

## Discussion of “Behavioral Economics”

*“Behavioral Economics” (Colin Camerer)  
and “Incentives and Self-Control”  
(Ted O’Donoghue and Matthew Rabin)*

**Ariel Rubinstein\***

### 1 WHAT IS THE BEHAVIORAL ECONOMICS “REVOLUTION”?

For me, economics is a collection of ideas and conventions that economists accept and use to reason with. Namely, it is a culture. Behavioral economics represents a transformation of that culture. Nonetheless, as pointed out by Camerer and Loewenstein (2003), its methods are pretty much the same as those introduced by the game theory revolution. At the core of most models in behavioral economics there are still agents who maximize a preference relation over some space of consequences and the solution in most cases still involves standard equilibrium concepts. However, the behavioral economists are not committed to what is usually referred to as rational motivations. An economic fable (or a model, as we would call it) that has at its core fairness, envy, present-bias and the like, is by now not only permitted but even preferred.

Why now? Perhaps, economists have finally realized that orthodox economic models are too unrealistic and dogmatic. And perhaps it is the result of our constant search for new directions in research. One might also ask why other ideas (such as those of bounded rationality) are less welcome than those of behavioral economics. I think that this is because the profession prefers progress in small steps. The models of behavioral economics are not that different from those of applied economics and thus are not perceived as a threat.

The extent of this transformation may go beyond the topics discussed in this session. For example, behavioral economics may influence the way economics is applied to political issues. It also reintroduces ideological questions such as to what extent governments should paternalistically “repair” biases and fallacies.

\* I would like to thank Rani Spiegler for the many discussions we had on behavioral economics during the last few years. I could not imagine writing this essay without the benefit of his comments, insights, ideas and criticism.

## 2 THEORETICAL BEHAVIORAL ECONOMICS

A paper in behavioral economics typically begins with a description of a real life phenomenon that cannot easily be explained by the standard rational man paradigm. To support the case, references are brought from research in psychology and sometimes even studies of animal behavior. In O'Donoghue and Rabin's discussion all that is borrowed from psychology is the idea that *some* people, in *some* cases, have *some* present-bias. I doubt we need psychological justification in order to make that assumption. Each of us can think of situations in which we exhibit present-bias. However, it is also easy to think of situations where we have *future*-bias. For example, say I have one piece of expensive chocolate. Whenever I am about to eat it, I think to myself: Why not leave it for the future and enjoy the feeling of expectation? The outcome: I leave it so long that it is no longer edible. The psychology literature tries to understand the circumstances where a phenomenon exists. But, in order to just assume that present-bias often exists, it is sufficient to cite casual observation as is done in other fields of economics.

The typical paper in this field then moves on to modeling the bias. The basic framework used to model present-bias is not new and goes back to Strotz (1956). What is new is that time inconsistency is being applied to a variety of economic settings. Ten years ago it was difficult to publish a paper in the QJE which included a "present-bias" assumption. These days it is almost impossible to publish a paper in that same journal that ignores present-bias, let alone one which criticizes the approach.

O'Donoghue and Rabin adopted the  $\beta, \delta = 1$  model. They repeatedly make the point that the standard time consistent model is wrong and that the  $\beta, \delta$  model is correct. I agree that the  $\beta, \delta$  model is a very interesting example of an analytically convenient functional form, but I find the claim that these models are more accurate and realistic to be misleading. Note that the introduction of time inconsistency requires the addition of strong assumptions about the way that different selves interact. In order to complete the model, behavioral economists resort to the standard assumptions. Usually these involve either naive or sophisticated agents. Naivete is not realistic since agents never learn. Sophistication is unrealistic since it suffers from the problems of subgame perfection. An agent is super-rational in the sense that he perfectly anticipates his future selves and arrives at equilibrium between them. Present-bias is a realistic phenomenon, but the combination of the  $\beta, \delta$  preferences with naivete or sophistication assumptions makes the model even more unrealistic than time consistency models.

O'Donoghue and Rabin go out of their way to beat, if I may use their own phrase, the "dead parrot" of full rationality. Of course there are many facts that are hard to reconcile with full rationality. But the psychology and economics literature has replaced a dead parrot with one that is equally dead. If the "time consistent" model is wrong, then the  $\beta, \delta$  model is equally wrong. It can easily be disproved experimentally (see Rubinstein (2003, 2004a)), but the  $\beta, \delta$  fans apparently prefer to ignore experimental evidence that does not go their way.

