Journal of Economic Literature 2013, 51(2), 528–543
http://dx.doi.org/10.1257/jel.51.2.528

Incorporating Limited Rationality into Economics†

Matthew Rabin*

Harstad and Selten (this forum) raise interesting questions about the relative promise of optimization models and bounded-rationality models in making progress in economics. This article builds from their analysis by indicating the potential for using neoclassical (broadly defined) optimization models to integrate insights from psychology on the limits to rationality into economics. I lay out an approach to making (imperfect and incremental) improvements over previous economic theory by incorporating greater realism while attempting to maintain the breadth of application, the precision of predictions, and the insights of neoclassical theory. I then discuss how many human limits to full rationality are, in fact, well understood in terms of optimization. (*JEL* B49, D01, D03, D81, D84)

1. *Introduction*

In a thoughtful and well-hyphenated article, "Bounded-Rationality Models: Tasks to Become Intellectually Competitive," Harstad and Selten (this volume) lay out some challenges to neoclassical economics—and a refreshingly serious challenge to those trying to improve it. I offer some thoughts on their article and their challenge. The set of topics I emphasize differs significantly from theirs. This is due in part to different interests and knowledge. It is also because of division of labor: Crawford's contribution to this forum, "Boundedly Rational versus Optimization-Based Models of Strategic Thinking and Learning in Games," offers perspectives on a set of strategic and market topics (e.g., asset bubbles) closer to those topics focused on by Harstad and Selten. I focus more on limits to rationality that might help explain consumer and savings behavior, risky choice, and other classical problems of individual decision making.

In footnote 1, Harstad and Selten correctly note that many researchers trying to improve the psychological realism of economics do not necessarily see themselves as bounded-rationality theorists outside the realm of neoclassical theory, broadly defined. Nor even do we all eschew optimization models. Harstad and Selten themselves note that there are virtues to optimization as a methodology. Instead, their article is explicit in understanding its virtues, and lays out the challenge for those proposing alternatives to

* University of California, Berkeley. I thank Tristan Gagnon-Bartsch for valuable research assistance, and Vince Crawford, Janet Currie, Erik Eyster, Tristan Gagnon-Bartsch, and Ron Harstad for helpful comments.

† Go to http://dx.doi.org/10.1257/jel.51.2.528 to visit the article page and view author disclosure statement(s).

match some of those virtues. The articles by Harstad and Selten and by Crawford together provide a very useful and thoughtful presentation of the necessity and prospects for departing from optimization models in the set of topics they discuss. I agree with many of their arguments. As such, my article is not so much a rejoinder to Harstad and Selten's article as it is an addendum listing some current and potential research incorporating limited rationality into a more neoclassical framework. By discussing some principles and examples of current research, I hope to give readers a richer sense of the benefits and surprising possibilities for using optimization models to understand the implications of limits to rationality. In the process, my article will make more salient some types of human errors that seem downplayed by the community of scholars grappling with bounded-rationality models.

As the other articles in this forum note, much of the recent literature modifying neoclassical economics is about improving the realism of preferences by altering the utility function in ways that are fully compatible with rationality, and hence quite obviously conducive to models that assume people optimize.[1] This article outlines a perspective that may appear more surprising: that many *limits* to rationality can be usefully studied with optimization models. Some of the examples are rather Procrustean: the underlying psychology does not have a very optimizing feel to it, but can be crammed into the optimization framework to leverage the power of the neoclassical model.[2] But other examples of human limits to rationality, I will argue, are genuinely *best* understood in terms of optimization models. As surely as optimization captures cases of fully rational behavior, it captures the psychology and resulting behavior of some limits to rationality.

The case for optimization models of limited rationality largely reflects an insight from psychology that is underappreciated by economists: not all limits to rationality are based on computational unmanageability. Many of the ways humans are less than fully rational are *not* because the right answers are so complex. They are instead because the wrong answers are so enticing. Human intuition leads us astray in all sorts of ways that are simply not well described in terms of the difficulty or complexity of problems that bounded-rationality models seem best suited for. The pervasiveness of bounds errors where people are daunted by the task of optimization, or are simply not geared to it, should not be doubted. But astray errors, as one might call them, likewise seem pervasive, and especially amenable to neoclassical modeling. We can capture many errors in terms of systematic mistakes in the proximate value function people maximize (quasi-maximization models) or where the beliefs incorporated into their maximizations are systematically distorted (quasi-Bayesian models). The general sensibility and emphasis of economic theorists, experimental economists, and others lead them to neglect important examples of limited rationality that are important for economics.

[1] As Harstad and Selten note in the context of preference reversals, some of the psychologically inspired new models of preferences may have properties that economists had previously associated with irrationality and which may require a somewhat different toolbox. Fully rational theories of belief-based utility, such as Kőszegi (2010) and Kőszegi and Rabin (2006, 2009), generate time inconsistencies and violations of other axioms that economists had been in the habit of associating with rationality.

[2] The distinction between optimization models and bounded-rationality models is not so clear; at some level, they are probably isomorphic to the types of limits I discuss in this article. Nor is the labeling important. But even when they are distinct, we can still leave open the question of what models to use. And, especially in those cases where we all agree that the optimization models are wrong, the different articles constituting this forum will help see a range of perspectives on the costs and benefits of moving forward with models that are wrong but useful.

Studying astray errors rather than bounds errors allows methodological advantages that facilitate meeting the challenges that Harstad and Selten lay out. They directly suggest that we should not be too eager to move away from optimization models. It is easier to study astray errors than bound errors using optimization models because the former involve behavior that is richly purposeful and driven by a logic that the economic agent finds compelling. Fully rational investors who understand randomness and who want to retire wealthy will maximize their risk-adjusted expected savings with investment strategies that are best given their (correct) notion of randomness. But investors who believe in a false law of averages, expecting that good luck tends to be followed by bad and bad luck tends to be followed by good, will also maximize their risk-adjusted expected savings given *their* perceptions of randomness. Would-be exercisers who rationally forecast systematic taste change due to habit formation and due to fluctuations in energy will plan to carefully develop an exercise habit and will also know that they should not radically change plans in reaction to temporary fluctuations in enthusiasm. But those who systematically underappreciate taste change will not make any plans based on inducing good habits but may make unfortunate long-term plans (such as joining a gym) based on falsely extrapolating temporary enthusiasm. Elaborate, goal-driven plans are the stuff of savvy maximizers who correctly understand both the laws of statistics and their own future utility. But elaborate, goal-driven plans are also the stuff of those who both misunderstand statistics and systematically mispredict their own future behavior.

