American Economic Review: Papers & Proceedings 2008, 98:2, 175–180
http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.2.175

Theories of the Mind

By Isabelle Brocas and Juan D. Carrillo*

Economics has traditionally relied on revealed preferences (and, occasionally, on verbal reports) to understand the desires of people. Another source of information has been developed in recent years: the direct observation of choice processes. This mechanism, possible thanks to the improvements in the designs and techniques to measure brain activity, is explored in the burgeoning field of experimental neuroeconomics (see Paul W. Glimcher and Aldo Rustichini (2004) and Colin Camerer, George Loewenstein, and Drazen Prelec (2005) for recent surveys).

In this article, we argue that the evidence on brain activity can also be used to build theoretical models that help us understand choices and predict behaviors. We label this research "Neuroeconomic Theory." In Section I, we describe the procedure, discuss some advantages over traditional methodologies, and establish some facts that motivate our approach. In Section II, we illustrate the methodology with two brain-based models of decision making.

I. What is "Neuroeconomic Theory"?

Neuroeconomic theory is an interdisciplinary line of investigation that combines research from neuroscience, neurobiology, and economics. Experimental neuroscience and neurobiology provide detailed evidence of the functionality, interconnectivity, and physiological constraints of the brain systems involved in decision making. Microeconomic theory supplies the toolkit to build simple optimization models that incorporate these network interactions and well-defined constraints into the mechanisms of choice.

* Brocas: Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089, and CEPR (e-mail: brocas@usc.edu); Carrillo: Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089, and CEPR (e-mail: juandc@usc.edu). For more information on the "neuroeconomic theory" project, visit the Web site of our laboratory TREND at http://www.neuroeconomictheory.org. We thank Ignacio Palacios-Huerta, Francesco Sobbrio, and Pierre-Olivier Weill for comments.

A. *An Alternative Approach*

Neuroeconomic theory provides parsimonious models of decision making capable of delivering qualitative behavioral predictions. Models are necessarily simplifications of an intricate reality. The objective in this research is not to model the physiological elements involved in a brain process (neurones, synapses, neurotransmitters) but, instead, to capture the fundamental properties of those processes. The models are still "as-if" representations of reality, and build on the assumption that a process is more likely to flourish if it utilizes its resources efficiently. So, in this respect, neuroeconomic theory follows the economic modelling tradition, and departs from the computational models of neural systems developed in neuroscience.

We argue that there are at least three advantages in approaching economic decision making from a neuroeconomic theory angle. The first one is rigor. Bounded rationality can be modeled in many different ways. The evidence from neuroscience provides precise guidelines vis-à-vis the constraints that should be imposed on decision-making processes. As an illustrative example, our limited ability to process information leads to judgment biases. To understand these biases, one may build widely different behavioral theories based on introspection and casual evidence: models with exogenous information processing costs, utility representations that incorporate preferences about information, non-Bayesian information updating technologies, etc. Naturally, each theory may have a different prediction, and may also presage a different behavior in other contexts. Which one is the most appropriate then? Although the criterion is ultimately empirical, neuroeconomic theory proposes to use the evidence from neurobiology to guide the modelling choices. For instance, in this example, the properties of neuronal cell firing that transform sensory perceptions into voluntary actions impose specific restrictions on the way individuals process information. In Brocas and Carrillo (2007), we incorporate

these constraints in an otherwise standard model of decision making.

The second advantage is foundational. By explicitly modeling the interactions between different brain systems, it is possible to provide microfoundations for some aspects of preferences traditionally considered exogenous. For example, it is well understood that, with some exceptions, individuals put lower weight in future events than in current events. Finer aspects of discounting are captured by exogenous parameters, even though they are supposed to be intrinsic elements of preferences. Neuroeconomic theory can help identify the systems involved in intertemporal decision making and derive discounting as a result of their interactions. This, in turn, may help understand more subtle features of time preferences (see Result 1 below).