The typical paper in this field then moves on to demonstrate that a particular standard economic result is not robust to the introduction of psychological bias. I cannot be more sympathetic to this research agenda and, in fact, I myself have been involved in research of a similar nature. In particular, I found O'Donoghue and Rabin's question regarding the design of incentive schemes to overcome time biases to be so interesting that it would remain of interest even if the authors add some modest reservations to their claims.

A major drawback of the behavioral economics models is that they lack both the elegance and generality that characterize the literature of general equilibrium and game theory. The typical paper is messy and terribly long. Simple ideas are lost in poorly formulated models and numerical examples.

If I were the authors, I would have presented the issue using the following simple model:

You bring your child to a street fair to buy him a toy. It is available in two versions –  $G$  and  $B$ . You are unable to assess the quality of the toy and rely on your child's judgement. There are  $T$  stands in a row and you plan to walk with your child from one edge of the row to the other. Each stand sells one of the versions and you assign probability 0.5 to the possibility that a particular stand sells  $G$ .

Both you and your child prefer to buy the  $G$  version rather than the  $B$  version. However, your child is a  $\beta, \delta = 1$  type. He is present-biased and will settle for the  $B$  version even if he expects to find the  $G$  version with very high probability at another stand. Without your intervention, the trip will end at the first stand and the child will buy whatever version the first stand is selling. You are not happy about this and seek a way to trick your child into making the right choice.

You have \$1 in your pocket. Your child prefers  $G + \$1$  in the future over  $B$  today. On the other hand, he prefers  $G$  right away to  $G + \$1$  in the future. A strategy determines, for every  $t$ , whether or not to pay the child a dollar if he buys the toy at the  $t$ 'th stand. Your aim is to maximize the probability that your child buys  $G$ .

If the child is naive and  $T$  is large, an optimal strategy would be to promise to pay the dollar if the child buys the toy during the last  $K$  periods where  $K$  is relatively small. The reason is that when facing  $B$  in the first phase, the naive agent expects to obtain  $G + \$1$  in the second phase with very high probability. Thus, he will pass over  $B$  in the first phase in anticipation of receiving  $G + \$1$  in the second phase. The result is that he will almost always buy  $G$  in the first  $T - K$  periods without the parent having to actually pay the \$1.

As for the sophisticated child, it is not difficult to see that you can at most increase the probability that he buys  $G$  to  $3/4$  by promising the \$1 if the child buys the toy at the second stand.

This simple example has every advantage over Rabin's presentation: it is simple, non-trivial and avoids the long calculations.

**Comment:** One criticism made of behavioral economics is the arbitrariness of the welfare criterion. If an agent is a collection of selves, then why should the utility of the first self be the basis for welfare considerations? If an agent's utility is also affected by disappointment or envy, why should a utility function devoid

of psychological effects serve as the welfare measure? Gul and Pesendorfer (2005) refer to several functional forms used in behavioral economics that have a psychological element. Gul and Pesendorfer point out that the behavior in a typical behavioral economics paper is consistent with many utility functions that have the assumed functional form and therefore welfare considerations cannot be derived from an agent's behavior. Since they only accept economic entities that are rooted in behavior, they are led to fundamentally reject the behavioral economics methodology.

I disagree with them. Behavioral economics makes it clear that one must make assumptions about the relation between the preferences that explain behavior and the preferences used in the welfare criterion. Even when an agent is perfectly rational in the sense that he systematically maximizes some function, it is not at all obvious that the utility function that explains his behavior should be inserted into the welfare considerations. An extreme case would be an agent who is acting against what he perceives to be his own best interests. That is, he maximizes the function  $-v$ , where  $v$  represents his own perceived interests. It would be absurd to use the function  $-v$  as a positive component in the welfare analysis. More generally, once psychological effects enter into the calculus, there is no escape from separating welfare and behavior. The behavioral economics literature uses specific functional forms to state a particular relation between behavior and welfare.

### 3 "RELYING ON EVIDENCE"

The "behavioral economics" literature relies heavily on evidence from experiments, animal behavior and neuroeconomics. I am not happy with the way it is done and would like to comment on each in turn.