A second methodological advantage of this approach is that it embraces with enthusiasm Harstad and Selten's goal of making alternative models competitive with the current paradigm. These models use an approach discussed in Rabin (forthcoming), which has been employed in many papers: formulate models that offer complete mappings from environment to outcome, using as much as possible the same set of variables as in traditional models, and embed both the earlier (neoclassical) model and the limited-rationality model as parameter values. The competition between the old and new models is then half done—we've shown the two models to be structurally comparable in terms of degrees of freedom, applicability, etc. And the second half of the competition is put into motion: the models can be compared and judged, in a fair fight, by establishing point estimates and confidence intervals on the parameter values.

Section 2 explains and illustrates this approach. I provide many examples of basic models that improve the realism of economics in very structured, disciplined—competitive—ways using optimization models that closely resemble existing models. And I lay out concrete general criteria for realism-improving theories that make explicit their power and scope—their ability to make precise predictions and their general applicability—and very explicit their relationship to existing models.[3] I also discuss more generally an orientation toward incremental progress in improving models that can be used to improve theoretical and empirical economics. Section 3 turns to the methodological and substantive case for optimization-based models of limited rationality, by briefly illustrating human biases and limits that, by dint of being driven by mistaken intuitions rather than automatic or rule-of-thumb behavior, are at their core optimization models. I conclude in section 4 with a very brief discussion of the movement toward integrating

[3] Much of my discussion in this section reflects arguments I have made elsewhere, especially Rabin (forthcoming).

new models of limited rationality into mainstream empirical and theoretical economics.

2. *Competitive Models*

Since all models are wrong, and all paradigms have their limits, it is always a difficult judgment call as to when and how to emphasize the severity of the shortfalls of prevailing approaches—and when to call for the development of a new paradigm. Nor is it even clear what is inside a paradigm and what is outside. And it certainly does not ultimately matter what labels we use. Especially in the domain of rationality (and behavioral economics), much unintentional harm and distraction and obfuscation has been caused by enthusiasm for categorizing models rather than elucidating specific empirical claims. Harstad and Selten thankfully avoid all that, and are thoughtful and balanced in their approach. But my own judgment is that their assessment of the neoclassical research program is too dire. We all in this forum believe working to reduce its shortcomings is worth the time of economists—more time than it is getting. But part of the case for the approach I advocate perhaps comes from greater confidence that the core of microeconomic theory and empirical work has been a success, and is continuing as a success. And, most importantly in this context, that it is amenable to rapid improvements by incorporating limited-rationality models.

On the other hand, the entire premise of their article is to acknowledge the merits of the neoclassical optimization paradigm and the burden on bounded-rationality models to add value to this approach. I am struck by Harstad and Selten's thoughtful observation that generations of scholars have put effort into bounded rationality without delivering a serious rival to the neoclassical model. The high standard they hold new approaches to is laudable when applied to one's own research approach, as they are doing, but may also bleed into underappreciation of progress in both their own and others' research programs. It is hard to know what counterfactuals to compare a program to. But this article contributes to the forum by enumerating some of the current models of limited rationality that are being introduced to improve economic theory and empirical economics. These models may still miss too much of the bounded in bounded rationality.

Rabin (forthcoming) proposes an approach to improving the psychological realism of economics while maintaining its conventional techniques, of formal theoretical and empirical analysis using tractable models focused on prediction and estimation. The emphasis is on developing models improving behavioral realism that can be used as inputs into economic theory by dint of their precision and broad applicability.[4] Striving for realism-improving theories to be maximally useful to core economic research suggests a particular approach: portable extensions of existing models (PEEMs). One should (a) extend the existing model by formulating a modification that embeds it as a constellation of parameter values with the

[4] The desire for maximal precision of new theories does not come from confidence that the theories are right. Besides enhancing their usefulness, in fact, one value of precision is that it highlights flaws to aid in further improvements. Indeed, in the context of this forum, I would emphasize a macro version of this claim: it may be that the most efficient and useful way to reach what may eventually feel like a new paradigm is in incremental steps. In fact, with caution given the brilliance of the researchers and research involved, it is my guess that the line of research in the tradition of Simon (1955) and Newell and Simon (1972) generates too much emphasis on shortcomings of existing models and—largely as collateral damage from its wonderful emphasis on evidence and intuition about the way people behave in real-world problem-solving—is too oriented toward lists of specific shortcomings. The greater emphasis on generalizable alternative predictions in the approaches inspired by Kahneman and Tversky (2000), and discussed in reviews by Rabin (1998, 2002b) and DellaVigna (2009), may have a greater target chance of both near-term improvements in workaday economic research and long-term improvements in the structure of economics.

new psychological assumptions as alternative parameter values, and (b) make it portable by defining it across domains using the same independent variables as in existing research, or proposing measurable new variables. To further their integration into mainstream economics, PEEMs lend themselves to two types of comparative statics. The first is to look within chosen environments at how predictions change with the parameter values. This allows us to test their empirical validity in comparison to existing theories by estimating those parameters, and to see their potential value added in comparison to existing theories by seeing how those parameters do and don't matter in important economic contexts. The second type of comparative statics is the more traditional one: fixing the parameter values that accord with the improved assumptions, how does changing the environment affect economic outcomes? Applying this form of comparative statics indicates that the improved psychological realism is ready for economic primetime. If one views hypothesis testing and comparative statics as empirical and theoretical competitions, then this approach sets up the competition Harstad and Selten talk about. Rabin (forthcoming) argues that this competition is an important way forward for improvements in economic models, disciplining both those proposing the theories and those doubting them in a way that pushes research toward constructive, normal-science debates.

An early and successful example of this methodology, in the domain of preferences under uncertainty, is the literature providing alternatives to classical expected utility. Beginning with Machina (1982), researchers have introduced realistic aspects of preferences precluded from the expected-utility formulation, ranging from nonlinear probability weighting (Tversky and Kahneman 1992 and Prelec 1998), to disappointment (Bell 1985, Loomes and Sugden 1986, and Gul 1991), to ambiguity aversion (Gilboa and Schmeidler 1989).[5] Although all these models were developed with elegant generality and formulated with axioms capturing basic principles, in the context of monetary risk they have all lent themselves to providing precise alternatives defined in the same set of situations to which classical wealth-based expected utility applies.

