The Neuroeconomic Theory of Learning

By Andrew Caplin and Mark Dean*

Paul Glimcher (2003) and Colin Camerer, George Loewenstein, and Drazen Prelec (2005) make powerful cases in favor of neuroeconomic research. Yet in their equally powerful defense of standard "Mindless Economics," Faruk Gul and Wolfgang Pesendorfer (forthcoming) point to the profound language gap between the two contributing disciplines. For example, for an economist, risk aversion captures preferences among wealth lotteries. From the neuroscientific viewpoint, it is a broader concept related to fear responses and the amygdala. Furthermore, as economic models make no predictions concerning brain activity, neurological data can neither support nor refute these models. Rather than looking to connect such distinct abstractions, Gul and Pesendorfer (forthcoming) argue for explicit separation: "The requirement that economic theories simultaneously account for economic data and brain imaging data places an unreasonable burden on economic theories."

We share the conviction of Glimcher (2003) and Camerer, Loewenstein, and Prelec (2005) concerning the potential value of neuroeconomics, yet we believe that the field will live up to its potential only if a common conceptual language can be agreed upon. Hence, we face the Gul and Pesendorfer challenge head on by developing theories that simultaneously account for behavioral and brain imaging data. The principal innovation lies in our use of the decision theorists' standard axiomatic methodology in this highly nonstandard setting. This removes any linguistic confusion by defining concepts directly in terms of their empirical counterparts. It also allows us to pinpoint how to design experiments directed to the central tenets of the theory, rather than to particular parametrizations. If these experimental tests reveal the theory to be wanting, then knowing which axiom is violated will aid in the development of appropriate alternatives.

For our initial foray into this line of research, we focus on learning theory (Caplin and Dean 2007a). This represents an ideal test case for the integrative methodology since neuroscientists have independently formulated a specific theory of neurological function, the "Dopaminergic Reward Prediction Error" (DRPE) hypothesis, which has important behavioral implications. Dopamine is a neurotransmitter for which release had previously been hypothesized to reflect "hedonia," as when a thirsty monkey is given a squirt of juice. Yet Wolfram Schultz, Paul Apicella, and Tomas Ljungberg (1993) found that if such a monkey learns to associate a tone with later receipt of fruit juice, the dopaminergic response occurs when the tone is heard, not when the juice is received. Dopamine somehow appears to signal changes in the anticipated value of rewards. Schultz, Peter Dayan, and P. Read Montague (1997) noted that a "prediction error" signal of this form is precisely what is needed in reinforcement algorithms to drive convergence toward a standard dynamic programming value function (Andrew Barto and Richard Sutton (1982). The DRPE hypothesis of neuroscience asserts that dopamine does measure a reward prediction error that is used to update an evolving value function. In addition to more standard tests, Mathias Pesseglione et al. (2006) have shown that neurological interventions aimed at the dopamine system can have an impact on the rate at which learning appears to take place, much as the theory suggests.

Economists have become interested in reinforcement learning for their own reasons. While the natural assumption within economics is that inference is Bayesian, this assumption has little predictive power in complex environments. The need for predictive models led economists toward the reinforcement model, and Ido Erev and Alvin E. Roth (1998) have demonstrated that simple variants of this model match the pattern of behavior in a wide variety of standard games. Camerer and Teck-Hua Ho (1999) enrich the model by adding counterfactual learning based on past outcomes. It is clear, however, that

* Caplin: Department of Economics, New York University, 19 West 4th St., New York, NY 10011 (e-mail: andrew.caplin@nyu.edu); Dean: Department of Economics, New York University, 19 West 4th St., New York, NY 10011 (e-mail: mark.dean@nyu.edu). Deep thanks are extend to Paul Glimcher and to all of the economists, neuroscientists, and psychologists who contribute to the vibrancy of the New York University Neuroeconomics Seminar.

these simple "history-based" models will match behavior only in relatively stable environments. We believe that neuroscientific research has the potential to suggest next steps in modeling learning in more complex environments. Indeed, Anthony Dickinson and Bernard W. Balleine (2002) have begun to uncover evidence suggesting that dopaminergic reinforcement is but one of several neurological modules related to learning.

