A Foundation for Behavioral Economics

Jessica L. Cohen and William T. Dickens
The Brookings Institution

Paper Presented at the Meetings of the American Economics Association

December 2001

Later Published in the *American Economic Review*, May 2002.

A Foundation for Behavioral Economics

Jessica L. Cohen and William T. Dickens*

The core theory of behavior in Economics, which structures inquiry and provides a framework for empirical analysis, is largely responsible for the success of the discipline. Behavioral Economics (BE) challenges this theory, but has failed to provide a coherent alternative. Consequently the influence of BE has been limited. In what follows we argue that Evolutionary Psychology (EP), suitably adapted, can provide at least a partial foundation for BE. Its methods offer a way of generating theories of the origins of anomalous behaviors and of testing those theories.

I. Behavioral Economics

BE has been most successful in documenting failures of the rational actor model (e.g. failures of expected utility theory, irrational cooperation, and time inconsistent preferences). However, attempts to incorporate these observations into theory have been *ad hoc*: either an anomalous behavior is induced by modifying the utility function or the behavior is simply assumed and implications derived. The lack of theoretical foundations causes a number of problems for BE. First, empirical analysis can show the inadequacy of mainstream theory, but it does little to help develop alternatives. Second, without a coherent theory it is difficult to develop new applications. Third, the policy influence of BE is limited by its inability to predict, beyond what has been observed, circumstances in which anomalous behavior will arise or how it will respond to policy changes. Finally, it is hard to judge the welfare implications of policy if we do not understand the origins of such behavior.

Bounded rationality has been proposed as a theory of BE, but has been mostly used to provide *ex post* rationalization for sub-optimal behavior, and has not been very successful in generating new theory or anticipating anomalies. This failure is inherent in the premise of the theory—that information-processing capability is a scarce resource. A theory of the allocation of reasoning capacity should describe an equilibrium if it is to apply to anything other than momentary states, and only a theory of the optimal allocation of capacity could be consistent with equilibrium. However, any attempt to compute an optimal allocation encounters a paradoxical infinite regress (Winter 1964, p. 252). To obtain an optimal allocation of rationality one must first allocate some of the resource to compute the optimal allocation. The decision of how much to allocate requires a prior allocation, which is itself another decision requiring a prior allocation and so on. Since an optimal allocation is impossible, bounded rationality is therefore a theory of ad hocracy—good enough decisions and rules of thumb—providing no insight into the nature of such rules.

Cultural evolution could provide a positive theory of behavior in a world of necessarily imperfect rationality, but there are problems adapting that theory to economics. Models of cultural evolution typically assume automatic behavior and extremely slow learning—assumptions that are irreconcilable with both standard economic theory and our experience of our own behavior. Further, even with cultural evolution as a theory of how rules of thumb emerge and change, bounded rationality has a number of additional shortcomings as a foundation for BE.

First, if limited information processing capacity was the primary reason for most anomalies, providing information or computational assistance should eliminate the

problem behaviors, but it often doesn't. Second, many anomalies appear universal in a way that is hard to square with them being due to *ad hoc* rules of thumb. Third, emotions play an important role in many anomalous behaviors in a way that cannot be anticipated by a theory that blames the anomalies on *ad hoc* adaptation to limited information processing ability.

II. Emotions

Emotions play an obvious role in anomalies arising from motivation conflict such as cognitive dissonance or impulse control. However, there is evidence that emotions play a role in information processing anomalies as well, several of which arise because information is ignored or improperly weighted in decision-making.

Damasio (1994, p. 170) describes how emotional "somatic markers" seem to signal the results of unconscious processes that judge the salience of different information to a decision. Emotions also seem to play a role in evoking rational thought. Most of us act habitually most of the time, but all of us have had the experience of being shocked into reconsidering some habitual behavior when we perceive that circumstances make the behavior inappropriate. The experience is not of rationally detecting some change. Rather, we experience a startling sense that something is wrong before we have any idea what it might be, which starts our minds racing to diagnose the problem.

Emotions also seem to play a role in stopping rational processes. People with damage to their frontal lobes are emotionally flat but cognitively able. They can solve finite logic problems, but have trouble stopping when faced with open-ended problems. Damasio (1994, p. 37) describes one such person who, when asked to choose between

two dates for an appointment, went on for over 15 minutes raising one consideration after another, only stopping when told which day to come.