A third advantage is the increased possibility of feedback between theory and experiments. Brain-based theoretical models provide new testable implications about the contribution of brain systems to different aspects of choice processes. Importantly, theoretical models are not as constrained as experimental models. It is always possible to make a theoretical environment more complex, but it is often difficult to implement it in the laboratory. Interesting theoretical predictions help identify which new experiments are worth conducting. Once these issues are addressed experimentally, other theoretical challenges are likely to emerge.

B. *Traditional Studies of Behavior*

Decision theory analyzes the general properties of individual decision making. The methodology broadly consists of defining axioms on individual choice and then providing utility representations that satisfy those axioms. Researchers in experimental economics have tested these theories and have documented significant violations of the axioms and the behaviors they predict. To address these empirical shortcomings, decision theorists have proposed new axioms, such as uncertainty aversion or set-betweenness. More recently, behavioral economists have also provided alternative models of decision making based on findings in psychology. The methodology here is quite different, as it generally presupposes the existence of utility functions, without deriving them from first principles.

A key aspect of decision theory is that it draws inferences exclusively from choices. There are two strands within this literature. The preference-based approach poses axioms on (non-observable) tastes, while the choice-based approach poses axioms on (observable) choices. If the rational paradigm holds, consistent choices correspond to rational preferences, and both formulations are equivalent. However, if the rational paradigm does not hold, there is not such a clear mapping between the choice-based and the preference-based theories of decision. We argue that, in this case, there is scope for a third approach that complements these two.

When the researcher focuses on choices and refrains from drawing inferences about preferences, theories are limited to the situations in which the choice of interest is observed. As a result, understanding why a behavior occurs in one particular situation and not in others is out of the scope of the analysis. In our view, this approach puts too much weight on categorizing situations and describing behaviors, and too little on understanding the underlying relationships between situations and behaviors.

When the researcher draws inferences about preferences from the observed choices, the modeling of these preferences is speculative. As argued in Section IA, the main problem is the existence of one rational paradigm and countless ways to depart from it. Different models can predict a given "non-rational" choice, but they will each provide a different prediction in other situations.

The traditional way to discriminate between models has been by comparison of choices across experiments. With the direct observation of choice processes (and therefore of the physiological constraints when acquiring and processing information), neuroeconomic theory provides a supplementary tool. The researcher is not bound anymore to choose between studying choices and studying preferences; nor does he have to speculate on how to model preferences in the latter case.

C. *Evidence of the Brain as an Organization*

The premise of neuroeconomic theory is the existence of multiple brain systems. Neuroscientists and neurobiologists support the idea of brain modularity. Some well-known processes affected by modularity include memory

(declarative versus procedural; see Russell Poldrack and Paul Rodriguez 2004), regulation of cognition (performance monitoring versus implementation of control; see Earl K. Miller and Jonathan D. Cohen 2001) and reward evaluation (immediate versus future; see Antoine Bechara 2005). In these and other settings, researchers have identified double dissociations using either imaging techniques or patients with brain lesions. At the same time, neuroscientists argue against a rigid, one-to-one mapping between system and function: each system performs different functions and each function needs the intervention of several systems.

There is also support for "strategic interactions" between systems. Different brain systems perform different, and sometimes incompatible, functions. As a result, the brain may have to select among competing options. Information processing is a typical example. Experimental research suggests the existence of a brain system that monitors the occurrence of conflicts and passes the information on to the centers that implement control over actions (Matthew M. Botvinick et al. 2001). This, however, does not imply that decision making is a fully decentralized process with numerous participants. Instead, for each particular decision, only a few systems play significant roles.

Finally, some readers may think that evolution should favor the development of single systems of increasing complexity and flexibility. Evolutionary biologists argue, on the contrary, that multiple systems will be the result of an evolutionary process whenever an adaptation that serves one function cannot, because of its specialization, serve other functions. In a classical work, David F. Sherry and Daniel L. Schacter (1987) discuss the case of memory, where selection pressure has resulted in the development of one system that encodes habits and another system that encodes episodic memories. Note, however, that natural selection promotes neither perfection nor optimal adaptation. Adaptation is constrained by the available heritable components. Also, it does not anticipate future requirements, and therefore does not equip organisms with the means to meet them.

