American Economic Journal: Microeconomics 2009, 1:2, 42–47
http://www.aeaweb.org/articles.php?doi=10.1257/mic.1.2.42

A Comment on Bernheim's Appraisal of Neuroeconomics†

By Faruk Gul and Wolfgang Pesendorfer*

This paper comments on "On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal" by B. Douglas Beinheim (JEL D01, D87)

Many academic economists are skeptical of philosophical arguments and methodological discussions. They perceive the success of their discipline to be rooted in hard-nosed empiricism and quantitative methods, and are loath to engage in vague, nonquantitative discussions considered the bane of humanities and philosophy. Phrases such as "the proof of the pudding is in the eating," or "the success of neuroeconomics will depend on whether or not it can ultimately deliver the goods" are used to brush aside methodological questions.

B. Douglas Bernheim's (2009) discussion of the current and potential contributions of neuroeconomics to positive economics provides an ingenious rhetorical innovation; one that accommodates hard-nosed economists and might render methodological analysis acceptable, if not agreeable—even to those who normally have little enthusiasm for such discussion. Bernheim replaces the vocabulary of philosophy and methodology with that of econometrics. He defines the prototypical task of positive economics as predicting some variable(s) y. Positive economics then postulates the form of the equation $y = f(x, w)$, measures the endogenous and exogenous variables (x and y), and eliminates the unobservable exogenous variables (w) from the equation by modeling them as random. An economic theory becomes a collection of conditional probability measures $\eta(\cdot \mid A)$, one for each x, where $\eta(A \mid x)$ is the probability that y will take on a value in the set A given observables x.

This formulation enables Bernheim to categorize and discuss the possible contributions of neuroeconomics. Can neuroeconomics render some unobservable exogenous variables observable? Can neuroeconomics remedy some of the econometric problems associated with estimating η, such as the omitted variables bias, endogeneity, and other common econometric problems that make it difficult to predict the impact of policy from an existing econometric model?

The approach serves Bernheim well because it enables him to be specific about where he expects significant contributions from neuroeconomics and where he does

* Gul: Department of Economics, Princeton University, Princeton, NJ 08540 (e-mail: fgul@princeton.edu); Pesendorfer: Department of Economics, Princeton University, Princeton, NJ 08540 (e-mail: pesendor@princeton.edu).

† To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at: http://www.aeaweb.org/articles.php?doi=10.1257/mic.1.2.42.

not. It also lets him identify and explain the sources of disagreement between advocates and critics of neuroeconomics. Regardless of their familiarity with this methodological debate, all readers are likely to benefit from Bernheim's willingness to engage the neuroeconomic research on its own terms. Moreover, his focus on the details of research, rather than concerns relating to philosophy and economics, lets him replace the sonorous tone of many methodology papers with the more agreeable tone of a survey article, a tone that many hard-nosed economists are likely to appreciate.

Bernheim considers examples of disagreement between enthusiasts and critics of neuroeconomics. For example, he discusses the observation by Samuel M. McClure et al. (2004) "that decisions activate distinct regions of the brain to differing degrees depending on whether they involve immediately available or delayed rewards." Bernheim notes,

> "The paper is sometimes interpreted as providing a neural test of the popular $\beta-\delta$ model of quasihyperbolic discounting. That interpretation is inappropriate. Even assuming that the observed neural activity encodes subjective valuations and that those valuations govern decision making (hypotheses that are not addressed by this particular experiment), the evidence does not rule out the possibility that valuations are time-consistent, or that any inconsistencies are harmonized by other structures" (Bernheim 2009, 20).

Hence, like Gul and Pesendorfer (2008), Bernheim notes that economics makes no predictions about the brain, and hence neuroeconomic evidence cannot falsify economic theories. Moreover, his well-chosen examples illustrate the gap between assertions about the brain (different parts of the brain encode utilities for immediate versus delayed rewards), and the intuitively related assertions about the corresponding economic behavior (nonexponential discounting in dynamic choice).