Unfortunately there is a profound language barrier that largely prevents economists from embracing the growing neuroscientific evidence on learning. In economics, concepts such as utility and reward are inferred from observed choices, while neuroscientists interpret them in relation to intuitions concerning the flow of experience (e.g., a squirt of juice is assumed to be rewarding to a thirsty monkey). In fact, many neuroscientific tests of the DRPE hypothesis take the perspective of "classical" or "Pavlovian" conditioning in which choice plays no role, rendering economic interpretation impossible. As with risk aversion, the fact that economic and psychological concepts have identical names does not imply identical interpretations.

Caplin and Dean (2007b) take an axiomatic perspective on the DRPE hypothesis and characterize its empirical implications for a data tape with combined information on choice and dopaminergic activity. If the data do not obey our axioms, then the DRPE model is fundamentally wrong, not merely misspecified. Our approach allows us to identify and rectify a problem in current quantitative tests of the DRPE hypothesis. In these tests, it is typical to treat neurologically measured dopaminergic signals as defined only up to linear transformations, with a quantitatively larger dopaminergic response identifying a larger reward difference. We pinpoint the somewhat harsh assumptions that are needed to justify this conclusion. Our central result justifies only an ordinal version of the DRPE hypothesis in which reward differences are ill-defined, just as marginal utility is ill-defined in ordinal characterizations. We strengthen the assumptions as required to justify use of dopamine as a measuring rod for differences in reward.

I. Basic Propositions

We develop the DRPE hypothesis for a case in which probabilities are objective and dopaminergic responses derive from realizations of specific lotteries over final prizes. An advantage of this simple lottery-outcome framework is that it avoids tying our formulation to a particular model of learning. We consider a setting in which the agent is either endowed with or chooses a specific lottery from which a prize is realized. We observe any initial act of choice among lotteries and the dopaminergic response when the prize is realized. Definition 1 lays out the various prize and lottery sets studied in the model as well as our idealized measure of the dopamine response rate.

DEFINITION 1: *The set of prizes is a compact metric space Z with generic element $z \in Z$. The set of all simple lotteries over Z is denoted Λ, with generic element $p \in \Lambda$. We define $e_z \in \Lambda$ as the degenerate lottery that assigns probability 1 to prize $z \in Z$ and the set $\Lambda(z)$ to comprise all lotteries with prize z in their support,*

$$\Lambda(z) \equiv \{p \in \Lambda | p_z > 0\} \subset \Lambda. \quad (1)$$

The function $\delta : M \rightarrow \mathbb{R}$ identifies the idealized dopamine response function (DRF), where M comprises all pairs (z, p) with $z \in Z$ and $p \in \Lambda(z)$.

The dopaminergic reward prediction error hypothesis hinges on the existence of a function defining the "expected" and the "experienced" reward associated with receipt of each possible prize from any given lottery. Under the assumption that the expected reward of a degenerate lottery is equal to the experienced reward of that prize, what we are looking for is a dopaminergic reward function $r : \Lambda \rightarrow \mathbb{R}$ which defines both the expected reward associated with each lottery and the experienced reward associated with each prize. A basic assumption is that this reward function contains all the information that determines dopamine release.

DEFINITION 2: *A function $r : \Lambda \rightarrow \mathbb{R}$ fully summarizes a DRF $\delta : \mathrm{M} \rightarrow \mathbb{R}$, if there exists a function $E : r(Z) \times r(\Lambda) \rightarrow \mathbb{R}$ such that, given $(z, p) \in M$,*

$$\delta(z,p) = E(r(e_z), r(p)), \quad (2)$$

where $r(Z)$ comprises all values $r(e_z)$ across degenerate lotteries, and $r(\Lambda)$ identifies the range across all lotteries. In this case, we say that r and E represent the DRF.

