicting Future Preferences" (with George Loewenstein and Ted O'Donoghue), Cornell University Working Paper, July 2002.
17. "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, forthcoming.
18. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'" (with Colin Camerer, Samuel Issacharoff, George Loewenstein and Ted O'Donoghue), *Penn Law Review*, forthcoming.

who gobbles up psychology.) A key insight from his readings was that people are neither selfish nor altruistic all the time. Rather, people engage in a type of reciprocal fairness: they are nice to people who are nice to them, but mean to people who harm them. The next question was how to capture this property in an economic model.

Matthew's trick is a "kindness" function, which measures whether one player is being nice or mean to another player. Niceness is a positive number and meanness is negative. Players get utility from material payoffs and from the *product* of how nice they are and how nice the other player is. This specification makes players want to behave nicely toward nice guys and *also* behave meanly toward jerks (since the product of two negative kindnesses is positive—revenge against an enemy increases utility).[1] Given

[1] Technically, kindness is defined by assuming that a player A's move, given A's belief about what B will do, effectively awards the other player B a payoff. Define a "fair" or kindness-neutral payoff as, say, some

Table 1
Prisoner's Dilemma with Fairness-Adjusted Payoffs

		Column C	*Column* D
Row	C	$4 + .75\alpha, 4 + .75\alpha$	$0 - .5\alpha, 6$
	D	$6, 0 - .5\alpha$	$0, 0$

these fairness-adjusted payoffs, Rabin then applies standard equilibrium concepts.[2] Some empirical bite comes from the assumption that the fairness terms are independent of the scale of payoffs, so that fairness becomes less important as monetary payoffs get larger.

A familiar example is the prisoner's dilemma. Table 1 shows equilibrium payoffs adjusted for fairness.[3] The Cooperate-Cooperate (C,C) money payoffs are (4,4), the Defect-Defect (D,D) payoffs are (0,0), and the Cooperate-Defect payoffs are (0,6). The term α is the relative weight on fairness compared to money.

Consider the Row player's payoffs. If she chooses C when the Column player chooses D, she suffers a fairness penalty of $-.5\alpha$ because Column played meanly. Conversely, if she expects the Column player to choose C, then reciprocating cooperation by also choosing C is nice (because it gives Column a better payoff than if Row defected). Column is also being nice by not taking advantage and defecting; the product of the mutual niceness adds a fairness bonus of $.75\alpha$ to the material payoff of 4, yielding a total payoff for Row in the (C,C) cell of $4+.75\alpha$. Thus, mutual cooperation is a fairness equilibrium when α is large enough (specifically, where $4 + .75\alpha$ or $\alpha > 3.33$). In this view, the prisoner's dilemma is a coordination game in which players try to coordinate their emotions or levels of niceness. This interpretation jibes nicely with the experimental observation that players are often "conditionally cooperative"—those who cooperate tend to expect others to cooperate too.

By transforming the prisoner's dilemma into an emotional coordination game with multiple equilibria, Rabin's approach can explain why there is a deep indeterminacy in what will happen. The same group of people might play cooperatively if they expect niceness, but might defect if they expect defection. This property of

---

average of the most and the least that A can "give" B through her own move. Then kindness is just the difference between the payoff that A's move gives to B and the fair payoff, suitably normalized so that kindness is between -1 and $\frac{1}{2}$. (Kindness does not lie between -1 and $+1$ for technical reasons.)

[2] Rabin used the "psychological games" approach for which Geanakoplos, Pearce and Staccheti (1989) deserve credit.

[3] The table is an abuse of notation because fairness-adjusted utilities depend on one's strategy, beliefs about others' strategies and beliefs about beliefs. Think of the adjusted payoff in each cell as the equilibrium payoff if choices, beliefs and iterated beliefs match. The upper right row payoff $0 - .5\alpha$ for example, is Row's fairness-adjusted payoff if she plays C, believes Column will play D and believes Column believes she will play D.

Table 2
Chicken with Fairness-Adjusted Payoffs

		Column Dare (D)	Chicken (C)
Row	Dare	$-2, -2$	$2, 0 - .5\alpha$
	Chicken	$0 - .5\alpha, 2$	$1 + .75\alpha, 1 + .75\alpha$

the model gracefully accommodates the empirical effects on cooperation rates of many variables that do not directly affect payoffs and that have long puzzled economists. For example, the variable that most changes in cooperation rates in a one-shot prisoner's dilemma game is not game-theoretic training, culture or payoffs—it is the ability of players to exchange costly, nonbinding communication that theorists call "cheap talk" before playing (Sally, 1995; Ledyard, 1995). Cheap talk can have an effect, in Rabin's model, because it addresses the players' central coordination problem—are we going to be nice to each other or not? Saying you will cooperate can shift another player's beliefs, which can lead her to cooperate and then make you want to cooperate as you promised to. (In game-theoretic jargon, promising to cooperate is a correlating device.) Similarly, cheap talk tends to shift behavior toward efficient outcomes in other coordination games (Camerer, 2003, chapter 7).