The universality of emotions, their involvement with many behaviors necessary for survival, and their somatic roots and expression suggest a biological basis. It seems reasonable to speculate that both our ability to experience emotions and many of the specific functions emotions serve are biological in origin, and thus have likely evolved. If understanding the role of emotions and other biological influences on tastes and capabilities is crucial to developing a positive theory of BE, then understanding how they have been shaped by evolution may be the basis for such a theory.

III. Evolutionary Psychology

EP arose as a way to explore the reasons for universal human behaviors that defy simple instrumental explanations.[1] The point of departure for any EP inquiry is to consider what survival/reproductive advantage a behavior might have bestowed on a hunter-gatherer in the Pleistocene. This epoch is chosen because it is the only period in human history long enough to have allowed substantial evolutionary adaptation.

Past uses of evolutionary theory to explain behavior have been criticized as unscientific post-hoc rationalization, but the methodology of EP goes further. Typically the theory is used to generate surprising predictions about contemporary behavior that can be tested in a laboratory. For example, an implication of the theory of cooperation with non-kin is that individuals must ultimately expect to get something in return for such behavior. If true, then people would have to be very good at detecting when others were not reciprocating to get any advantage from such behavior. This leads to the prediction that people should have evolved to be very good at detecting cheating in social exchange.

The best example of the success of EP is the large number of experiments demonstrating that people are uncannily better at solving complex logic problems when they are framed as detecting cheating then when the problem takes any other form (Cosmedes and Tooby 1992).

There is a problem adapting the methods of EP to Economics. Consistent with how we experience our own behavior, Economics views individuals as making choices. In contrast, in evolutionary models behavior is instinctual – it is simply the consequence of our genetic make-up. If we are going to use EP to rationalize BE we will need to reconcile instinct and choice.

IV. Instinct and Choice

In economics if two people behave differently it is because of differences in their tastes or the opportunities available to them. A natural first step towards introducing genes into economics would be to assume that they affect our tastes and our capabilities. This way of thinking about how genes affect behavior is more consistent with economic thinking and with how we experience our own behavior than the view of behavior as instinctual. More important, many behavioral economic experiments show that anomalous behaviors, while surprisingly consistent, can sometimes be shaped in complex ways by economic incentives. The approach to understanding behavioral anomalies that we are proposing would allow a positive theory of the malleability of the anomalous behavior.

It is obvious how this approach can be useful for structuring analysis of behavioral anomalies that can be induced by changing the utility function, but we think that an evolutionary approach can also elucidate anomalies related to information

processing. We may not be able to be rational about the allocation of our computational capabilities, but it is possible that we could have evolved mechanisms to accomplish very efficient allocations. However, these mechanisms would have been adaptive in different circumstances than those we face today.

We previously described how emotions and unconscious processes seem to be involved in initiating rational thought processes, screening the information used in those processes, and deciding when the process should end. Thinking about the circumstances in which such regulatory mechanisms evolved could shed light on their design. Further, mechanisms that were well adapted to life in hunter-gatherer tribes might be maladaptive today when circumstances demand very different sorts of reasoning. Considering those differences could help us anticipate when our decision-making will be inadequate. Finally, understanding what cues rational rather than habitual action, what makes some information particularly salient, and what signals sufficient consideration of a problem, could illuminate when behavioral anomalies will be flexible and when they will not, and how they might be remedied or circumvented.

This view of behavior is profoundly different from the standard view in economics, but often won't lead to different predictions. In many cases where there are opportunities for learning we expect behavior to converge to what would be predicted by the model of the omniscient omnipotent optimizer (albeit with a utility function that looks different from that of the standard model). This is a desirable characteristic of a behavioral framework since that model often predicts well in such circumstances. However, taking an evolutionary approach to understanding our preferences, abilities,

and thought processes may help us generate a theory of the limits of rationality and how those limits might change.

Economists who borrow evolutionary methods for developing theory will also want to borrow EP's method for testing those models. While traditional methods can still be used, looking for surprising predictions of evolutionary models that can be tested in labs is a powerful additional method.

V. An Example

Bewley (1999) finds that employers expect wage cuts to provoke angry reactions from workers. The role of anger in this reaction suggests the possibility that cuts are provoking an instinctual reaction with an evolutionary explanation. What follows is a sketch of how one might apply the method we are proposing.