A DRPE representation rests not only on the ability to use reward computations to understand all dopaminergic responses, but also on these responses having appropriate order properties. Intuitively, the dopaminergic response should be strictly higher for a more rewarding prize than it is for a less rewarding prize, and from a less rewarding lottery than from a more rewarding lottery. In addition to depending on the existence of an appropriate reward function, the DRPE hypothesis rests on the assumption that, if expectations are met, the dopaminergic response does not depend on what was expected. We say that a DRF admits a reward prediction error representation if we can find *r* and *E* functions that satisfy all three conditions.

DEFINITION 3: *Functions r and E, which represent a DRF, respect dopaminergic dominance if E is strictly increasing in its first argument and strictly decreasing in its second argument. They satisfy no surprise constancy if $E(x,x) = E(y,y)$ for all $x, y \in r(Z)$.*

DEFINITION 4: *A DRF $\delta : \mathrm{M} \rightarrow \mathbb{R}$ admits a dopaminergic reward prediction error (DRPE) representation if there exists functions $r : \Lambda \rightarrow \mathbb{R}$ and $E : r(Z) \times r(\Lambda) \rightarrow \mathbb{R}$, which (i) represent the DRF; (ii) respect dopaminergic dominance; and (iii) satisfy no surprise constancy.*

It is clear from the definition that if $r : \Lambda \rightarrow \mathbb{R}$ forms part of a DRPE representation of a DRF $\delta : \mathrm{M} \rightarrow \mathbb{R}$, then so does any function $r^* : \Lambda \rightarrow \mathbb{R}$ that is a strictly increasing monotone transform of *r*. Hence, this representation does not allow one to treat dopamine as an invariant measure of reward differences. We develop an additive representation that embodies the minimum requirement for using dopaminergic response to animate the notion of reward differences. We also develop an expected reward representation that is entirely analogous to the expected utility representation from choice theory.

DEFINITION 5: *A DRF $\delta : \mathrm{M} \rightarrow \mathbb{R}$ admits a dopaminergic additive reward prediction error (DARPE) representation if there exists a function $\mathrm{r} : \Lambda \rightarrow \mathbb{R}$ and a strictly increasing function $G : r(Z) - r(\Lambda) \rightarrow \mathbb{R}$, such that, given $(z, p) \in M$,*

$$\delta(z,p) = G(r(e_z) - r(p)), \tag{3}$$

where $r(Z) - r(\Lambda)$ comprises all values $r(e_z) - r(p)$ across $(z,p) \in M$.

DEFINITION 6: *A DRF $\delta : M \rightarrow R$ admits a dopaminergic expected reward prediction error (DERPE) representation if there exists functions $r : \Lambda \rightarrow \mathbb{R}$, $E : r(Z) \times r(\Lambda) \rightarrow \mathbb{R}$ that form a DRPE representation of the DRF in which*

$$r(p) \equiv \mu_p[u] \text{ all } p \in \Lambda,$$

for some function $u : Z \rightarrow \mathbb{R}$, where $\mu_p[u]$ denotes the expected value of $u : Z \rightarrow \mathbb{R}$ with respect to the lottery p.

For economic interest to be warranted, there must be a connection between the dopaminergic reward and choice. The simplest such connection occurs if choices among lotteries can be modeled as deriving from maximization of the DRPE reward function. While this case is of obvious interest to economists, a more standard scenario involves dopamine as simply one component of a richer overall process of learning and of choice.

DEFINITION 7: *The choice correspondence C is defined on domain $Q = \{X \in 2^{\Lambda}$ with $|X| < \infty\}$, with $C(X) \subseteq X$ denoting the set of lotteries chosen from any finite subset of the lottery space. A DRF $\delta : M \rightarrow R$ and a choice correspondence C admit a choice-consistent DRPE representation if δ admits a DRPE representation, and for all $X \in Q$,*

$$C(X) = \arg\max_{p \in X} r(p). \tag{4}$$

Caplin and Dean (2007a) characterize all the representations above. In the present paper, we outline conditions for the basic DRPE representation. There are three axioms that are intuitively necessary for $\delta : M \rightarrow R$ to admit a DRPE. The first ensures coherence in dopaminergic responses to prizes: if one prize is more of a positive surprise than another when received from some lottery, it must be so for any other lottery. The second provides the analog condition with respect to lotteries: if an outcome *z* leads to a bigger dopamine release when obtained from one given lottery than from some other lottery, the same must be true for any other outcome that is in the support of both lotteries. The third characterizes the dopamine function as having equivalent value all along the 45-degree line.