Another variable that is irrelevant in standard game theory is how the prisoner's dilemma game is described to players. In experimental studies, if the prisoner's dilemma is described as a "community game," players cooperate more than if it is described as a "Wall Street game" (Ross and Ward, 1996). In Rabin's approach, labels can matter just as they do in Schelling's (1960) famous "focal point" examples—because labels can influence players' expectations about whether *others* will behave nicely or meanly. Indeed, in Rabin's fairness-adjusted prisoner's dilemma, those expectations are self-fulfilling because players are trying to match their niceness with others.

Another interesting example is "Chicken,"[4] with fairness-adjusted payoffs shown in Table 2. In this game, the two choices are to play "Chicken" or "Dare." The Nash equilibria are (C,D) and (D,C)—one player backs down and chooses C(hicken) when the other is expected to choose D(are). However, when the fairness factor α is large enough (above 4), a player who expects the other to pick D would rather choose D, earning -2, than choose C and let the mean player take advantage, suffering a loss of $\$.5\alpha$ Similarly, when $\alpha > 4/3$, then reciprocating the choice of Chicken with Chicken pays, because it repays the other player's kindness. Thus, if fairness effects are large, then either (C,C) and (D,D) are fairness equilibria.

[4] Many of the small gems to be found in reading Matthew's papers are in the footnotes. For example, in this paper, when he introduces the game "Chicken," he offers the following footnote: "While I will stick to the conventional name for this game, I note that it is extremely speciest—there is little evidence that chickens are less brave than humans and other animals."

This example is important because the set of outcomes allowed by fairness is *completely the opposite* of the standard equilibrium outcomes. In this sense, Chicken is the *best* game to use to contrast fairness and pure self-interest, a better game than ultimatum bargaining, prisoner's dilemmas and other games that have been much more thoroughly studied. The game also captures both the mutually happy and mutually angry aspects of social preference, like a couple who sacrifice to please each other, only to end up in an ugly "War of the Roses" divorce in which their sole goal is to harm the other person who harms them. The off-diagonal cells are the thin line between love and hate. Moreover, experimental evidence shows that the Chicken-Chicken and Dare-Dare outcomes are more likely than the Nash outcomes (C,D) and (D,C) (Rutstrom, McDaniel and Williams, 1994).

An intuitive way to think about Matthew's fairness model is that it combines two forces: material payoffs and the concern for mutual fairness that leads (in equilibrium) to mutual kindness *or* mutual unkindness. In the prisoner's dilemma, material concerns lead to mutual defection, while fairness leads to either both parties cooperating or both parties defecting—but undermines the asymmetric (Cooperate, Defect) outcomes. In Chicken, material concerns lead to the asymmetric outcomes of (Dare, Chicken) or (Chicken, Dare), but fairness undermines these outcomes and favors either (Dare, Dare) or (Chicken, Chicken).

Rabin proves several propositions about the existence and characterization of fairness equilibrium. He also shows that firms cannot sustain monopoly prices because fair-minded buyers will "reciprocate" what they perceive as unfair price-gouging by withholding demand. (This result happens in experiments, too, as in Ruffle, 2000.) An application to gift exchange in employment also shows how a worker might reciprocate a high wage with high effort, even when workers are free to shirk and there are no reputational advantages to working hard, as observed in many experiments (for example, Fehr and Gächter, 2000).

Matthew's fairness paper also highlights another characteristic of most of his work: He is fearless about proposing a model that is provocative and an important start, but admittedly wrong in some dimensions. His paper includes a long discussion of the model's flaws, which is both humble and a wise strategy for fending off criticism by preempting it.

The Pre-Behavioral Rabin

Before taking the plunge with his fairness paper, Matthew had been honing his craft as a game theorist working on topics such as communication and signaling that were, at that time, closer to the mainstream in game theory. These early papers often show flashes of the same flair for invention that characterizes all of Matthew's work—his ability to invent useful new constructs.

For example, his paper on "Reneging and Renegotiation" [1] contributes to the literature on renegotiation in the theory of repeated games and introduces

formal concepts of temptation (the biggest short-term gain from cheating on the implicit agreement) and credibility (if a player cheats once his credibility is shot).

Another early interest of Rabin's is communication in games. Before the 1980s, theorists suspected that either preplay communication didn't matter at all (unless promises are binding) or that communication could guarantee equilibrium and efficiency. A flurry of research in the 1980s and 1990s suggested that the theoretical value of communication lies in between those extremes. In a paper in this journal, Farrell and Rabin [8] summarize what was learned. When players have a common incentive to communicate private information, then they are likely to do so. But when players' incentives are not aligned, the effects of communication are severely undermined. It is often impossible to rule out "babbling equilibria" in which players expect their messages to be ignored, so they have no incentive to say anything meaningful. To refine out these equilibria, either evidence, a theory of meaning or some auxiliary assumptions are needed. For example, Rabin [3] tackles this problem by considering the effects of communication about intentions. The bad news from this literature is that a limited amount of communication may not help the players. The good news is that communication *eventually* does help. As Farrell and Rabin [8, p. 116] put it, "[T]he worst a player can do is to give up and say to the other, 'You choose,' which at least assures enough communication can always keep players from slipping into an outcome which is bad for both of them." Costa-Gomes (2002) showed how data from many experiments corroborate Rabin's prediction.