II. Modelling the Brain

In this section, we delineate two simple models that incorporate some brain-based constraints into an otherwise standard individual decision-making problem. The main goal is to illustrate the kind of conclusions that can be obtained using our approach. Sections IIA and IIB closely follow Brocas and Carrillo (forthcoming). Due to space considerations, we present only an overview of the results. We refer the reader to the paper for the more exhaustive analysis.

A. *Discounting as an Information Problem*

The literature in neuroscience provides evidence of informational asymmetries in the brain. Because neural connectivity is a limited resource, most brain areas are unidirectionally linked to others. These restrictions act as physiological constraints on the flow of information, and result in limited awareness of motivations for decisions. As an example, studies show activation of the ventral striatum in response to learning, even in the absence of conscious knowledge (Scott L. Rauch et al. 1997). It has also been argued that two different systems, the amygdala and prefrontal cortex, are responsible for evaluating information related to immediate and future prospects (Bechara 2005). This view suggests a temporal evaluation conflict between an impulsive and a reflective system.

In our first model, we use a multiple brain systems approach of individual decision making to study "informational" and "temporal" conflicts. We consider a simple, consumption and labor setting. The individual lives T periods. At each period $t \in \{1,\ldots,T\}$, he works $n_t \in [0,\overline{n}]$ and consumes $c_t \geq 0$. For each unit of labor, he obtains one unit of income that can be spent in any period. Markets are perfect and the interest rate is positive. To incorporate the conflicts and following the evidence from neuroscience, we split the individual into two systems. First, a myopic, informed system (called "agent," he) whose preferences at date t are summarized by:

$$U_t = \theta_t u(c_t) - n_t,$$

with $u' > 0$ and $u'' < 0$. Also, $\theta_t \in [\underline{\theta},\overline{\theta}]$ is privately known and represents the marginal value of consumption at t. Second, a forward-looking, uninformed system (called "principal," she) who weighs equally the utility of all agents. She maximizes the expected utility of all agents:

$$S = \sum_{t=1}^{T} E[\theta_t u(c_t) - n_t],$$

under the budget constraint that links lifetime consumption and lifetime labor.[1] If the principal knew θ_t, the "existence" of agents would be irrelevant. Given the positive interest rate, she would concentrate work in early periods and consume at each date according to the marginal valuation. The consumption and labor decision would thus be, to some extent, independent. More interesting is the asymmetric information case. There, the principal proposes at each date a menu of incentive compatible pairs $\{(c_t(\theta_t), n_t(\theta_t))\}_{\theta_t \in [\underline{\theta}, \bar{\theta}]}$ and allows the agent to decide which one is picked. Overall, the problem is analogous to a contract with hidden information. We obtain the following result.

RESULT 1: *At each date, higher consumption is allowed in exchange for higher labor. Intertemporal choices exhibit properties consistent with positive discounting, decreasing impatience, and higher impatience for goods where the valuation has higher variability.*

By definition, the myopic and forward-looking systems disagree on the desired levels of consumption and labor, so the principal cannot impose her preferred choices. It is not optimal to delegate the decision to the agent either, because the latter will overconsume and underwork. Another option would be to select the levels of consumption and labor that maximize her expected utility. Result 1 shows that the optimal strategy is different: it consists of offering several pairs characterized by a positive link between consumption and labor within each period, and letting the agent choose. Overall, a self-disciplining rule of the form "work more today to consume more today" emerges *in equilibrium* as a response to the temporal and informational conflicts.