AXIOM 1 *(A1: Coherent Prize Dominance): Given* $(z,p),(z',p'),(z',p),(z,p') \in M$,

$$\delta(z,p) > \delta(z',p) \Rightarrow \delta(z,p') > \delta(z',p'), \tag{5}$$

AXIOM 2 *(A2: Coherent Lottery Dominance): Given* $(z,p),(z',p'),(z',p),(z,p') \in M$,

$$\delta(z,p) > \delta(z,p') \Rightarrow \delta(z',p) > \delta(z',p'). \tag{6}$$

AXIOM 3 *(A3: No Surprise Equivalence): Given* $z, z' \in Z$,

$$\delta(z',e_{z'}) = \delta(z,e_z). \tag{7}$$

While necessary, A1–A3 are not sufficient for a DRPE representation, due to the fact that the domain of the dopamine function differs across prizes. These domain differences allow A1–A3 to be consistent with the cycles of apparent dopaminergic dominance, and other conditions inconsistent with the DRPE (Caplin and Dean (2007a). In that paper, we show that the following continuity conditions suffice to establish availability of such a representation.

AXIOM 4 *(A4: Uniform Continuity): The function* $\delta : M \rightarrow \mathbb{R}$ *is uniformly continuous in the appropriate metric.*

AXIOM 5 *(A5: Separation): Given* (z,p), $(z',p) \in M$,

$$\delta(z,p) \neq \delta(z',p) \Rightarrow \inf_{\{p' \in \Lambda | (z,p'),(z',p') \in M\}} |\delta(z,p') - \delta(z',p')| > 0. \tag{8}$$

THEOREM 1: *Under A4 and A5, a DRF* $\delta : M \rightarrow \mathbb{R}$ *admits a DRPE representation if and only if it satisfies A1–A3.*

II. Next Steps

Caplin and Dean (2007a) provide a theoretical framework within which to test the DRPE hypothesis. Our next step is to generate data with which to perform such a test. To this end, we are working with Glimcher at the Center for Neural Science at New York University on the laboratory implementation of our axiomatic framework. We are currently refining an experimental design that will allow us to record brain activity as lotteries are chosen and as they resolve. This takes place while the subject is lying in a functional Magnetic Resonance Imaging (fMRI) scanner. We use these data to construct an empirical analogue of the δ function.

Assuming that we find supporting evidence for the DRPE hypothesis, our next step will be to extend our theoretical and empirical work to allow for dynamic environments in which learning takes place. Richer axioms will be developed to connect the observed patterns of choice and dopaminergic response, enabling us fully to characterize reinforcement learning models for such data.

The drive to deepen understanding of the neurological and behavioral aspects of learning will ultimately force researchers to look beyond the dopamine system. Dickinson and Balleine (2002) have developed a two-process theory of the motivation, with only one of these two processes involving the dopamine system. In association with this theory, Balleine (2005) is currently investigating the development of "habits" as opposed to more flexible responses, and an associated "supervisory function" that determines the extent to which habitual behavior is called into play in any given situation. Depending on specific neurological manipulations, there can be wholesale changes in the extent to which past rewards dominate future behavior. In a multimodal framework, our understanding of behavior will be greatly enhanced if we can develop neurological evidence pinpointing the extent to which different modes are in operation.

We believe that the axiomatic approach to neuroeconomics may be of value in areas other than learning. Kenway Louie and Glimcher (2007) present intriguing preliminary evidence of a neurological switch that occurs in the process of decision making, and that may possibly signal the process of contemplation has come to an end. This suggests the value of studying the neurological and behavioral aspects of the process of "arriving at a decision." The hope is that neuroscientific measurements will provide systematic, if limited, access to the search process that presages the act of choice. This would provide us with new channels for understanding when and how information provided to a decision maker is mapped into the ultimate decision.