In another paper on communication much admired by formal theorists, Rabin and Sobel [7] tackle a subtle problem in the logic of signaling games. Signaling games often have implausible equilibria because they presume that if a sender does something unexpected, then the receiver makes a strange guess about what "type" of player the sender might be. For example, in simple models of education as a signal of ability, there can be equilibria in which nobody gets educated because people who get educated are thought by employers to be dumb rather than smart; so it doesn't pay for smart people to go to college. Rabin and Sobel present a theory of play in which sender types are tempted to deviate from this equilibrium and show which equilibria are viable after this theory of play. The results are surprising, deep and required a lot of new mathematical machinery to prove.

The Psychological Rabin, Post-Fairness

After the publication of the fairness paper [2] in the *American Economic Review*, Matthew turned his attention almost exclusively to psychology and economics, and we devote the rest of this paper to that work.

Quasi-Bayesian Expectations

Bayesian updating is at the heart of most information economics and, more generally, any model that incorporates rational expectations. A problem for

economists has been that even if they were inclined to believe that agents do not typically make Bayesian forecasts, there has not been an alternative model that is both tractable and descriptive. Matthew's contribution is a "quasi-Bayesian approach" in which some cognitively plausible error or miscalculation is permitted, and then the technical apparatus of Bayesian updating is applied to the miscalculated numbers.

His paper with Schrag on "confirmatory bias" [12] illustrates the quasi-Bayesian approach. They first point out evidence suggesting that people are prone to misread vague evidence as confirming, rather than disconfirming, what they believe. To model this, they assume there are two hypotheses, A and B, and the diagnosticity of signals is a parameter θ; that is, conditional on A being true, a supportive signal is generated with probability θ. Therefore, $\theta = 0.5$ means signals don't tell you anything. Rabin and Schrag start with the premise that if people think hypothesis A is more likely than B, then evidence that is consistent with A is always encoded correctly, but evidence that disconfirms A is mistakenly encoded as being consistent with A q percent of the time ($q = 0$ is just Bayes' rule with no bias). This kind of encoding bias is deeply rooted in perception and can be viewed as adaptive behavior when faced with cognitive constraints. People literally do not process brief images they do not expect; and conversely, after buying a new car, it is common to notice more examples of the same car while driving around.

Some interesting results follow. For example, feedback and experience can make people who started out believing in A overconfident about A. The most striking result is that, if the signals are weak enough, then even a small degree of confirmation bias keeps agents from *ever* figuring out the truth, even with infinitely many signals. A fashionable theory in financial economics to explain the massive volume of speculative trade is that agents have different "priors" (Harris and Raviv, 1993). But agents in financial markets who have been trading for years don't hold priors, per se—in Bayesian language, they are posterior beliefs, not priors. For example, many money managers have been firm believers in "value" strategies for over 20 years, while many others have believed in "growth" strategies for the same length of time. As stocks go in and out of favor, the two groups trade with one another. Rabin and Schrag's theory puts a firm foundation beneath these differences-of-opinion stories by explaining precisely how two people who begin work on Wall Street with different ideas about how markets work can persistently disagree throughout their entire careers. Confirmation bias can also explain how an incorrect theory (pick your favorite example to insert here) can persist for decades or even centuries in spite of massive contrary evidence.

Law of Small Numbers

Matthew also applies the quasi-Bayesian approach in his paper [17] on the "law of small numbers." This "law" is a term coined by Kahneman and Tversky, half-facetiously, to describe the human tendency to jump too quickly to conclusions from small samples. The law of small numbers is illustrated by the "gambler's fallacy"—the mistaken (and irresistible) belief that a roulette wheel that has come up red several times in a row is "due" for a black outcome. This mistake is

manifested in actual betting behavior. Consider New Jersey's pick-three numbers game, in which everyone who bet on the winning three-digit number shares the pool (minus the state's healthy cut of half). A couple of days after a particular number wins, bets on that same number dip to about 60 percent of their typical level, then slowly rebound after about eight weeks (Terrell, 1994). Racetrack bettors also bet less on unlikely "longshots" in later races if a longshot won earlier in the day (Metzger, 1985), as if "lightning doesn't strike twice."

Rabin models the law of small numbers by assuming that people mistakenly conceive of independent Bernoulli trials as draws *without replacement* from an urn with N balls. (The model generalizes Bayes' rule because it reduces to Bayes when N is infinite.) The model is custom-built to capture the gambler's fallacy, because sampling without replacement creates negative autocorrelation between draws. But the model also predicts "over-inference bias." For example, suppose biased investors are considering mutual funds. They will be surprised by a streak of successes and will mistakenly conclude that a fund with a short winning streak must be good. On the opposite end, they will write off a fund as a loser too quickly after a short slump. This kind of "overinference" leads to "false variation bias," or a belief that the dispersion of true skill is greater than it really is. Rabin also shows that these mistakes can lead people to switch funds too frequently. In addition, "overinferrers" become pessimistic because their fund history tends to be dominated by losers that were only held briefly, so they think most funds are no good.