#### (i) Experiments

An experiment in economics starts from a basic intuition. Anybody can do an experiment but doing a good one that exposes a psychological phenomenon in a crystal clear manner is an art. If there is value in doing experiments, rather than making do with causal empiricism, it is only when they are done and assessed very carefully. It is my impression that intuitive and "sexy" results are gladly accepted by behavioral economists without sufficient criticism.

Let me illustrate the point with the well-known paper by Gneezy and Rustichini (2000). Camerer describes the paper in the following words: "To discourage parents from picking their children up late, a day-care center instituted a fine for each minute that parents arrived late at the center. The fine had the perverse effect of *increasing* parental lateness. The authors postulated that the fine eliminated the moral disapprobation associated with arriving late and replaced it with a simple monetary cost that some parents decided was worth incurring. Their results show that the effect of price changes can be quite different than in economic theory when behavior has moral components that wages and prices alter." Gneezy and Rustichini (2000) is indeed based on a very appealing

intuition. The behavioral economics literature has wholeheartedly adopted the paper. Camerer, Loewenstein and Rabin (2003) included this paper in their impressive collection of selected papers and Levitt and Dubner (2005) discussed it in Chapter 1 of their new best seller *Freakonomics*.

Being a skeptic, I found it difficult to believe that the experiment could have been carried out as described. I know Israel quite well. It is a country where rules are rarely enforced. It is hard for me to believe that teachers would really fine a parent who is ten minutes late. In my experience, any excuse for lateness is accepted. Furthermore, it is impossible for me to imagine that Israeli teachers would have kept even roughly accurate records of late arrivals with noisy parents crowding around the entrance of the school to take home their screaming kids.

Therefore, I at least want to know what the procedure was for collecting data. The paper does not provide such details. In correspondence, one of the authors claimed that professional standards had been maintained. Apparently, an RA went to the schools once a week and asked the assistant teacher who was late the previous week. There was no attempt to control the accuracy of the RA's records. Oddly, I was not allowed to talk with the teachers.

The authors claim that their findings are "statistically significant." But, as always, the statistical calculations do not take into account the reliability of the collected data. What is the probability that teachers actually implemented the incentive scheme and that assistant teachers accurately reported late arrivals?

Again, I do not doubt the truth and originality of the conclusion that introducing a small fine might be counterproductive. In fact, Heyman and Arieli (2004) even demonstrated the point experimentally. But I am in doubt as to whether the conclusions were indeed confirmed experimentally in Haifa kindergartens.

Who is to blame? The overly motivated authors; the refereeing process, which puts too much trust in authors' data; myself, since I knew about the faulty procedure and did not bother to write a comment (which would probably have been rejected); and what is most relevant to the current discussion – those behavioral economists who gave the paper wide exposure without critical assessment.

**Comment:** The reader is encouraged to read Gneezy and Rustichini's response (see <http://arielrubinstein.tau.ac.il/papers/WC05/GR.pdf>). The fact that the authors feel that all standard professional rules were kept strengthens my argument (which is not directed at any particular experimenter) regarding the uncritical way in which experimental results are accepted.

I should clarify that, at the time, one of the authors agreed that he should delay publication until a new experiment with better monitoring of data collection was conducted. Eventually, the paper was published as is.

I was pleased to learn that the authors decided to redo the experiment more carefully and that a draft of the paper reporting the new results is available (see <http://arielrubinstein.tau.ac.il/papers/WC05/GR1.pdf>). The burden of proof is on the authors, and I cannot dismiss my doubts about the method of data collection in both experiments. Although the authors report that they are still working

on the data and that they have asked the teachers for permission to be interviewed by me, the author who directed the experiment has just informed me that he has lost the names of the kindergarten teachers who participated in the experiment.

## (ii) Animal Behavior

The behavioral economics literature often quotes results from animal behavior to support an assumption. Laibson (1986) quotes results on pigeon behavior (see Ainslie (1992)) as support for the  $\beta, \delta$  model. Glimcher, Dorris and Bayer (2005) cite Harper's (1982) ducks who reach Nash equilibrium in an allocation game in a pond at Cambridge. Camerer quotes Dorris and Glimcher (2004) who claim that: "Monkeys can also learn to approximate mixed strategies in games."

Why is animal behavior relevant? I have no idea. If the behavioral economists are trying to say that the behavior of human beings is rooted in their physical nature, I imagine they are right. Indeed, we are just flesh, blood, and neurons. Even if we consider these experimental results relevant, a skeptical approach is recommended here as well.