In his discussion of positive neuroeconomics, Bernheim makes at least two other important observations. First, he criticizes an analogy in a paper by Colin F. Camerer, George Loewenstein, and Drazen Prelec (2004), who suggest that neuroeconomics will benefit consumer theory by unlocking its black box (the brain) the same way that producer theory benefited from the theory of the firm which was previously black-boxed as a production function. Bernheim notes that the analogy is misleading since: "In developing the theory of the firm, economists were not motivated by the desire to improve the measurement of reduced form production functions relating output to labor and capital. Rather, questions pertaining to the internal workings of the firm (unlike those pertaining to the internal workings of the mind) fall squarely within the historical boundaries of mainstream economics because they concern organized exchange between individuals. (...) In contrast, the mind is not an economic institution, and exchange between individuals does not take place within it" (Bernheim 2009, 5–6).

Finally, Bernheim offers the somewhat farfetched but revealing hypothetical of a graduate student who observes that he can "predict milk purchases accurately with a single variable, whether the customer reaches out to grab a carton of milk." He concludes that:

> "The historical objective of positive economists is to improve the prediction of choice ($\mathbf{y}$) from standard exogenous variables ($\mathbf{x}$), such as prices,

> taxes, income, gender, age, and so forth. The observation that one can more accurately predict choice from endogenous neural variables (**z**) simply does not speak to that objective" (Bernheim 2009, 13).

The hypothetical makes two distinct points. First, endogenous neural variables are often too close to the behavior that is being predicted. Hence, any observer with access to such x's would also observe the corresponding y's. Therefore, these x's would not be useful for forecasting the y's. The second point is that whether or not these variables are useful for forecasting, often economists are interested in the interaction of the underlying economic variables and not the intervening neural variables.

Bernheim's analysis illustrates both the strengths and weaknesses of his rhetorical device. Identifying all positive economics with an econometric forecasting exercise enables him to discuss specific research contributions in detail and on their own terms. But, on occasion, understanding his arguments requires the reader to keep in mind that there is more to economics than forecasting. Concepts such as sunk cost, fixed cost, externality, private information, and reservation utility are abstractions and not ready-made variables awaiting measurement. The success of an economic research project often depends on the researcher's ability to formalize and quantify a particular economic abstraction within a particular application. Developing a sensible framework within which neuroeconomic variables and standard economic variables interact is perhaps the most daunting challenge for positive neuroeconomics. Up to this point the bulk of neuroeconomic research falls into one of two categories: (i) monitoring brain activity while the individual is making a particular economic decision and (ii) interpreting and analyzing fairly standard economic models with insights gained from recent developments in the type of research described in (i). Undoubtedly, such research can be illuminating but is a far cry from the integrated neural and economic models that some have in mind when contemplating the future of neuroeconomics. The farther we move away from reduced-form models that resemble a fairly nontheoretical forecasting exercise, the more difficult the integration becomes.

Bernheim's rhetorical device sidesteps the methodological difficulties associated with this integration at the cost of making his overall negative assessment difficult to understand. Bernheim writes that he is not convinced "that [neuroeconomics] is likely to become a central or indispensable component of standard positive economics, or that it will revolutionize the field in some fundamental way" (Bernheim 2009, 29). Readers who do not understand the need for a conceptual framework are likely to respond to Bernheim with questions such as "are there uses for exogenous neuroeconomic variables?" with "why not just throw them into the regression and find out?" Bernheim's reply is that, so far, the opportunities to throw them in and get something useful have come up fairly infrequently. But we suspect that his negative verdict on positive neuroeconomics also relies on an unstated observation. For mainstream economics, identifying the brain regions involved in economic decisions, or including novel noneconomic (for example, neural) variables in a regression, is of relatively minor importance. Therefore, even if there are more uses for such variables in the future, this will not lead to the revolution in economics that many neuroeconomists expect.

After concluding his discussion of positive neuroeconomics, Bernheim offers the following formulation of the prototypical task of normative economics. "Suppose a planner (P) must choose between two actions, A and B, the consequences of which affect only one individual (I)...." (Bernheim 2009, 35) what criteria should the planner use for making his choice? He considers three possible candidates for a welfare criterion that the planner might use to decide what is best for I. The first is what he calls "neural utilitarianism," (Bernheim 2009, 29) (i.e., P should decide what to choose by studying I's brain). The second is what he calls "happiness-based welfare analysis," (Bernheim 2009, 31) (P should ask I whether A or B would make him happier or has made him happier in the past and choose accordingly). The third is "choice-based welfare analysis," (P should let I make the choice or choose what I would have chosen if he could make the choice). Hence, Bernheim defines normative economics as an exercise in grounding policy analysis in moral philosophy.