We think this model is a big part of the explanation for why there are so many mutual funds (more than the number of traded stocks) and why new ones are created all the time. It could also be applied to domains like the market for managers, sports coaches or even mates. Rabin also shows how the model can explain the well-known fact that stock markets underreact in the short-term and overreact in the long-term to public information such as earnings announcements. In Rabin's model, investors expect mean-reversion, so they aren't impressed by a short streak of good earnings (they underreact); but they also overinfer that a firm must be terrific after a long streak of good earnings, and they overreact (Barberis, Shleifer and Vishny, 1998).

Rabin's law of small numbers paper illustrates an important feature of his modeling: The models not only express a psychological regularity (the law of small numbers), they actually *do* new psychology, by predicting a fresh bias (false variation). Psychologists had not predicted this bias, perhaps because the insight comes from mathematics, which is not a standard engine of discovery in psychology. Rabin also provides a unifying explanation of why there can be both a gambler's fallacy in betting on lottery numbers and a persistent belief in positive autocorrelation—a mythical "hot hand" in sports performance (Gilovich, Vallone and Tversky, 1985). (Contrary to popular belief, basketball players who hit a couple shots in a row are *not* more likely to hit their next shot, even in free-throw shooting or experiments where there is no defensive adjustment to cancel out performance momentum.) Economists wonder why people don't learn to correct their mistakes, and Rabin's models can explain why. Mistakenly expecting outcomes to even out, then observing

that they do not, can lead to switching from one mistake to another and coming to believe that the time series must have "momentum." As Rabin explains, "Faced with actual independence of signals, people develop a bogus belief in a form of positive autocorrelation in signal generation that to them explains the missing negative autocorrelation they expected due to gambler's fallacy."

Risk Aversion

Risk aversion is a cornerstone concept of economics. The traditional explanation for risk aversion has been based on the diminishing marginal utility of wealth. Since von Neumann and Morgenstern, the standard practice in economics has been to assume agents maximize expected utility, and the traditional way of implementing that assumption is to assume agents maximize the expected utility of final wealth. What most economists have failed to realize is that this standard approach is incompatible with much of the behavior it is used to model. This result is demonstrated in Rabin's short *Econometrica* paper [13].

Take the example of someone being offered a coin flip bet: heads, he wins $105; tails, he loses $100. This bet has little appeal to most people; if you have any doubt of this fact, ask the students in your class, or your spouse, hairdresser or plumber. When confronted with such commonplace behavior, economists haul out their standard explanation and say that this behavior is "risk aversion." Most of us have drawn the familiar concave utility of wealth curve, labeling W and drawing a linear chord below it, to explain to our students why this framework is obviously the explanation for such behavior.

Matthew made calculations about what such a choice would imply about that utility of wealth function. His insight was that curves that are slightly concave locally are *extremely* concave globally. How concave? What he shows is that if someone turns this bet down at all wealth levels, then that person will also turn down an opportunity to flip a coin in which she risks losing $2,000, but stands to gain all of Bill Gates's fortune and then some (more precisely, an infinite amount). Of course, no one with nontrivial wealth (or borrowing power) would turn the second bet down, but most will turn the first bet down. There is no trick here, so as economists, we have a problem. To a first approximation, agents who maximize their expected utility of wealth must be virtually risk neutral for bets that are small relative to their wealth. Therefore, when we see people declining attractive smallish bets, the explanation must be something other than the one normally used.

Do people turn down bets of the $105/$100 variety? Yes, frequently. There is voluminous experimental evidence of such behavior using both real and hypothetical choices; for a recent illustration, see Holt and Laury (2002), who find that the risk aversion displayed in answers to hypothetical questions becomes even more pronounced for real stakes. Behavior outside the lab is also abundant. Consider the behavior of buyers of automobile collision insurance. Based on data from 1994–1996, more than half of the purchasers of collision insurance elected a deductible of $250 or less. A typical consumer could save about $80 a year by increasing the deductible from $250 to $500 (Grgeta and Thaler, 2003). Similarly, extended

warranties for small-ticket items such as cell phones is a big-ticket business for major electronics retailers. Unless we posit that retailers and insurance companies are selling these policies at a loss, it is difficult to explain why risk-neutral consumers would buy them.

Rabin's conclusion is that economists need something other than expected utility of wealth to explain what has been conventionally interpreted as risk aversion for moderate stakes. He thinks (and we concur) that one ingredient that is necessary to explain this behavior is what Matthew has called "piecemeal preferences." The idea is that decisionmakers tend to view decisions one at a time, and independent of the rest of their life, rather than incorporating the decision at hand with the rest of life's portfolio of risks. This behavior is also called myopic loss aversion, narrow framing or bracketing, or decision isolation. Matthew joined Daniel Read and George Loewenstein to write a comprehensive survey of this topic [10].