Recently, I attended Colin's lecture on Neuroeconomics. I was impressed by a video of Chen's monkey experiment ( $n = 3$ ), which Colin presented (see Chen, Lakshminarayanan and Santos (2005)). In each set of this experiment, a monkey was given 12 tokens. Each of two experimenters held either one piece of apple or one piece of candy. The monkey would then choose to which experimenter it gives a token for what he was holding in return. This process was repeated 12 times until the monkey had spent all its tokens. It is truly amazing to watch a monkey pay for food. It would only be more amazing to watch a monkey following a Wall Street ticker tape and trading options.

The experimenters report that they waited until the monkey's behavior stabilized (no data on this phase was made available) and then the average of the monkey's performance over a week was calculated. The monkeys clearly satisfied Walras' Law. No wonder, they were hungry. Then, the experimenters doubled the number of apple pieces the monkey could get for one token. The number of tokens was adjusted so that his previous average bundle was on his new budget curve. The three monkeys, on average, increased their apple consumption. Two of them spent more tokens on apples. Chen et al. make the following claim: "We show that standard price theory does a remarkably good job of describing capuchin purchasing behavior; capuchin monkeys react rationally to both price and wealth shocks." Camerer praises the result. National Public Radio concludes that "Humans Ape Monkey Market Decisions."

In correspondence with one of the authors, I found out a few things that were not reported in the paper.

First, the monkey was eating what he "bought" immediately. Thus, he was not selecting a bundle. He was actually involved in a series of binary choices interpreted only by the authors as a choice of bundles. The fact that a series of such choices is very different from a one-time choice is in fact something Simonson

(1990) demonstrated in a beautiful experiment which Camerer, Loewenstein and Rabin (2003) themselves refer to. When having to choose between the same two snacks day after day, children usually chose the same snack. When they had to plan in advance, they diversified much more. Thus, there is no reason to link successive choices to the choices made in the standard consumer model.

Second, the authors state that they are reporting on average consumption following the stabilization of choice (the allocation of the 12 tokens). On my first reading, I presumed that the stability of choice was maintained once achieved. The paper only presents average consumption without the raw data itself. When I obtained the data, I in fact did not find any stability in the choices following the "stabilization."

Thus, the most one can say is that when faced with a sequence of choices between a piece of apple and a piece of candy, the three monkeys chose the slice of apple with frequencies of 47%, 42% and 51% respectively; and when faced with a sequence of choices between two pieces of apple and one piece of candy they chose the slices of apple with frequencies 69%, 64% and 50% respectively. Not remarkable, nothing to do with prices and budget sets, nothing to quote. Just a story about three poor monkeys, three overly motivated researchers and some non-critical behavioral economists.

### (iii) Neuroeconomics

I am sure that Neuroeconomics will be the subject of one of the sessions at the next World Congress. However, here again we have a field that oversells itself.

Consider, as an example, the fMRI study by Sanfey et al. of the ultimatum game with \$10. Camerer, Loewenstein and Perlec (2005) refer to it as "one of the most striking neuroscientific findings about game theory." It is claimed that the activity in an area in the brain that is associated with negative feelings (in a statistical sense) is greater when an individual considers an offer of \$2 than when he considers an offer of \$5. It is also greater among individuals who accept a \$2 offer than among those who reject it. Sanfey et al.'s interpretation of the observation is that in response to an unfair offer, the brain struggles to resolve the conflict between the desire to accept a monetary reward and the "insult" of being treated unfairly. I bet they are right! What other considerations can cross the mind of a player who is offered \$2 out of \$10? If confirmed by fMRI studies, it would be a cause for celebration among fMRI researchers. But the results themselves are not that clear.

Problems: If there are two centers in the brain that express conflicting emotional and cognitive motives, which brain center resolves the conflict? Does the increase in brain activities come before the subject's discussion or does it reflect his feeling after he made up his mind? If the finding is correct, would one not expect the time response of those who accept an insultingly low offer to be shorter than that of those who reject it? In fact, I found (among 3700 subjects) the distributions of response time of those who accept the offer and those who reject it to be amazingly similar (see Rubinstein (2004b)). In any case, the

findings far from justify the title "The Neural Basis of Economic Decision-Making in the Ultimatum Game."