The model of reference-dependent preferences developed in Kőszegi and Rabin (2006, 2007, 2009) likewise attempts to bring some of the insights from Kahneman and Tversky's (1979) prospect theory toward general applicability. In a simple linear form of the model, they take any classical economic situation and make an alternative prediction based on the idea that people don't only care about absolute consumption, but about consumption relative to reference levels. Under uncertainty or in deterministic settings, when trading off consumption in different dimensions (quality versus price on a given consumer item, or different items in a bundle of goods), people are more bothered by losses relative to a reference point than pleased by gains. A key to the PEEMishness of the model is that reference points are pinned down: they assume people's reference points are their expected consumption, and make a unique prediction in most settings. This is done by assuming a person behaves according to preferred personal equilibrium: she chooses her favorite consumption plan such that, if she has the reference point corresponding to expecting to follow the plan, she will do so. Kőszegi and Rabin (2006) show that, in static settings with no uncertainty, the prediction of the model corresponds to the classical predictions of consumer theory, but it makes precise alternative predictions in settings with either dynamic choice or where there is intrinsic uncertainty or surprise. The model

[5] See Harless and Camerer (1994) for an excellent conceptual and empirical review of these models.

is parameterized by $\eta \geq 0$, the degree to which one cares about news about gains and losses, $\lambda \geq 1$, the degree to which one is more bothered by unexpected losses in consumption relative to unexpected gains, and $\gamma \in [0, 1]$, the degree to which people care about news about future consumption relative to news about contemporaneous consumption. (Classical reference-independent preferences correspond to $\eta = 0$.)

Perhaps the most active area of formulating general alternatives to more traditional neoclassical models is in departures from narrow self-interest. Following a long tradition in economics, recent models building from Bolton (1991) and Rabin (1993) have tried to study social preferences based on laboratory data. Inspired by simple models by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), Charness and Rabin (2002) specify a simple form where subjects who get more money than others put weight $\rho > 0$ on those others' payoffs and $1 - \rho$ on their own payoff, and when behind the other subject put weight $\sigma > 0$ on those others and $1 - \sigma$ on their own payoff.[6] These simple models ignore intentions-based reciprocity motives; theories such as the full model contained in Charness and Rabin (2002), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) follow Rabin (1993) in formally modeling reciprocity concerns into the utility function.

Closer to the topic of this forum, models have been developed to capture departures from full rationality. Crawford (this forum) discusses various approaches; two models that most closely fit the formal PEEM framework in this domain are McKelvey and Palfrey's (1996) quantal-response equilibrium and Eyster and Rabin's (2005) cursed equilibrium. In McKelvey and Palfrey, a parameter $\lambda \geq 0$ captures the propensity of players to make random mistakes, with an equilibrium assumption pinning down other subjects' beliefs; in Eyster and Rabin's (2005) cursed equilibrium, parameter $\chi \in [0, 1]$ captures the degree to which players neglect the informational content in other players' behavior.[7] Although, as with Nash equilibrium, they do not always predict a unique equilibrium, they are defined in all games solely by the parameters and the structure of the game. As such, they provide exact competitors to Bayesian Nash equilibrium and other classical game-theoretic solution concepts.[8]

Moving to limited-rationality models of individual choice, several formal models fall into the category of quasi-maximization models I discuss below. Loewenstein, O'Donoghue, and Rabin (2003) develop a model where economic actors underappreciate taste change. Relying on the timing of utility and using a simple linear form, the model predicts a particular form of people having irrational beliefs of their own future behavior: they believe their future tastes will be more like their current tastes than they are. The error is parameterized by $0 \leq \alpha \leq 1$, where $\alpha = 1$ means a person fully projects

[6] Although it is doubtful the parameters are very stable outside the lab, in a broad class of games estimates are often around $\rho = 0.4$ and $\sigma > 0$ is very close to zero. Notwithstanding a heavy emphasis in early literature on the idea that subjects dislike getting less money than other subjects, a small but positive value of σ is clearly more common once allocational preferences are disentangled from reciprocity.

[7] See also Esponda (2008) for a variant of cursed equilibrium. Although applied solely in simplified form to a model of naive social learning, Eyster and Rabin (2008) define for all games a form of naive, face-value inference, parameterized by $\upsilon \in [0, 1]$.

[8] The alternative cognitive-hierarchy models discussed by Crawford (this forum), like Stahl and Wilson (1994), Camerer, Ho, and Chong (2004), Crawford and Iriberri (2007), and Crawford, Costas-Gomes, and Iriberri (2013) are not fully closed in the way I emphasize here insofar as the level-0 play is not derived formulaically from the game. But they can be completed with any level 0, and also typically provide unique, and hence more disciplined, predictions once a level-0 is chosen, than do cursed equilibrium and quantal-response equilibrium.

her current tastes onto her future tastes, and $\alpha = 0$ implies full rationality. These models show that adding realistic systematic biases in how we predict our future utility can compete on power and scope with the full-rationality model. Barberis and Huang (2009) develop a formal parameterized model of narrow bracketing, following on the tradition of Benartzi and Thaler (1995), Kahneman and Lovallo (1993), and Read, Loewenstein, and Rabin (1999), whereby agents maximize prospect-theory preferences but fail to consider how gains and losses across different choices might cancel each other out. Barberis and Huang assume that people partially maximize the normatively appropriate combined outcomes from separate choices, but erroneously partially maximize the separate choices, parameterizing the degree of narrow bracketing by the parameter $v \in [0, 1]$. Surely the most successful PEEM to date is the simple model developed by Strotz (1956) and rejuvenated by Laibson (1997), on what O'Donoghue and Rabin (1999) refer to as present bias: parameter β captures short-term impatience, and parameter $\hat{\beta}$ captures misprediction of future impatience. These models are applied widely, improving economic research by any standard that embraces the value of conceptual and empirical insights on economically important topics.