Procrastination and Self-Control: Doing it Now or Later

Matthew has written several papers on procrastination and self-control with Ted O'Donoghue. In contrast to much of Matthew's other behavioral theorizing, Matthew and Ted were relative latecomers to this particular party. Early work by Strotz (1956) and Pollak (1968) introduced the idea of time-inconsistent preferences, and David Laibson (1997) had recently rekindled interest in this topic with an explicitly behavioral approach. So why did Matthew take up this topic? We suspect that as a self-proclaimed avid procrastinator, he couldn't resist the temptation to dig into the problem (eventually).

The point of departure for this research is the observation that people exhibit a specific type of time inconsistency that O'Donoghue and Rabin [11] have dubbed present-biased preferences and what Laibson calls quasi-hyperbolic discounting. (These two models have identical structures.) In choosing between a smaller reward right now versus a bigger one tomorrow, many people opt for the small immediate reward. However, if the choice is between a smaller reward in 30 days and a larger reward in 31 days, the same people choose the larger award. This behavior produces *dynamic inconsistency,* because one choice is preferred if it is off in the future, but the other choice is preferred in the present. An agent who has present-biased preferences predictably changes her mind over time. Agents who discount the future using exponential discounting, as normally assumed in economic theory, do not display this inconsistency. The now standard approach to modeling such present-biased preferences is the so-called beta-delta model. In this model, preferences can be represented by

$$\text{For all } t,\ U^t(u_t, u_{t+1}, \ldots, u_T) \equiv \delta^t u_t + \beta \sum_{T=t+1}^{T} \delta^T u_T \qquad \text{where } 0 < \beta,\ \delta \leq 1.$$

In this model, δ represents long-run time consistent discounting, whereas β represents the bias for the present. If $\beta < 1$, then preferences are present biased.

A question that has long puzzled theorists is whether agents are aware of this predictable pattern in their own behavior. Most economists assume that people are "sophisticated" about their time inconsistency and so will seek external self-control to limit their own future behavior. O'Donoghue and Rabin were surprised that economists had rarely explored the implications of the alternative assumption that agents "naively" assume their current preferences over future options will remain constant (for an exception, see Akerlof, 1991). So in their early papers, O'Donoghue and Rabin spell out the implications of the assumptions of extreme sophistication and naiveté. They find, surprisingly, that naive agents need not behave in crazy ways, and they sometimes even take actions that seem more sensible than those sophisticated agents take.

O'Donoghue and Rabin [11] consider three types of agents: those who are *time consistent*, *sophisticates,* who realize their preference change over time, and *naifs*, who think they are time consistent.[5] They then consider a series of Rabinesque examples, beginning with this one: "Suppose you usually go to the movies on Saturdays, and the schedule at the local cinema consists of a mediocre movie this week, a good movie next week, a great movie in two weeks, and (best of all) a Johnny Depp[6] movie in three weeks. Now suppose you must complete a report for work within four weeks, and to do so you must skip the movie on one of the next four Saturdays. When do you complete the report?" The example is made precise by assuming that $\delta = 1$, $\beta = 1/2$ and the valuations of the mediocre, good, great and Depp movies are 3, 5, 8 and 13.

The time consistent agents, who always do the rational thing, skip the movie in the first week and do the report. The naifs procrastinate until the last Saturday. Each week they (wrongly) think they will skip the movie the following week, so on the last Saturday they have to finish the report and never get to see the best movie. Counterintuitively, the sophisticates do the report in week *two* rather than in week one like the time consistent agents. Here's why: The period-1 sophisticate correctly predicts that he would have self-control problems on the third Saturday and would see the great movie, forcing him to do the report on the last Saturday at a high opportunity cost (if he hadn't done it earlier). However, the period-1 sophisticate also correctly predicts that knowing about period-3 self-control problems will induce him to do the report on the second Saturday. So the period-1 sophisticate can safely procrastinate for one week.

If you think this result is odd, consider the next example. Now you have a coupon to see exactly one of the movies in the next four weeks, and your allowance does not permit you to pay for a movie. Which movie do you see? Predictably, the time-consistent agents will wait and see the great Depp movie. Naifs see the merely great (value 8) movie: On the first two Saturdays, they skip the mediocre and good

[5] In the early draft of this paper, these types were labeled O'Donoghues, Laibsons and Rabins, respectively. Without going into details here, let's just say that the labels are apt.