We can reasonably speculate that Bewely's workers and Pleistocenian hunter-gathers faced social settings in which cooperation could be beneficial, but exploitation of cooperation by a partner could be costly. Generating cooperation in Prisoner's Dilemma games is difficult, particularly when the game is played anonymously with a finite number of repetitions. Yet experiments show that people often cooperate in these circumstances (Fehr and Gächter, 2000). At least part of the reason for this cooperation is that people fear punishment. However, in the case of finitely repeated and/or anonymous games this requires agents to follow through on punishments that are not credible (e.g. punishments in the last round of the game or in circumstances in which there is no opportunity for reputation).

Punishing under these circumstances seems to involve a display of anger. Suppose we interpret the emotional state of being angry as involuntarily induced by

cognition of being cheated in a social exchange, and that its effect is to make those experiencing it suffer negative utility unless they punish the suspected cheater. The ability to experience anger becomes functional in this context, since it communicates a commitment to punish even in circumstances where punishment is "sub-optimal", and thus facilitates cooperation.

If we stopped here we too could be accused of post-hoc rationalization, but we would require an evolutionary explanation to pass two additional tests. We would want to be able to show rigorously that such a genetic tendency to experience anger could have invaded and stabilized in a population, and we would want to find implications of the theory that could be tested. We have not yet developed a model that would accomplish the first task, but we believe that the evolutionary model of reciprocal preferences presented in Bowles and Gintis (2000) could be reinterpreted with minor modifications and extended to demonstrate how anger at cheaters could spread through a population and be sustainable in equilibrium.

We would also want to test the hypothesis. A test that is parallel to that employed by Tooby and Cosmides (1992) to investigate cheater-detection would be a test of anger-detection in the context of social exchange. Experiments have shown that people are good at discerning between pictures of people experiencing real emotions and those faking emotions. If we are right that anger is crucial to facilitating exchange then people may be particularly good at judging the authenticity of anger when the judgement is called for in the context of a social exchange problem.

Returning to our motivating question, if wage cuts induce anger because they are viewed as defaults on exchange, then we cannot expect this reaction to be easily

“unlearned.” If anger imposes a utility cost when punishment is not meted out, then people will be able to restrain their anger in situations where expressing it could be very costly, but anger-induced behavior will appear irrational. In fact, if anger originated as a commitment device it will be valuable only if it is very difficult to suppress. Thus, if our hypothesis about wage cutting is correct, we would not expect resistance to cuts to be easily overcome or circumvented.

VI. Conclusion

BE suffers for lack of theoretical foundations, but the methods of Evolutionary Psychology may help provide a foundation. Understanding how we evolved to make the best use of our limited capacity for rationality may be the key.

References

Barkow, J., Cosmides, L. and J. Tooby, eds. *The Adapted Mind*. Oxford: Oxford University Press, 1992.

Bewley, Truman, *Why Wages Don't Fall During a Recession*, Cambridge: Harvard University Press, 1999.

Bowles, Samuel and Herbert Gintis, "The Evolution of Reciprocal Preferences," University of Massachusetts Working paper, 2000.

Cosmides, L. and J. Tooby, "Cognitive Adaptations for Social Exchange," in *The Adapted Mind*. Barkow, J., Cosmides, L. and J. Tooby, eds. Oxford: Oxford University Press, pp. 163-228, 1992.

Damasio, Antionio R., *Descartes' Error*, New York: Putnam, 1994.

Fehr, Ernst and Simon Gachter "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, Summer 2000, 14(3), pp. 159-182.

Winter, Sidney G., Jr. "Economic 'Natural Selection' and the Theory of the Firm," *Yale Econ. Essays*, 1964, 4(1),pp.225-72.

Notes

* Cohen: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142; Dickens: Brookings, 1775 Massachusetts, Washington DC 20036. The material in this paper is taken from our longer paper "Instinct and Choice in Cynthia Garcia Coll ed. *NATURE AND NURTURE: THE COMPLEX INTERPLAY OF GENETIC AND ENVIRONMENTAL INFLUENCES ON HUMAN BEHAVIOR AND DEVELOPMENT* Erlbaum (forthcoming).

1. See Barkow et. al. (1992) for a number of examples.