Recent papers also develop formal models of biases in statistical reasoning, although typically in less general settings than the models above. Grether (1980, 1992) models base-rate neglect for the purposes of empirical testing in a simple way, where decision-makers apply Bayes Rule in updating, but, in the logarithmic form of Bayes Rule, weigh the base rate by only $\alpha < 1$. Proper Bayesian updating corresponds to $\alpha = 1$, while $\alpha = 0$ corresponds to maximal base-rate neglect in which priors are neglected altogether. Building on an approach pioneered by Barberis, Shleifer, and Vishny (1998), several papers have developed quasi-Bayesian models of updating, discussed below. Rabin (2002a) captures the idea that people believe in the Law of Small Numbers (a term first formulated by Kahneman and Tversky 1973) whereby people erroneously exaggerate the likelihood that underlying population proportions will show up even in small samples. Rabin and Vayanos (2010) extend this model to capture the related gambler's fallacy, whereby people expect recent positive outcomes to be counteracted in the short term by negative ones; one parameter is used to determine how intensely an agent expects recent signals to be counteracted, and a second parameter captures the memory for how long ago agents expect signals to influence current outcomes. Benjamin, Rabin, and Raymond (2012) model a non-belief in the law of large numbers whereby decisionmakers underappreciate the central limit theorem, and assign positive probability to even huge samples not closely resembling the underlying population.

Although the goal is to provide new models that are fully competitive with earlier economic theory, there is a catch to all these models. Even when the independent variables are imported from regularly used economic theories, the extended models often require interpretations that go beyond the original model. Several of the non-expected-utility models, and the game-theoretic models by McKelvey and Palfrey (1996) and Eyster and Rabin (2005), permit literal extensions (when literally applied) with no additional structure. But almost all the other models do rely on adding further structure than is a literal part the previous neoclassical model they are modifying. The model of present bias, for instance, requires interpretations of the time-delivery of utility flows that need not per se be considered a primitive of models without present bias. Fortunately, the timing of utility is obvious, observable, and well understood in prevailing contexts.

Models of social preferences likewise require cardinal interpretations of material payoffs, such as money, before those models that specify exact parameters are pinned down. Likewise, models like Kőszegi and Rabin (2006) that in principle pin down exact alternative predictions in all situations require a cardinal interpretation of the consumption utility as it appears in economic models. In addition, an assumption of what aspects of consumption constitute distinct hedonic dimensions must be specified in a way not demanded of classical consumer theory. Yet these too are rarely that ambiguous or subject to hidden specification mining in particular applications. Models of probability judgment that depart from Bayesian reasoning, on the other hand, often require even more ancillary assumptions specifying how decisionmakers frame hypotheses. Models of narrow framing, like Barberis and Huang (2009), must make strong assumptions about how people separate their decisions.

In other cases, the parameters are likely to be affected by factors outside the model to the point where they are unstable. Care must be taken, however, not to invoke the instability of specifications to favor worse stable models over better stable models simply because that is what earlier models assumed. For instance, the model that parameterizes inattentiveness to the information contained in others' behavior, as formalized by Eyster and Rabin (2005), is clearly missing variation due to factors missing from the model, such as experience of economic actors and the salience of the information. This means assuming the empirical average $\chi = 0.5$ (say) rather than a less-pinned-down $\chi \in [0, 1]$ may sacrifice too much realism for the sake of precision—but assuming a universal $\chi = 0$ as current game theory does sacrifices even more.

Insofar as these new models leave open parameters, they will of course increase degrees of freedom in making predictions. Although research is necessarily still at the stage of estimating parameters (and, obviously, building new models and modifying previous iterations), a theme of the literature is that the improvements are coming from precise alternatives, not degrees of freedom. The variant of prospect theory in Kőszegi and Rabin (2006, 2007) corresponding to $\eta = 1$, $\lambda = 3$, $\alpha = 0.88$ probably fits large ranges of data better than classical reference-independent model corresponding to $\eta = 0$. Eyster and Rabin (2005) propose that estimating a type of information neglect by $\chi = 0.5$ fits evidence across games better than the fully rational $\chi = 0$. Models of present bias and naivety specifying $\beta = 0.7$, $\hat{\beta} = 0.8$ surely fit most data better than the classical parameters $\beta = \hat{\beta} = 1$. Indeed, in this domain a good guess is that improved assumptions will fit behavior better while *removing* a degree of freedom: instead of trying to fit data with wildly varying values for the traditional discount factor, δ, once we use better values of β and $\hat{\beta}$ we can also restrict our models to the more reasonable yearly $\delta \approx 0.95$. Laboratory data and early field estimates suggest that a good guess about the degree of underappreciation in taste predictions across settings as modeled by Loewenstein, O'Donoghue, and Rabin (2003) is that a misprediction parameter of $\alpha = 0.5$ may generally fit better than the fully rational $\alpha = 0$. Rabin and Vayanos (2010) and subsequent work estimate that the gambler's fallacy of $(\alpha, \delta) = (0.2, 0.7)$ fits better than the Bayesian $\alpha = 0$. Benjamin, Rabin, and Raymond (2012) estimate the degree of nonbelief in the law of large numbers across a wide range of experiments at $\psi \in [7, 15]$ rather than the Bayesian model of $\psi = \infty$. Such numbers will surely not hold up or be stable, and several are simply best guesses I am listing here. But the effort to do so reflects an increasing standard in the literature that some fixed parameterization of alternative models ought to fit better than the rational model across settings, rather than performing well solely from degrees of freedom or selectivity of examples.

This standard, in turn, reflects a growing emphasis in this research program on making pervasive improvements rather than finding particular cases where new models do well. It is right to cut slack for new insights during early stages of development—evidence-gathering and theory-forming under extreme and selective circumstances can be the best way to understand the logic and potential of new models. But if the long-run goal is for newer models that explain things better than the traditional model, care should be taken to investigate whether the new models improve insight on average.

I believe that this is especially important in the context of judging bounded-rationality models. In my view, many new models and explanations for experimental findings look artificially good and artificially insightful in the very limited domain to which they are applied. Although this is a problem for all sorts of bounded-rationality and behavioral-economic theories, and a ubiquitous problem for less formal theories people provide for their experimental results, it may be that procedurally rational models may be especially susceptible to the illusion of explanatory power. Roughly put, the models inherently steer us toward judging them solely in the specific contexts where the procedures make basic rational sense and seem intuitively likely—omitting all the cases where they predict badly. In contrast, although (say) the tendency to infer too much from small samples is centrally mediated by whether we pay attention to the data, in virtually every situation where economists had previously applied the Bayesian-updating model, the law-of-small-numbers inference model may improve our predictions. I am not as sanguine as others that bounded-rationality models of the sort emphasized in much of the literature have the language or aspiration to apply so broadly. We are still likely to benefit from having them as judiciously applied addenda to rational models. My worry is that, short of extremely tight empirical attention, applications of the models as replacements for the neoclassical paradigm may worsen our predictions in general.