[6] For a picture of Johnny Depp, look at Rabin's website and click on "picture of Matthew on a good hair day."

movies incorrectly believing they will wait to see the Depp movie. However, on the third Saturday they give in to temptation to go to the merely great movie. What about the so-called sophisticates? Under the concrete values assumed in this problem, they see the worst movie, in the first week! The period-2 sophisticate realizes that his period-3 counterpart will give in to temptation and see the merely great movie, so the period-2 self will choose to go to the good movie. But the period-1 sophisticates anticipates this behavior by his period-2 self, and so he decides to attend the mediocre movie. The ability to do backward induction, combined with realism about future self-control problems, dooms the sophisticate to see a worse movie than the naif does.[7]

It turns out that being sophisticated is not all it is cracked up to be. Sophisticates sometimes complete an unpleasant task sooner than they would if they had no self-control problem, but may end up consuming tempting goods later than they would if they had no self-control problems. This leads O'Donoghue and Rabin to investigate models somewhere between naive and sophisticated. A conception they pursue in recent work has agents realize they have self-control problems, but are too optimistic about their ability to resist temptation in future periods. ("I know I have slipped up occasionally before, okay, more than occasionally, but next time will be different.") This approach is sensible to pursue and is consistent with some empirical evidence about deadline setting and health club pricing (Ariely and Wertenbroch, forthcoming; Della Vigna and Malmendier, 2002).

Projection Bias

Everyone has heard that if you go to the grocery store famished, you buy more food. It turns out that this is not an old husband's tale. One study finds that if shoppers are given a muffin to eat before shopping, they buy fewer impulse items (items not on their shopping list) than those who were not given a muffin (Gilbert, Gill and Wilson, 2002). But how can this pattern persist if we all know about it? The answer is that we must not be completely "sophisticated" about this phenomenon; that is, even though we know we buy more when we are hungry, we underestimate just how much more.

Loewenstein, O'Donoghue and Rabin [16] dub this phenomenon "projection bias." Projection bias means we think that our preferences in other states will be closer to our current preferences than they actually will be. If we go into a restaurant feeling like we haven't eaten all day, we order a rich soufflé, underestimating how stuffed we might feel when the soufflé arrives. Their paper argues that this bias about imagining how hungry or full we feel in the future extends to many other domains.

The projection bias has much in common with the present-biased model of

[7] The fact that foresight can undermine the sophisticates' self-control is prominent in clinical studies of addiction. Addicts often relapse because they reason that some day they will fall off the wagon, so they might as well start right away. Combating this belief is why Alcoholics Anonymous encourages drinkers to stay sober "one day at a time."

intertemporal choice. In the present-biased model, people overweight the utility of current consumption relative to future periods. With projection bias, people overweight the relevance of their current state (hungry, aroused, excited, depressed) in their attempts to predict their preferences in future states. If we are depressed now, we find it hard to imagine that we will ever be in the mood to go to a party. If we are not aroused now, we find it hard to believe that we would fail to practice safe sex when we are aroused. The model proposed by Loewenstein, O'Donoghue and Rabin captures these intuitions and discusses examples that include overconsumption early in the life cycle (underappreciating habit formation) and drug addition. They then discuss welfare implications of the model. One clear policy implication is support for "cooling off periods." If people tend to buy a car or vacation time share property in the heat of the moment, then regret it soon after when they have calmed down, a three-day cooling off period offers them a chance to regain their senses (and does little harm if they made a good choice they don't want to reverse).

Morality

It is hard to spend much time in the "People's Republic of Berkeley," as Matthew does happily, and not think at least a little about morality. Matthew has two papers that do so.

One paper [4] is about cognitive dissonance and social change. Consider some morally dubious activity such as littering, writing late referee reports or eating more than your share of the jumbo shrimp at a cocktail party. Rabin assumes that people select a level of this activity, say X, although they believe the morally legitimate level is Y, which may be less than X. People get utility from activity level X, suffer painful "cognitive dissonance" if X is greater than Y, and also incur a psychic cost (or conformity penalty) if their moral belief Y is different from the average moral belief in the community, Y'. He shows, counterintuitively, that increasing the disutility from dissonance can backfire and lead people to do *more* of the bad activity X. Why? Raising the dissonance—by social shaming or public-service ad campaigns, for example—does have direct effects of lowering the level of the activity X and the morally legitimate norm for the activity Y. But it also has an indirect effect through the psychic cost. Since it is painful to reconcile one's belief with an accepted norm, one solution to the suggestion that Y should be low is for everyone to give up and backslide morally, choosing a higher value of Y', which then licenses people to believe Y is acceptable and choose a higher level of X. Rabin shows that this outcome can be an equilibrium. Take energy-hogging sports-utility vehicles as an example. Moral preaching that sports utility vehicles are bad could result in an equilibrium where people rebel, regard sports utility vehicles as less bad than before (because the psychic cost of reconciling pro-SUV beliefs is too high for everyone) and drive more sports utility vehicles. Another example is public looting. Even if everyone thinks looting is wrong, *how wrong* it is—or more precisely, how wrong it *feels*—depends on how many others are looting. One response to increased moral scolding is for everyone to think others might loot more and, due to the conformity effect, then everyone *will loot* more. While the model is stylized, it both

follows Rabin's colleague George Akerlof's interest in drawing social phenomena into economics and anticipates the recent interest among many other economists in social norms and influence.