Consumption in this model exhibits properties that are consistent with recent theories of discounting. By construction, the choice when θ is known to the principal is equivalent to that of an individual without conflict and no discounting. The choice when θ is unknown is characterized by increased early consumption, so it is observationally equivalent to that of an individual without conflict and positive discounting. In other words, Result 1 suggests that discounting can be endogenously derived from the primitives of the model (informational asymmetry) rather than imposed as an ad hoc feature of preferences. More interestingly, the consumption pattern is also consistent with decreasing impatience, that is, with a period-to-period discount rate that falls monotonically. Finally, the theory has a third testable implication: individuals exhibit most impatience in activities where the marginal valuation of consumption has the greatest variability.

B. *Self-Discipline as an Incentive Scheme*

Another strand of the neuroscience literature documents a discrepancy between brain systems in the relative importance attached to tempting versus non-tempting goods. Roughly speaking, the nucleus accumbens and amygdala tend to overemphasize the pleasure of tempting goods, while the prefrontal cortex is responsible for overriding ill-motivated impulses (Kent C. Berridge and Terry E. Robinson 2003).

Our second model investigates how this "incentive salience" conflict interacts with the "informational" conflict introduced previously. Formally, we abstract from the temporal dimension and assume that the individual allocates a fixed amount of resources between a tempting good x and a non-tempting good y during a single period. The cognitive system (or principal) is interested in optimizing consumption given how much each good is actually enjoyed. Her preferences are captured by

$$W(x, y; \theta) = \theta a(x) + b(y),$$

with $a' > 0$, $a'' < 0, b' > 0, b'' < 0$. The impulsive system (or agent) has a biased motivation, that is, a willingness to engage in excessive consumption of the tempting good compared to how much it is really liked. His preferences can be described by

$$V(x, y; \theta) = \alpha\, \theta\, a(x) + b(y) \quad \textit{with} \;\; \alpha > 1.$$

An interesting situation arises when the cognitive system can impose her preferred choices, but the impulsive system has a superior knowledge

[1] See Richard H. Thaler and Hersh M. Shefrin (1981) and Drew Fudenberg and David K. Levine (2006) for related two-system models (with full information and only one activity).

of θ. As before, the agent may use this private information to misrepresent his desires. Again, the problem boils down to a contract with hidden information, where the principal designs the menu of incentive-compatible pairs (x, y) that maximizes her expected utility $E_\theta[W(x, y; \theta)]$. Our second result is the following.

RESULT 2: *The reflective system sets a consumption cap for the tempting good but, otherwise, delegates the consumption choices to the impulsive system.*

Under complete information, biased motivations are irrelevant since the principal can impose her preferred choices. Under incomplete information, full delegation results in excessive consumption of the tempting good. However, even though the interests of principal and agent are not aligned, they are not opposed either. Increasing the consumption of the tempting good does not reduce the welfare of the principal, it simply does not increase it at the same rate. This is in contrast with the standard contracting literature. As a consequence, distorting the first-best allocation is not efficient, because it would require wasting valuable resources. Our model shows that, in the optimal contract, the intervention of the principal consists only in setting a maximum consumption for the tempting good, anticipating that the agent will incur some moderate excesses.

C. *Other Applications*

The conflicts introduced above can help rationalize other choices that may seem suboptimal. We now discuss briefly some natural extensions.

Choice Bracketing.—Consider an individual with fixed resources who allocates expenses between two pleasurable goods (e.g., entertainment and clothing). As in Section IIA, if decisions are delegated, the informed myopic system will overspend resources. Alternatively, the forward-looking but uninformed system can limit the per-period budget of each good to its expected optimal level. This ensures that resources are smoothly allocated over time, but it prevents consumption peaks whenever they are optimal. Following Result 1, we conjecture that the optimal solution will be to propose a menu of alternatives and allow the agent to choose which one he prefers. To satisfy incentive compatibility, higher levels of consumption of one good will be paired with lower levels of consumption of the other good. Thus, a rule of the form "spend less on entertainment this week if you want to spend more on clothing this week" arises in equilibrium.