To my mind, there is a simple and common reason why some bounded-rationality models are so unlikely to perform well across contexts: too much realistic rationality is being left out. Many models of reinforcement learning, for instance, assume neither a priori rationality nor anything like Nash equilibrium. Yet virtually all applications—either by design in the laboratory or by intuition in thought experiments—presort the environment so that there is little scope for irrationality to lead people far astray. Of the thousands of disastrous things all of us *could* do in virtually every new situation we face in life—including all situations economists care about—basic reason whittles away all but a few. That is, assumptions of rationality and Nash equilibrium do much of the heavy lifting of restricting predictions, and only then do other theories improve our predictions further. It is good, practical science to focus on what the existing neoclassical models are missing. But when it comes to formulating the general models that are serious competitors to full rationality, recognition of the pervasive power of rationality and equilibrium suggests we need the new models to be far closer to the old models than often proposed. Although in the interim it makes sense to formulate models that make improvements in specific contexts, one lesson from this perspective is that researchers ought to imagine how they might eventually combine new theories with a core of strong rationality. I believe the types of models I discuss above and below are more oriented toward that. Indeed, the PEEMs almost by structure have the feature that they replicate the rational model in all the ways that are not the specific focus of the modifications.

3. *Optimal Models of Limited Rationality*

I now turn to the case for modeling certain types of economically relevant errors in terms of optimization models, based on the joint attractions of psychological realism and methodological benefits.

Economists have an intuition and disposition to model bounded rationality in terms of complexity and inattention. The idea is often that people are cognitive misers: we humans wisely choose not to figure out all the exact right solutions to all the problems that face us.[9] This intuition is often on target. But one of the biggest lessons of some branches of psychology is how many errors are simply not in any useful sense conceptualized this way. The degree to which attention to small-scale risks reflects rational maximization of news-utility preferences or reflects errors is a difficult question. But evidence suggests that small-scale risk aversion is not only an error, but a rather effortful one: consumers would be better off simply never buying small-scale insurance and extended warranties. And they would leave the store sooner. Among the many errors small investors make, the inability to see incredibly complicated arbitrage opportunities in real time might be costly. Whether it is true and important that arbitrage opportunities are left on the table by those that don't see them, a far more pervasive error is that people tend to see patterns that are not there. Research on illusory correlation (e.g., Chapman and Chapman 1969 and Hamilton and Gifford 1976), the gambler's fallacy (see, e.g., Bar-Hillel and Wagenaar 1991), and the hot-hand fallacy (see, e.g., Gilovich, Vallone, and Tversky 1985) all concentrate on biased tendencies to see imaginary patterns. The rational way to do the familiar task of predicting the likelihood that the next flip of a coin is tails has a happy simplicity to it.[10] But people have far more complicated beliefs that different answers are needed for predictions following *HTHHHH* versus *HTTHTH* versus *THTHTH*.[11] And traders appear not to ask the very simple question who in the market is trading with me, and why?, or to come up with the near inevitable and perspective-changing answer that it is somebody who is himself trying to make money from different beliefs about an asset. The logic of realizing that in the future you'll have just the same self-control problem as now is no more complicated than the belief that beginning tomorrow you will reform your ways; just less psychologically compelling. The logic that you will crave food in the future when hungry, or (if addicted) crave a drug when you've gone too long without a hit, is cognitively available to all as a familiar fact of life; but it does not grab our attention at moments of decisionmaking when we are currently sated with food or drug.

Nor do all errors come from lack of effort to overcome them, or intellectual or physical laziness. People may not spend much effort predicting coin flips, but many spend huge amounts of time trying to see the patterns in stocks that, for any plausible money-making opportunities, are best approximated by a random walk. Not only may we maintain naiveté about our self-control problems in defiance of any attentive learning rule—we may put a lot of intellectual energy into not learning. The right explanation for misbehavior could eliminate convoluted, complicated,

[9] Economists are, I think, drawn to complexity explanations for limited rationality for several reasons. The main and best reason is its realism: it is intuitive and it is true that many of our limits come from the sheer impossibility of attending to everything or fully maximizing when we do pay attention. But it also seems to have an appeal based on its affinity with computer science and math, rather than psychology, and its greater amenability to deriving assumptions from first principles rather than empirical evidence.

[10] 0.5.

[11] Subjects tend to predict lower and lower chances of tails going from the first to the third sequence.

and time-consuming stories; the 1,000 different reasons why you didn't start your diet or dissertation the previous 1,000 days could be replaced with the simpler right answer—you didn't have the willpower. Or you think you'll not drink or have unsafe sex or smoke when you go out tonight, rather than think the equally simple (and truer) hypothesis that you'll likely do the same things you always do.[12] Such effortful errors are most strikingly inconsistent with the bounded-rationality metaphor when they involve a great deal of planning and actions. If per DellaVigna and Malmendier (2006) we actively join gyms and make other long-term plans for virtuous activities in irrational ways, it would appear that there is a whole lot of optimization going on. Just the wrong kind.

As noted earlier, one major class of optimization-based errors can be conceptualized as very close to utility maximization: situation by situation, people have reasonably focused goals, and maximize those goals reasonably well. We can model a person as engaging in traditional constrained maximization at each moment in time. But we specify the exact mistake the person is making in which function she is maximizing, or in what choice set she is choosing from. The three main classes of such quasi-maximization errors are what might be called (a) narrow bracketing, (b) present bias and hyperbolic discounting, and (c) projection bias in predicting future utility. Narrow bracketing is where people maximize their true utility functions among each choice set they focus on, but don't integrate others. Present bias is where, moment by moment, a person maximizes full intertemporal utility, but at each moment tends to overweight current utility. (And may mispredict the propensity to do so in the future.) Projection bias is where, because of current tastes or current focus, people (actively or passively) mispredict the utility of future situations. Formally, all these errors can be conceptualized as follows. As a function of a person's environment, the person makes choices from a choice set, and maximizes a particular goal: $\mathrm{Max}_{x \in X}\, v(x)$. But people may either focus on the wrong choice set (narrow bracketing), overweight near-term utility (present bias) or mispredict future utility (projection bias).