Of course, brain studies are fascinating, probably more so than even, say, economics. I would not be surprised if brain studies eventually change our view of decision making. However, I have yet to come across a single relevant insight produced by these studies. The popularity of brain studies might have to do with the obsession among many economists of becoming scientists.

How can serious researchers draw such hasty conclusions from so little data? Well, without the support of fMRI studies, I dare say economists are human beings with ordinary emotions and aspirations. Once I asked a researcher in the field whether he feels comfortable about making conclusions from such a small sample and relying on statistical inferences which are not well understood. He said, "Yes I know, but if I don't do it someone else will and I will not be the first."

#### 4 FINAL NOTE

To conclude, some of us (see Rubinstein (2001), Harrison (2005) and Shaked (2005)) have already voiced our dismay with the rhetorics of behavioral economics. It is nice that economic research gets on the cover of *Newsweek*, is quoted by the *Economist* and is given airtime by NPR. There is no reason for economics to hide behind the traditional barriers. We should widen the range of economic discourse but not by lowering our standards. For behavioral economics to be a revolutionary program of research rather than a passing episode, it must become more open-minded and much more critical of itself.

#### References

- Ainslie, G. (1992). *Picoeconomics*, Cambridge: Cambridge University Press.
- Camerer, C., G. Loewenstein (2003). "Behavioral Economics: Past, Present, Future", in Camerer, C., G. Loewenstein, and M. Rabin, *Advances in Behavioral Economics*, Princeton University Press.
- Camerer, C., G. Loewenstein, and D. Prelec (2005). "Neuroeconomics: How Neuroscience Can Inform Economics", *Journal of Economic Literature*, 43, 9–64.
- Camerer, C., G. Loewenstein, and M. Rabin (2003). *Advances in Behavioral Economics*, Princeton University Press.
- Chen, M.K., V. Lakshminarayanan and L. Santos (2005). "The Evolution of our Preferences: Evidence from Capuchin-Monkey Trading Behavior", mimeo.
- Dorris, M.C. and P.W. Glimcher (2004). "Activity in Posterior Parietal Cortex is Correlated with the Subjective Desirability of an Action", *Neuron*, 44, 365–378.
- Glimcher, P.W., M.C. Dorris and H.M. Bayer (2005). "Physiological Utility Theory and the Neuroeconomics of Choice", *Games and Economic Behavior*, 52, 213–256.
- Gneezy, U. and A. Rustichini (2000). "A Fine Is a Price", *Journal of Legal Studies*, 29, 1–18.
- Gul, F. and W. Pesendorfer (2005). "The Idea of Neuroeconomics", a lecture in the Nemmers conference.

- Harrison, G.W. (2005). "Book Review: Advances in Behavioral Economics", *Journal of Economic Psychology*, 25, 793–795.
- Harper, D.G.C. (1982). "Competitive Foraging in Mallards: 'Ideal Free' Ducks", *Animal Behavior*, 30, 575–584.
- Heyman, J. and D. Ariely (2004). "Effort for Payment: A Tale of Two Markets", *Psychological Review*, 15, 787–793.
- Laibson, D. (1996) "Hyperbolic Discount Functions, Undersaving, and Savings Plans", NBER Working Paper 5635.
- Levitt, S. and S. Dubner (2005). *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, William Morrow (a HarperCollins imprint).
- Rubinstein, A. (2001). "A Theorist's View of Experiments", *European Economic Review*, 45, 615–628.
- Rubinstein, A. (2003). "Economics and Psychology? The Case of Hyperbolic Discounting", *International Economic Review*, 44, 1207–1216.
- Rubinstein, A. (2004a). "Dilemmas of An Economic Theorist", mimeo.
- Rubinstein, A. (2004b). "Instinctive and Cognitive Reasoning: A Study of Response Times", mimeo.
- Sanfey, A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom, and J.D. Cohen (2003). "The Neural Basis of Economic Decision-Making in the Ultimatum Game", *Science*, 300, 1755–1758.
- Shaked, A. (2005). "The Rhetoric of Inequity Aversion", mimeo.
- Simonson, I. (1990). "The Effect of Purchase Quantity and Timing on Variety Seeking Behavior", *Journal of Marketing Research*, 27, 150–162.
- Strotz, R. H. (1956). "Myopia and Inconsistency in Dynamic Utility Maximization", *Review of Economic Studies*, 23, 165–180.