Disciplining behavioral theories through brain-based models of decision-making

Isabelle Brocas and Juan D. Carrillo
University of Southern California
and Centre for Economic Policy Research

Mark Dean noted in his chapter that some economists are reluctant to embrace Behavioral Economics as they do not find much value in extended models that depart from the standard rational approach. In this commentary, we argue that Neuroeconomics addresses precisely some of these criticisms of Behavioral Economics (for a more detailed discussion, see Brocas and Carrillo (2008a)).

Let us borrow the terminology of the chapter. If we think of an economic model as a mapping f between the set of environmental conditions **X** and the set of behaviors **Y**, standard 'rational' economics presupposes the existence of a well-behaved function f^*. Unfortunately, that function often fails to represent the observed realizations of elements from these sets. To cope with this mismatch, behavioral theories have spawned new models, that is, new functions f^{**} that are a better fit to the data.

This line of research has helped clarify several economic phenomena but is subject to a few shortcomings. First, f^{**} remains a shortcut of the true unobserved underlying process. As it is not possible to observe directly the causes for the better performance of f^{**}, the motivation remains "inspirational". Second, there often exist multiple mappings that can fit the sample of interest, echoing different motivations. A new sample will then be compatible with only a subset of those mappings or will require yet other mappings. As a consequence, even though f_1^{**} may work well in some situations, it may be unclear why this is the case. Moreover, there may be an observationally equivalent f_2^{**} that does as well and is closer to the true underlying process. Naturally, if two models worked equally well in all samples, choosing between the two would not be relevant from a purely economic perspective, but otherwise it is important to reveal the best mapping. Overall, behavioral theories are often criticized on the grounds that they impose much less discipline than standard axiomatic approaches.

Neuroeconomics offers a solution through an additional set of data obtained via measurements of brain activity during choice. These measurements show which brain regions are activated when a decision is made and how these regions interact with each other. Therefore, they help us understand how information about elements of **X** is processed and acted upon to deliver a decision in **Y**. This knowledge, in turn, can be used to build a model that represents the particular mechanism followed by brain processes and that is capable of predicting observed behavior. Contrary to behavioral theories, the model does not rely on "armchair inspirations" or "plausible assumptions" but rather on existing and documented biological properties of brain processing. Said differently, the methodology helps deriving f^{**} as the result of an underlying brain mechanism.

The methodology used in neuroeconomic theory has two advantages. First, evidence from the brain sciences provides precise *guidelines vis-à-vis the constraints that should be imposed on decision-making processes*. This can help uncover the "true" motivations for the "wrong" choices and improve the predictive power of the theory. For example, behavioral theories that account for biases in judgment build on specific models of non-Bayesian updating. Instead, neuroeconomic theory builds a model based on the documented physiological properties underlying learning and belief formation. In principle this can help pinpoint biological foundations for anomalous choices. Indeed, research in neurobiology demonstrates that the brain cannot encode all the information contained in a signal. If we consider a simple binary choice problem, a decision is triggered when 'enough' information supporting one alternative is obtained. The brain uses a variety of biological mechanisms to filter information in a constrained but efficient way. In a recent paper (Brocas and Carrillo, 2011) we show that these properties of the brain result in a behavioral tendency to stick too often to first impressions, to have opinions that are prone to polarization, and to hold beliefs that are affected by the order in which information is presented. The existing evidence about choices supports all of these behavioral biases.

In a related vein, behavioral theories build on "cost" specifications to rationalize observed performance discrepancies in multi-tasking environments. Even though the theories can be adjusted to fit the data well, they do not come with any evidence that such costs exist. By contrast, it is possible to rely on neuroscience evidence to model the actual processes behind multi-tasking. In Alonso, Brocas and Carrillo (2011) we model the brain as an organization in which a coordinator allocates the (scarce) resources to systems responsible for different tasks. The paper shows that it is optimal to impose on each system a resource limit that depends negatively on the amount of resources needed by other systems. Performance is therefore not simply the result of processing costs but rather the solution of a constrained optimization problem. Interestingly, the model provides foundations for inertia. Indeed, the optimal dynamic allocation rule is such that the amount of current resources allocated to a system is increasing in its past needs.

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 delay 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. 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. In a sense, models are modified to fit behaviors. 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 another that can form a mental representation of future rewards. Using this approach we derive from first principles three properties of dynamic choices commonly observed in the data: positive discount rate, decreasing impatience, and heterogeneity of discount rates across activities (Brocas and Carrillo, 2008b).

In conclusion, neural data should be regarded as an invaluable resource for understanding the underlying processes leading to choices. They encourage the formulation of new rigorous and reliable theories capable of explaining and predicting individual behavior and strategic choices in a disciplined way. Brain-based models of decision-making will help identify the reduced-form model f^{**} that should be used in each setting and the reasons why this is the case.

References

Alonso, R., Brocas, I. and J.D. Carrillo (2011) "Resource Allocation in the Brain", *CEPR D.P. 8408*.

Brocas, I. and J.D. Carrillo (2008a) "Theories of the Mind", *American Economic Review: Papers & Proceedings* 98(2), 175-180.

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

Brocas, I. and J.D. Carrillo (2011) "From Perception to Action: an Economic Model of Brain Processes", forthcoming in *Games and Economic Behavior.*