Present bias and naiveté about present bias as modeled by Laibson (1997) and O'Donoghue and Rabin (1999, 2001) is the most well known of these quasi-maximization errors. The third example, projection bias as modeled by Loewenstein, O'Donoghue, and Rabin (2003), applies in situations where tastes change over time, either due to temporary fluctuations (such as mood swings, or satiation and deprivation of food or drugs), or due to longer term adaptation (such as to our standard of living) or habit formation (such as exercise, or addiction to a drug). Changing tastes are a fact about utility, *not* an indication of irrationality. The error: people systematically underappreciate (even very predictable) changes in their tastes, and hence falsely project their current tastes onto the future. Formally, Loewenstein, O'Donoghue, and Rabin (2003) offer an alternative to the rational-choice assumption

[12] There are two important caveats to this perspective. First, it is slightly misleading to frame things in terms of how simple an answer is. An answer can be simple without it being easy to figure out. But many of these errors seem robust to simple communication of the right answer by a trustworthy person in a way that could happen only when a person is enticed by a countervailing intuition. Second, we need to be somewhat careful that the simplicity of a correct answer in a particular class of situations doesn't require the complex and time-consuming task for people to identify what situations they are in. Maybe the fact that we should not extrapolate our current tastes in cases where our tastes predictably change belies the fact that on average our future tastes will be our current ones; simply assuming no change may be a simple rule of thumb. This too seems wildly miscalibrated. "Ignore current hunger" seems not only a simpler rule of thumb than "extrapolate current hunger," but it also seems the mastering of improved rules like this is almost always employed in comparably important confrontations with binary variation in life that do not implicate the hold that current cravings have on us.

that people maximize state-dependent utility, which can be represented (without discounting) by

$$U^t = \sum_{\tau=t}^{T} u(\mathbf{c}_\tau, \mathbf{s}_\tau),$$

where $u(\mathbf{c}_\tau, \mathbf{s}_\tau)$ is her instantaneous utility in period τ, and $\mathbf{c}_\tau$ is the person's period-τ consumption vector which includes all period-τ behavior relevant for current or future instantaneous utilities. The vector $\mathbf{s}_\tau$ is the person's state in period τ, which incorporates all factors that affect instantaneous utility besides current consumption. They assume that at time t a person maximizes

$$\tilde{U}^t = \sum_{\tau=t}^{T} \tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau \,|\, \mathbf{s}_t),$$

where $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau \,|\, \mathbf{s}_t)$ is the prediction at time t in state $\mathbf{s}_t$ of $u(\mathbf{c}_\tau, \mathbf{s}_\tau)$. Whereas rational expectations posits that $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau \,|\, \mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_\tau)$, Loewenstein, O'Donoghue, and Rabin (2003) assume that a person may suffer from *simple projection bias* if there exists $\alpha \in [0, 1]$ such that for all $\mathbf{c}$, $\mathbf{s}$, and $\mathbf{s}'$,

$$\tilde{u}(\mathbf{c}, \mathbf{s} \,|\, \mathbf{s}') = (1-\alpha)\, u(\mathbf{c}, \mathbf{s}) + \alpha\, u(\mathbf{c}, \mathbf{s}').$$

This posits that at each moment in time people have well-defined perceived utility from their courses of action, and take actions to maximize their long-run happiness. This is right or wrong in all the ways the rational model can be right or wrong, differing solely from this general principle of projecting current tastes into the future. Projection bias can help understand mispredictions of food (Read and van Leeuwen 1998) and drug (Badger et al. 2007) cravings, underappreciation of long-run habit formation (Levy 2010 and Acland and Levy 2011), catalog clothing purchases (Conlin, O'Donoghue, and Vogelsang 2007), and demand for swimming pools and convertible car roofs based on temporary weather fluctuations (Busse et al. 2012).

There are also now a class of models that capture errors in people's statistical reasoning that fully accord to neoclassical models of preference optimization, but simply build in a systematic misunderstanding of statistics. In these quasi-Bayesian models, people maximize expected utility (or any other appropriate utility function), but form their expectations based on wrong models of the world. Given prior probabilistic beliefs $\pi(h)$ about hypotheses $h \in H$, how does a person form updated beliefs, $P(h^* | e)$, about the likelihood that a particular hypothesis, $h^* \in H$, is true upon observing evidence e? Proper Bayesian updating implies that after observing information $e_1, e_2, \ldots$ a person forms posterior beliefs

$$B(h^* | e_1, e_2, \ldots) = \frac{\pi(e_1, e_2, \ldots | h^*)\pi(h^*)}{\sum_{h \in H} \pi(e_1, e_2, \ldots | h)\pi(h)}.$$

Some errors in statistical reasoning can be understood in terms of an alternative functional form of how people combine conditional probabilities and priors, by assuming that people use some other functional form

$$P(h^* | e_1, e_2, \ldots) =$$

$$f(\{\pi(e_1, e_2, \ldots | h)\}_{h \in H}, \{\pi(h)\}_{h \in H}).$$

Most famously, people suffer from base-rate neglect: they underuse the priors $\{\pi(h)\}_{h \in H}$ implying updated beliefs have functional form

$$P(h^* | e_1, e_2, \ldots) = \frac{\pi(e_1, e_2, \ldots | h^*)[\pi(h^*)]^\alpha}{\sum_{h \in H} \pi(e_1, e_2, \ldots | h)[\pi(h)]^\alpha},$$

with $\alpha \leq 1$.