The broader backdrop to our research is our belief that the momentum to broaden the empirical basis of decision theory is unstoppable. We see the appropriate response as being to embrace rather than to resist this domain expansion. While Gul and Pesendorfer are correct to point out that language barriers impede interdisciplinary collaboration, these barriers are far from impenetrable. Our goal is not only to penetrate these barriers, but also to show that there is much to be learned through incorporation of nonstandard data into economic theory. Caplin and Dean (2007b) provide an entirely separate example focused around the question of how long a subject takes to make a decision. Caplin (forthcoming) argues against imposing any ex ante constraints on the types of data that can potentially be included in axiomatic models, even if the only goal is to provide insight into standard choice behavior. Our current argument that the neuroeconomic language barrier can be breached in the case of a specific neurological measurement represents only the tip of the methodological iceberg.

REFERENCES

Balleine, Bernard W. 2005. "Neural Bases of Food-Seeking: Affect, Arousal, and Reward in Corticostriatalimbic Circuits." *Physiology and Behavior*, 86(5): 717–30.

Balleine, Bernard W., A. Simon Killcross, and Anthony Dickinson. 2003. "The Effect of Lesions of the Basolateral Amygdala on Instrumental Conditioning." *Journal of Neuroscience*, 23(2): 666–75.

Barto, Andrew, and Richard Sutton. 1982. "Simulation of Anticipatory Responses in Classical Conditioning by a Neuron-Like Adaptive Element." *Behavioral Brain Research*, 4(3): 221–35.

Camerer, Colin, and Teck-Hua Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica*, 67(4): 827–74.

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

Caplin, Andrew. Forthcoming. "Economic Theory and Psychological Data: Bridging the Divide." In *Perspectives on the Future of Economics: Positive and Normative Foundations*, Vol. I, ed. Andrew Caplin and Andrew Schotter. Oxford: Oxford University Press.

Caplin, Andrew, and Mark Dean. 2007a. "Dopamine and Reward Prediction Error: An Axiomatic Treatment." Unpublished.

Caplin, Andrew, and Mark Dean. 2007b. "Revealed Preference and Time to Decide." Unpublished.

Dickinson, Anthony, and Bernard W. Balleine. 2002. "The Role of Learning in the Operation of Motivational Systems." In *Steven's Handbook of Experimental Psychology, Volume 3: Learning Motivation and Emotion,* ed. Randy Gallistel and Hal Pashler. New York: John Wiley & Sons.

Erev, Ido, and Alvin E. Roth. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review*, 88(4): 848–81.

Glimcher, Paul W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics.* Cambridge, MA and London: MIT Press.

Gul, Frank, and Wolfgang Pesendorfer. Forth coming. "The Case for Mindless Economics." In *Perspectives on the Future of Economics: Positive and Normative Foundations.* Vol. I, ed. Andrew Caplin and Andrew Schotter. Oxford: Oxford University Press.

Louie, Kenway, and Paul W. Glimcher. 2007. "Temporal Discounting Activity in Monkey Parietal Neurons during Intertemporal Choice." Unpublished

Pessiglione, Mathias, Ben Seymour, Guillaume Flandin, Raymond J. Dolan, and Chris D. Firth. 2006. "Dopamine-Dependent Prediction Errors Underpin Reward-Seeking Behaviour in Humans." *Nature,* 442(7106): 1042–5

Schultz, Wolfram, Paul Apicella, and Tomas Ljungberg. 1993. "Responses of Monkey Dopamine Neurons to Reward and Conditioned Stimuli During Successive Steps of Learning a Delayed Response Task." *Journal of Neuroscience*, 13(3): 914–23.

Schultz, Wolfram, Peter Dayan, and P. Read Montague. 1997. "A Neural Substrate of Prediction and Reward." *Science*, 275(5306): 1596–99.

This article has been cited by:

1. Don Ross. 2008. TWO STYLES OF NEUROECONOMICS. *Economics and Philosophy* **24**:03, 473. [CrossRef]