Neuroeconomic Theory:
Using Neuroscience to Understand the Bounds of Rationality

Isabelle Brocas and Juan D. Carrillo
(University of Southern California, CEPR and TREND Institute
http://www.neuroeconomictheory.org)

Published in http://www.voxeu.org/index.php?q=node/4758

Economic agents often fail to make 'rational' decisions. They are subject to multiple biases that affect the way they perceive events, act upon them and learn from experience. Most of these anomalies are extraordinarily recurrent and documented in real world and laboratory environments by behavioral data at the individual and aggregate levels. To cite just a few, individuals have systematically biased beliefs about the prospects of their entrepreneurial endeavors, under-save for retirement, and subscribe to mortgages they cannot afford.[1] These behaviors cannot be ignored since they may have disastrous consequences for the economy, as we recently witnessed during the subprime mortgage crisis. Knowing what type of mistakes and biases are prevalent is an important starting point. However, the main challenge is to understand why they emerge so that they can be predicted and, possibly, avoided.

Economics has always relied on a careful modeling of decision-makers. They are described by utility functions that represent their goals, and they interact at (Nash) equilibrium. The discrepancies between theoretical predictions and observed behavior have haunted the field for many decades. To cope with this problem, behavioral economists have developed new theories of decision-making capable of fitting behavioral data better than traditional models. The methodology consists in building models to demonstrate the relationship between a cause (e.g., a preference over an object or a cognitive limitation) and a behavioral anomaly. This line of research formulates possible explanations for behavioral data, but it is subject to some shortcomings. In particular, the cause is generally not observable, and there is no evidence of the relationship provided by the model. Most notably, the freedom provided by the introspection method leads to a model selection problem. Also, the cause of the behavioral anomaly may simply lie elsewhere.

Neuroeconomics offers a solution through an additional set of data obtained via a series of measurements of brain activity at the time of decisions. *Experimental neuroeconomics* can be seen as a subfield of experimental economics, where behavioral data is enriched with brain data. *Neuroeconomic theory* proposes to build brain-based models capable of predicting observed behavior.

Experimental neuroeconomics has been the object of a passionate debate between those who claim that this new data is irrelevant for economics and those who think it is essential.[2] Given the young age of the field, it seems early for a definite assessment. The discussion has also been centered on empirical issues regarding the collection method, amount, cost, and quality of brain data whereas, quite surprisingly, the broad implications have not received as much attention. Indeed, the new set of

[1] See e.g. the series "Anomalies" in the *Journal of Economic Perspectives* 1987-91 and 1995-2001.We will not discuss here whether these behaviors are 'irrational' or simply result from preferences different from those that are usually assumed.
[2] See e.g. Gul F. and W. Pesenderfor (2008) "The Case for Mindless Economics" in: The Foundations of Positive and Normative Economics, by A. Caplin and A. Schotter (eds.). Oxford U. Press and Camerer, C., Loewenstein, G. and D. Prelec (2005) "Neuroeconomics: How Neuroscience Can Inform Economics", *J. of Economic Literature*, 43, 9-64.

data provided by experimental neuroeconomics will shed light on the causes of behavior (and therefore of the behavioral anomalies) and help build new theories capable of explaining and predicting decisions, a long term goal of economics. Neuroeconomic theory offers to do precisely this. So far, research in that direction has been very limited and its impact has been largely ignored.

The objective of neuroeconomic theory is to build models based on evidence from the brain sciences, such as experimental neuroeconomics, but also other fields in neuroscience and neurobiology. Measurement of brain activity provides information about the underlying mechanisms used by the brain during choice processes. In particular, it shows which brain regions are activated when a decision is made and how these regions interact with each other. This knowledge can then be used to build a model that represents this particular mechanism. Contrary to behavioral economics, the model does not rely on introspection or plausible assumptions but rather on an existing and documented biological property of the brain.

The methodology used in neuroeconomic theory has two advantages. Primarily, evidence from the brain sciences provides precise *guidelines vis-à-vis the constraints that should be imposed on decision-making processes*. This can help unfolding the "true" motivations for the "wrong" choices and improve the predictive power of the theory. Behavioral theories that account for biases in judgment build on specific models of preferences over beliefs or non-Bayesian updating processes. Rather than guessing a cause for biases, neuroeconomic theory builds a model based on the existing physiological properties underlying learning and belief formation. In principle this can help pinpoint biological foundations for anomalous choices. For example, research in neurobiology demonstrates that the brain cannot encode all the information contained in a signal. A decision is triggered when 'enough' information supporting one alternative is obtained and the brain uses a variety of biological mechanisms to filter information in a constrained optimal way. In a recent paper we show that these properties of the brain result in a behavioral tendency to confirm initial priors.[3] Behavioral data reports precisely that individuals stick too often to first impressions. These confirmatory biases may all emerge from the same set of physiological information processing constraints. Further work in that direction may help uncover the causes of other biases and determine whether they are all related to the same physiological limitations.

The second advantage is that by explicitly modeling physiological properties, it is possible to provide *foundations for some elements of preferences traditionally considered exogenous*, such as risk-aversion, ambiguity-aversion or time-preference rates. Choices involving risk, uncertainty or time-delays may require complex trade-offs. Measures of brain activity allow us to determine if the evaluation process is centralized or if different brain systems compete to influence the final decision. Neuroeconomic theory proposes to model the actual brain organization, determine the behavior that emerges from it, and evaluate which theory fits best. Discounting provides an illustrative case study. The standard neoclassical theory derives time-preference rates from a set of axioms on the preferences of individuals. A nice property of these axioms is that discounting must be represented by a time-consistent function. To account for the observed tendency of individuals to procrastinate, behavioral economists have modified this function by introducing a parameter of time-inconsistency whereas decision theorists have modified the original axioms. In both cases, the motivation for the new theories is a behavioral observation that cannot be reconciled with the original theory. Instead, our recent research uses neurobiological evidence to model inter-temporal choices as the result of a conflict between two brain systems, one interested in immediate gratification and one that can form a mental representation of future rewards. Using this approach we are able to derive from first

[3] Brocas, I. and J.D. Carrillo (2009) "From Perception to Action: an Economic Model of Brain Processes" *mimeo, USC.*

principles three properties of dynamic choices commonly observed in the data: positive discount rate, decreasing impatience, and heterogeneity of discount rates across activities.[4] A similar methodology can be applied to rationalize other observed characteristics of preferences.

In conclusion, neuroeconomic theory will soon play a crucial role in the building of new reliable theories capable of explaining and predicting individual behavior and strategic choices. The main message is that the individual is not one coherent body. The brain is a multi-system entity (with conflicting objectives, restricted information, etc.) and therefore the decision-maker must be modeled as an organization. We conclude with an analogy. Before the so-called modern theory of the firm, organizations were modeled as individual players characterized by an input-output production function. The systematic study of interactions between agents and decision processes within organizations (acknowledging informational asymmetries, incentive problems, restricted communications channels, hierarchical structures, etc.) led to novel economic insights. Applying a similar methodology to study individual decision-making is, in our view, the most fruitful way to understand the bounds of rationality.

[4] Brocas, I. and J.D. Carrillo (2008) "The Brain as a Hierarchical Organization" *American Economic Review* 98(4), 1312-46.