More interestingly, many errors can be modeled as people engaging in putatively proper Bayesian updating, but with a precise way in which they either misobserve or misunderstand how that evidence relates to the hypotheses. Then researchers understand the implications of the error by studying a form of Bayesian updating that embeds the error. There are two categories of such quasi-Bayesian updating that researchers have employed. Warped-model Bayesians update according to

$$P(h^*|e_1, e_2,\ldots) = \frac{\tilde{\pi}(e_1, e_2,\ldots|h^*)\pi(h^*)}{\sum_{h\in H}\tilde{\pi}(e_1, e_2,\ldots|h)\pi(h)},$$

where $\tilde{\pi}(e_1, e_2, \ldots|h)$ is a false (but internally consistent) model of how signals are generated. This approach was pioneered by Barberis, Shleifer, and Vishny (1998), and adopted by Rabin (2002a) and Rabin and Vayanos (2010) to capture false overconfidence that small samples resemble underlying proportions, and Benjamin, Rabin, and Raymond (2012) to capture the false belief that large samples might not resemble underlying proportions. Other errors are captured by assuming that people are information-misreading Baye-sians, updating according to

$$P(h^*|e_1, e_2, \ldots) =$$

$$\frac{\pi(f(e_1), f(e_1, e_2), \ldots..), ..|h^*)\pi(h^*)}{\sum_{h\in H}\pi(f(e_1), f(e_1, e_2), \ldots..), ..|h)\pi(h)},$$

where $f(e_1), f(e_1, e_2)$ are misinterpreted signals. Rabin and Schrag (1999) capture psychological evidence on confirmatory bias by assuming that people tend to misread signals as supporting earlier hypotheses. Once a teacher conjectures a student is smart rather than stupid, he may interpret complicated answers in class during the semester as subtle rather than confused. Mullainathan (2002) models naive memory problems by assuming people forget some signals—but update as if the absence of signals contains information. A teacher may not remember all the good answers a student provides in class, but may assess the student harshly because he thinks a smart student should have given more good answers than he remembers her giving.

4. *Conclusion*

One welcome feature of the article by Harstad and Selten and the reply by Crawford is the attempt to relate models of bounded rationality to important economic phenomena, such as asset-market bubbles. Although I am not as optimistic as they are that the laboratory evidence they bring to bear provides the right building blocks for new models of behavior that drive real-world bubbles, the focus on the economic implications of new theories of choice is an important step. This of course comes at a cost if it is too soon to try to apply our models, or if economists studying real-world empirical applications are not the target consumers of our new models. Yet thinking about the implications of new models in work-aday applied-theory and empirical economics is a tremendously important component of the program to improve psychological realism in economics. First, it should not be controversial that mainstream economic researchers *are* the (eventual, downstream) target consumers. Second, it will provide guidance on which of the infinite ways we can improve the realism of economic assumptions are likely to be most important. Third, it will provide some guidance on the methodology researchers should employ to improve realism and the types of models that will be most useful.

It is early days for the applications of many optimization models of limited rationality, but I do think the more intimate ties between empirical economics and some of these models bodes well, in comparison to bounded-rationality models, for both their applicability and the speed at which the models themselves can be improved.[13] It

[13] But see Spiegler (2011) for an excellent exploration of the potential of both genres of models to provide insights into economic institutions and outcomes.

would be bad judgment to bet confidently against the potential insights of a research line that has generated so much excitement among economic theorists. And it would also be bad empirical analysis—since models of bounded rationality have already inspired new insights. But I think as we move forward with economic implications, with theoretical comparative statics and empirical measurement, some of the abstract and intuitive arguments we have all provided in this forum will be given more productive structure. Many of the new optimization models are being integrated rapidly into economic analysis to the benefit of both economics and the new models. For models of present bias, Strotz (1956) and Laibson (1997) started with the important application to pared-down models of savings and consumption. But the literature has since expanded rapidly to applications in too many domains of economics to list here. Angeletos et al. (2001) show how it can help understand housing, credit-card, and other markets in the United States. And there is research now studying its implications for default effects and commitment devices in savings, as well as for smoking behavior, optimal taxation, etc. Models of the role of narrow bracketing in loss aversion, from risky choice (e.g., Benartzi and Thaler 1995) to labor supply (e.g., Camerer et al. 1997 and Crawford and Meng 2011), are also surely going to improve empirical work. Analysis of the role of mispredicting future preferences, now studied in the context of clothing, car, and housing choices (see Conlin, O'Donoghue, and Vogelsang 2007 and Busse et al. 2012) are in earlier days. Little has yet been done to integrate statistical errors, or models of how people are neglectful and irrational in extracting information from other economic actors in strategic and market contexts. But integrating the potential of portable models of erroneous beliefs—e.g., to form more realistic notions of the beliefs of small investors in asset markets—is clear.

References

Acland, Dan, and Matthew Levy. 2011. "Habit Formation, Naiveté, and Projection Bias in Gym Attendance." Unpublished.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg. 2001. "The Hyberbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives* 15 (3): 47–68.

Badger, Gary J., Warren K. Bickel, Louis A. Giordano, Eric A. Jacobs, George Loewenstein, and Lisa Marsch. 2007. "Altered States: The Impact of Immediate Craving on the Valuation of Current and Future Opioids." *Journal of Health Economics* 26 (5): 865–76.

Barberis, Nicholas, and Ming Huang. 2009. "Preferences with Frames: A New Utility Specification That Allows for the Framing of Risks." *Journal of Economic Dynamics and Control* 33 (8): 1555–76.

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

Bar-Hillel, Maya, and Willem A. Wagenaar. 1991. "The Perception of Randomness." *Advances in Applied Mathematics* 12 (4): 428–54.

Bell, David E. 1985. "Disappointment in Decision Making under Uncertainty." *Operations Research* 33 (1): 1–27.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics* 110 (1): 73–92.

Benjamin, Daniel J., Matthew Rabin, and Collin Raymond. 2012. "A Model of Non-belief in the Law of Large Numbers." Unpublished.

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

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–93.

Busse, Meghan R., Devin G. Pope, Jaren C. Pope, and Jorge Silva-Risso. 2012. "Projection Bias in the Car and Housing Markets." Unpublished.

Camerer, Colin F., Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics* 112 (2): 407–41.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics* 119 (3): 861–98.

Chapman, Loren J., and Jean P. Chapman. 1969. "Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs." *Journal of Abnormal Psychology* 74 (3): 271–80.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–69.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97 (4): 1217–49.

Crawford, Vincent P. 2013. “Boundedly Rational versus Optimization-Based Models of Strategic Thinking and Learning in Games.” *Journal of Economic Literature* 51 (2).

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri. 2013. “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications.” *Journal of Economic Literature* 51 (1): 5–62.

Crawford, Vincent P., and Nagore Iriberri. 2007. “Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?” *Econometrica* 75 (6): 1721–70.

Crawford, Vincent P., and Juanjuan Meng. 2011. “New York City Cab Drivers’ Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income.” *American Economic Review* 101 (5): 1912–32.