In the second paper on moral preferences [5], Rabin draws an interesting distinction between agents who trade off raw preferences and moral concerns and agents who make choices according to moral "rules." Matthew first shows that behaviors guided by distaste for acting immorally, or by rules, are observationally equivalent. For example, working as a prostitute might result from simply hating poverty more than sin, as a madam once famously remarked, or by having a moral benchmark that permits prostitution. This observational equivalence has allowed economists to remain comfortable modeling such choices as the result of smooth tradeoffs ("there's a price for everything") rather than bound by moral constraint, though the latter viewpoint is popular among sociologists and philosophers.

But Matthew shows that preference- and rule-based modeling approaches are not exactly the same, in a subtle and surprising way. Suppose agents aren't sure how morally dubious an activity is, but can find out before they make a choice. (Think of those earnest modern hippies who ask a restaurant waiter whether the ahi tuna was line-caught rather than netted—since netting harms porpoises.) Preference-guided people dislike feeling that they are doing harm, so they will be more willing to find out the true harm they are causing and won't do the activity if they find out it is too harmful. Rule-bound people will, in contrast, either i) protect their ability to do the behavior by not finding out, if they plan to do it or ii) will seek information to permit themselves possibly to do it even if they think they probably shouldn't. (Perhaps they normally don't eat ahi tuna, but the diner at the next table seems to be enjoying hers quite a bit!) As a result, the rule-bound agents will cause more social harm than preference-driven agents, because the rule-bound agents sometimes avoid finding out how bad their behavior can be.[8]

This difference provides a rationale for social policy that forces people to find out about the consequences of their actions, since rule-bound people will not seek information, and their reluctance to find out is socially harmful. A dramatic modern example is AIDS testing. People unknowingly infected with AIDS may put others at risk by sharing needles or engaging in high-risk sex. Rabin's model explains why many of these people don't want to be tested to see if they have the disease: If they tested positive, they would have to stop doing something pleasurable to themselves (and bad for others); but without testing, they can do so without qualms (if they think they are low risk). Such questions are central to many economic questions: half of sub-Saharan Africa is infected with HIV; the infection rates are growing ferociously in Russia and other countries; and infection rates remain high in many other countries or subpopulations, like certain American cities.

[8] See Dana, Weber and Kuang (2003) for an interesting experiment that illustrates why search for information distinguishes moral rules and preferences.

The Character of Matthew

Matthew has been working on a book tentatively entitled *Psychological Foundations of Economic Theory*. How he came to write that book sounds like an example from one his papers. Once upon a time, he was asked to write an essay on Daniel Kahneman and Amos Tversky [6]. Matthew began with characteristic seriousness and produced a draft that came in at about 100 pages, roughly four times too long for the book. During the writing of this draft, Matthew undoubtedly felt that he was not working on this chapter, but rather was procrastinating about doing something else, something more important, like getting tenure. (For years, he had a cardboard clock in his office that was labeled the "tenure clock.") He later turned the long version of the chapter into a 1998 survey for the *Journal of Economic Literature* [9], but even that published survey was much trimmed down from its longer version. The much-too-long version seemed like a good start on a book. When this book is complete, we can say with confidence that it will be the place students begin if they want to learn about psychology and economics. And if it takes Matthew a long time to finish, he can say he was doing empirical work on procrastination the whole time!

We have tried here to give readers a sense of what Matthew's research has been about, and we have dropped some hints about what a character he is. We would need to be novelists to do full justice to the latter topic. Matthew is, well, complicated. He loves bright colors. At workshops, he always has a collection of colored pens and colorful toys to fiddle with. His hand-drawn overhead slides are crammed with colors. (A good sign that the world is ending will be when Matthew adopts PowerPoint.) He usually wears very loud tie-dye T-shirts purchased on Telegraph Avenue in Berkeley. He may get some of the credit or blame for the fact that there are still so many vendors in that locale.

Matthew is simultaneously very serious about his research, and very funny about everything else. The working title of one of his early papers [3] (later published in the *Journal of Economic Theory*) was "Preplay Communication: A Lengthy-Paper Approach." He jokes that he became a game theorist because his father played "rock, paper, scissors" with real rocks, and he wanted to lose less often. Although he is a fine speller, and a fanatic about the proper use of hyphens, he never signs his name at the end of e-mails quite correctly (some recent examples include: maththew, mathweh, mathtehw, mathhew). He is a big fan of *Monty Python* and the *Simpsons*, and manages to work suitable references to them into his work. Rabin and Thaler [14] made use of the famous *Monty Python* "dead parrot" sketch in an "Anomalies" column in this journal based on Matthew's risk aversion research. In a recent law review paper [18] on paternalism, Matthew coined the term "faith-based anti-paternalism" to refer to the view that paternalism is by definition bad since people always choose what is in their best interest. He then managed to insert as footnote 97 the following commentary on whether people who buy extended warranties should be considered idiots: "In a classic *Simpsons* episode, Homer was having a crayon hammered into his nose to lower his I.Q. (Don't ask.)

The writers indicated the lowering of his I.Q. by having Homer make ever stupider statements. The surgeon knew the operation was complete when Homer finally exclaimed: 'Extended Warranty! How can I lose?'"