Addiction.—Suppose that the endowment is allocated between a normal and an addictive good, where current consumption of the addictive good both decreases the future total utility and increases the future marginal utility of consuming it. As in Section IIA, the principal cares about the stream of utilities, whereas the agent cares only about current utility. This means that the conflict between the two systems is most problematic for the addictive good, making delegation particularly dangerous in this setting. Based again on previous results, we conjecture that the optimal strategy for the principal will be to allow full freedom for normal goods and enforce strict prohibition for addictive ones.[2]

Information Flows.—The incentive salience model presented in Section IIB rests on a bidirectional link between the motivationally biased agent and the welfare maximizing principal. Unfortunately, it is difficult to determine if this is the most appropriate assumption concerning neural connectivity. It may then be interesting to consider alternative formulations regarding how information flows between systems. For example, the problem can also be modeled as a "cheap talk" game (Vincent P. Crawford and Joel Sobel 1982). Borrowing from this literature, our conjecture is that the agent will choose to transmit the same message for all valuations within a compact set and different messages for valuations in different sets. Although, this mechanism seems somewhat abstract, it has a natural interpretation: it captures the tendency of individuals to reduce complex choices to a few options and pick one of them depending on the specifics of the situation.

[2] B. Douglas Bernheim and Antonio Rangel (2004) propose a different model of addiction, also based on neuroscience evidence.

III. Concluding Remarks

In recent years, research in the brain sciences has advanced significantly our understanding of brain functionality in choice processes. It is an exciting time to develop new ways of modeling decision making based on that evidence. The purpose of this article is twofold. First, it illustrates how the models and techniques used to analyze multiagent relationships can be applied (and adapted) to study individual decision making. Second, it discusses the strengths of this methodology as a complement to the more traditional approaches.

REFERENCES

Bechara, Antoine. 2005. "Decision Making, Impulse Control and Loss of Willpower to Resist Drugs: a Neurocognitive Perspective." *Nature Neuroscience*, 8(11): 1458–63.

Bernheim, B. Douglas, and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, 94(5): 1558–90.

Berridge, Kent C., and Terry E. Robinson. 2003. "Parsing Reward." *Trends in Neurosciences*, 26(9): 507–13.

Botvinick, Matthew M., Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen. 2001. "Conflict Monitoring and Cognitive Control." *Psychological Review*, 108(3): 624–52.

Brocas, Isabelle, and Juan D. Carrillo. 2007. "Reason, Emotion and Information Processing in the Brain." Center for Economic Policy Research Discussion Paper 6535.

Brocas, Isabelle, and Juan D. Carrillo. Forthcoming. "The Brain as a Hierarchical Organization." *American Economic Review.*

Camerer, Colin F., George Loewenstein, and Drazen Prelec. 2005. "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature*, 43(1): 9–64.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–51.

Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review*, 96(5): 1449–76.

Glimcher, Paul W., and Aldo Rustichini. 2004. "Neuroeconomics: The Consilience of Brain and Decision." *Science*, 306(5695): 447–52.

Miller, Earl K., and Jonathan D. Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience*, 24(3): 167–202.

Poldrack, Russell, and Paul Rodriguez. 2004. "How Do Memory Systems Interact? Evidence from Human Classification Learning." *Neurobiology of Learning and Memory*, 82(3): 324–32.

Rauch, Scott L., Paul J. Whalen, Cary R. Savage, Tim Curran, Adair Kendrick, Halle D. Brown, George Bush, Hans C. Breiter, and Bruce R. Rosen. 1997. "Striatal Recruitment During an Implicit Sequence Learning Task as Measured by Functional Magnetic Resonance Imaging." *Human Brain Mapping*, 5(2): 124–32.

Sherry, David F., and Daniel L. Schacter. 1987. "The Evolution of Multiple Memory Systems." *Psychological Review*, 94(4): 439–54.

Thaler, Richard H., and Hersh M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2): 392–406.