DellaVigna, Stefano. 2009. “Psychology and Economics: Evidence from the Field.” *Journal of Economic Literature* 47 (2): 315–72.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. “Paying Not to Go to the Gym.” *American Economic Review* 96 (3): 694–719.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. “A Theory of Sequential Reciprocity.” *Games and Economic Behavior* 47 (2): 268–98.

Esponda, Ignacio. 2008. “Behavioral Equilibrium in Economies with Adverse Selection.” *American Economic Review* 98 (4): 1269–91.

Eyster, Erik, and Matthew Rabin. 2005. “Cursed Equilibrium.” *Econometrica* 73 (5): 1623–72.

Eyster, Erik, and Matthew Rabin. 2008. “Naïve Herding.” Unpublished.

Falk, Armin, and Urs Fischbacher. 2006. “A Theory of Reciprocity.” *Games and Economic Behavior* 54 (2): 293–315.

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics* 114 (3): 817–68.

Gilboa, Itzhak, and David Schmeidler. 1989. “Maxmin Expected Utility with Non-unique Prior.” *Journal of Mathematical Economics* 18 (2): 141–53.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. “The Hot Hand in Basketball: On the Misperception of Random Sequences.” *Cognitive Psychology* 17 (3): 295–314.

Grether, David M. 1980. “Bayes Rule as a Descriptive Model: The Representativeness Heuristic.” *Quarterly Journal of Economics* 95 (3): 537–57.

Grether, David M. 1992. “Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence.” *Journal of Economic Behavior and Organization* 17 (1): 31–57.

Gul, Faruk. 1991. “A Theory of Disappointment Aversion.” *Econometrica* 59 (3): 667–86.

Hamilton, David L., and Robert K. Gifford. 1976. “Illusory Correlation in Interpersonal Perception: A Cognitive Basis of Stereotype Judgments.” *Journal of Experimental Social Psychology* 12 (4): 392–407.

Harless, David W., and Colin F. Camerer. 1994. “The Predictive Utility of Generalized Expected Utility Theories.” *Econometrica* 62 (6): 1251–89.

Harstad, Ronald M., and Reinhard Selten. 2013. “Bounded-Rationality Models: Tasks to Become Intellectually Competitive.” *Journal of Economic Literature* 51 (2).

Kahneman, Daniel, and Dan Lovallo. 1993. “Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking.” *Management Science* 39 (1): 17–31.

Kahneman, Daniel, and Amos Tversky. 1973. “On the Psychology of Prediction.” *Psychological Review* 80 (4): 237–51.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (2): 263–91.

Kahneman, Daniel, and Amos Tversky, eds. 2000. *Choices, Values, and Frames*. Cambridge; New York and Melbourne: Cambridge University Press; New York: Russell Sage Foundation.

Kőszegi, Botond. 2010. “Utility from Anticipation and Personal Equilibrium.” *Economic Theory* 44 (3): 415–44.

Kőszegi, Botond, and Matthew Rabin. 2006. “A Model of Reference-Dependent Preferences.” *Quarterly Journal of Economics* 121 (4): 1133–65.

Kőszegi, Botond, and Matthew Rabin. 2007. “Reference-Dependent Risk Attitudes.” *American Economic Review* 97 (4): 1047–73.

Kőszegi, Botond, and Matthew Rabin. 2009. “Reference-Dependent Consumption Plans.” *American Economic Review* 99 (3): 909–36.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics* 112 (2): 443–77.

Levy, Matthew R. 2010. “An Empirical Analysis of Biases in Cigarette Addiction.” Unpublished.

Loewenstein, George, Ted O’Donoghue, and Matthew Rabin. 2003. “Projection Bias in Predicting Future Utility.” *Quarterly Journal of Economics* 118 (4): 1209–48.

Loomes, Graham, and Robert Sugden. 1986. “Disappointment and Dynamic Consistency in Choice under Uncertainty.” *Review of Economic Studies* 53 (2): 271–82.

Machina, Mark J. 1982. “‘Expected Utility’ Analysis without the Independence Axiom.” *Econometrica* 50 (2): 277–323.

McKelvey, Richard D., and Thomas R. Palfrey. 1996. “A Statistical Theory of Equilibrium in Games.” *Japanese Economic Review* 47 (2): 186–209.

Mullainathan, Sendhil. 2002. “A Memory-Based Model of Bounded Rationality.” *Quarterly Journal of Economics* 117 (3): 735–74.

Newell, Allen, and Herbert A. Simon. 1972. *Human Problem Solving*. New York: Prentice Hall.

O’Donoghue, Ted, and Matthew Rabin. 1999. “Doing It Now or Later.” *American Economic Review* 89 (1): 103–24.

O’Donoghue, Ted, and Matthew Rabin. 2001. “Choice and Procrastination.” *Quarterly Journal of Economics*

116 (1): 121–60.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica* 66 (3): 497–527.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281–1302.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.

Rabin, Matthew. 2002a. "Inference by Believers in the Law of Small Numbers." *Quarterly Journal of Economics* 117 (3): 775–816.

Rabin, Matthew. 2002b. "A Perspective on Psychology and Economics." *European Economic Review* 46 (4–5): 657–85.

Rabin, Matthew. 2013. "An Approach to Incorporating Psychology into Economics." *American Economic Review* 103 (3): 617–22.

Rabin, Matthew, and Joel L. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics* 114 (1): 37–82.

Rabin, Matthew, and Dimitri Vayanos. 2010. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies* 77 (2): 730–78.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty* 19 (1–3): 171–97.

Read, Daniel, and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118.

Spiegler, Ran. 2011. *Bounded Rationality and Industrial Organization*. Oxford and New York: Oxford University Press.

Stahl, Dale O., II, and Paul W. Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." *Journal of Economic Behavior and Organization* 25 (3): 309–27.

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

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323.

This article has been cited by:

1. Ronald M. Harstad,, Reinhard Selten. 2013. Bounded-Rationality Models: Tasks to Become Intellectually Competitive. *Journal of Economic Literature* **51**:2, 496-511. [Abstract] [View PDF article] [PDF with links]
2. Vincent P. Crawford. 2013. Boundedly Rational versus Optimization-Based Models of Strategic Thinking and Learning in Games. *Journal of Economic Literature* **51**:2, 512-527. [Abstract] [View PDF article] [PDF with links]