Matthew is also a very sweet and modest person. (Empirical proof is that he'll be embarrassed to read this article.) He is extremely generous with this time to both students and colleagues (though he probably thinks of it just as more procrastination).

Matthew the person is important as an intellectual leader who showed how to begin the long task of making economic models more realistic and better. Because economists have been building models on the foundations of optimization and equilibrium for a long time, no one will be able to come up with an alternative approach that does as much with so little. The research agenda of adding psychological realism to economics is one of complicating the model to incorporate the complex truths about people, much as other domains of economics have complicated the models to incorporate complex truths about institutions. The idea that firms maximize value is simple. Adding principal-agent problems and information asymmetries adds realism at the cost of simplicity; but there is no doubt that adding these complications has proved worthwhile. Similarly, incorporating bounded rationality, self-control problems and other-regarding behavior to simpler models of human behavior will make models more complicated, but improve their predictive power. (If not, don't complicate the model!)

We hope young economists will follow Matthew's lead. If they do, they will make economics more true and more fun.

■ *Thanks to Bob Gibbons, George Loewenstein, Ted O'Donoghue and Joel Sobel for helpful anecdotes and technical comments and to the* JEP *editors for doing their usual magic.*

References

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review*. 81:2, pp. 1–19.

Ariely, Dan and Klaus Wertenbroch. Forthcoming. "Procrastination, Deadlines, and Performance: Using Public Commitments to Regulate One's Behavior." *Psychological Science*.

Barberis, Nicholas, Andrei Shleifer and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal Financial Economics*. September. 49:3, pp. 307–43.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

Costa-Gomes, Miguel A. 2002. "A Suggested Interpretation of Some Experimental Results on Preplay Communication." *Journal of Economic Theory*. May, 104:1, pp. 104–36.

Dana, Jason, Roberto Weber and Jason Xi Kuang. 2003. "Do People Value Being Fair or Not Being Unfair? Behavior Inconsistent with 'Fairness Preferences.'" Working paper.

Della Vigna, Stefano and Ulrike Malmendier. 2002. "Self-Control in the Market: Evidence from the Health Club Industry." Working paper, University of California, Berkeley.

Fehr, Ernst and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of

Reciprocity." *Journal of Economic Perspectives.* 14:3, pp. 159–81.

Geanakopolos, John, David Pearce and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior.* March, 1:1, pp. 60–80.

Gilbert, Daniel T., Michael J. Gill and Timothy D. Wilson. 2002. "The Future is Now: Temporal Correction in Affective Forecasting." *Organizational Behavior and Human Decision Processes.* 88:1, pp. 430–44.

Gilovich, Thomas, Robert Vallone and Amos Tversky. 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology.* July, 17, pp. 295–314.

Grgeta, Edi and Richard H. Thaler. 2003. "Estimating Risk Aversion from the Purchases of Automobile Collision Insurance." University of Chicago working paper.

Harris, Milton and Arthur Raviv. 1993. "Differences in Opinion Make a Horse Race." *Review of Financial Studies.* 6:3, pp. 473–506.

Holt, Charlie A. and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review.* December, 92:5, pp. 1644–655.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics.* May, 112:2, pp. 443–77.

Ledyard, John. 1995. "Public Goods: A Survey of Experimental Reserach," in *Handbook of Experimental Economics.* J. Kagel and A. Roth, eds. Princeton: Princeton University Press, pp. 111–94.

Metzger, Mary Ann. 1985. "Biases in Betting: An Application of Laboratory Findings." *Psychological Reports.* June, 56, pp. 883–88.

Nisbett, Richard E. and David E. Kanouse. 1968. "Obesity, Hunger, and Supermarket Shopping Behavior." *Proceedings of the Annual Convention of the American Psychological Association.* 3, pp. 683–84.

Pollak, Robert A. 1968. "Consistent Planning." *Review of Economic Studies.* April, 35, pp. 185–99.

Ross, Lee and Andrew Ward. 1996. "Naive Realism: Implications for Social Conflict and Misunderstanding," in *Values and Knowledge.* T. Brown, E. Reed and E. Turiel, eds. Hillsdale, N.J.: Lawrence Erlbaum Associates, pp. 103–35.

Ruffle, Bradley J. 2000. "Some Factors Affecting Demand Withholding in Posted-Offer Markets." *Economic Theory.* 16:3, pp. 529–44.

Rutstrom, Lisa, Tanga McDaniel and Melonie Williams. 1994. "Incorporating Fairness into Game Theory and Economics: An Experimental Test with Incentive Compatible Belief Elicitation." Unpublished manuscript, University of South Carolina, Department of Economics.

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

Schelling, Thomas. 1960. *The Strategy of Conflict.* Cambridge, Mass.: Harvard University Press.

Strotz, Robert H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies.* 23:3, pp. 165–80.

Terrell, Dek. 1994. "A Test of the Gambler's Fallacy—Evidence from Parimutuel Games." *Journal of Risk and Uncertainty.* May, 8, pp. 309–17.