Bernheim dismisses the first two options in favor of a modified version of the third, *informed* choice-based welfare analysis. To illustrate this option, Bernheim supposes I:

> "chooses A when the alternatives are described verbally, and B when they are described partly verbally and partly in writing. Which choice is the best guide for public policy? If we learn that the information was provided in a dark room, we would be inclined to respect the choice of A, rather than the choice of B. We would reach the same conclusion if an ophthalmologist certified that the individual was blind, or, more interestingly, if a brain scan revealed that the individual's visual processing circuitry was impaired" (Bernheim 2009, 35–36).

Hence, Bernheim sees an important but limited role for neuroeconomics in normative analysis. It can determine when I is informed (or impaired) and when she is uninformed, and thus determine which of her choices should be identified with her welfare and which should not.

Two features of Bernheim's moral philosophical argument and his preferred welfare criterion stand out. The first is that it is an expression of subjective preference or Bernheim's own stance rather than an argument. At critical points of his discussion, Bernheim candidly states what he, as the planner, would do and makes no attempt to argue that this is what a social planner should do: "Personally, I would be unwilling to overrule the individual's choice and declare him better off with B. On the contrary, I would be inclined to assume that the neural measure overlooked some consideration that was important to the individual" (Bernheim 2009, 33).

Thus, Bernheim assumes that the reader shares his presumption that a choice-based welfare criterion is appropriate for the planner and offers to place the word "informed/unimpaired" in front of choice as a friendly amendment. The second interesting feature of Bernheim's discussion is that this friendly amendment turns out to be a Trojan horse. For it soon becomes apparent that Bernheim's leading example, that of a decision maker choosing in the dark, is not really an example but a grand metaphor that can accommodate every situation. Virtually any choice that the individual makes can be deemed uninformed and disregarded, and any choice that Bernheim considers appropriate can be defined as *the* informed decision in the *same* situation. Consider the following example, "a recovering alcoholic drinks whenever

he socializes with drinkers, but at other times would happily impose upon himself a binding commitment not to drink in such situations. Because those choices pertain to precisely the same actions and circumstances, there is plainly a conflict. We resolve that conflict in favor of the precommitment, on the grounds that a decision to drink taken in the presence of a cue (social interaction with drinkers), associated with the consumption of an addictive substance (alcohol), is influenced by a neural forecast that is most likely distorted due to the substance's neurobiological properties" (Bernheim 2009, 36).

Getting rid of all alcohol in the house is a different choice than drinking club soda on a night out with friends. The difference is clear from the fact that many people find the first much easier to do than the second. Bernheim acknowledges this difference only as a factor for processing information and not as one that affects preferences. To put it differently, when he disagrees with a choice, he identifies the accompanying factors as "cues;" that is, as impediments to decision making (i.e., incorrect information, broadly defined), and ignores the utility interactions between these factors and the consumption choices. Hence, interpreting certain environmental factors as cues (i.e., as obstacles to making the correct decision) versus complements (as factors that influence which decision is correct) is Bernheim's device for summarily passing judgment on difficult moral philosophical issues without acknowledging them as such.

Even if we grant that actions at the commitment stage and actions at the consumption stage represent the same choice, designating the former as informed and the latter as uninformed seems arbitrary and unjustified. Bernheim designates the decision to drink as uninformed because of studies concluding that a particular mechanism in the brain "overestimates" the amount of hedonic reward that alcohol will provide. This conclusion is susceptible to the same criticisms that Bernheim levies at those who interpret the McClure et al. (2004) study as evidence of the $\beta-\delta$ model. Even assuming that the observed neural activity can be interpreted as information processing, and that this information governs decision making, the evidence does not rule out the possibility that better information about the hedonic consequences of alcohol is processed by other structures and harmonized (or aggregated) with the output of the hedonic forecasting mechanism. Conversely, given the standards that Bernheim has adopted, it is easy to come up with some psychological or neural mechanism to support the conclusion that the commitment choice is also "uninformed," "made in the dark," or suffers from biased hedonic forecasting.[1] This flexibility allows the researcher to designate virtually any preconceived notion of the right choice as informed.

Bernheim predicts that the biggest future contribution of neuroeconomics will likely be in normative economics; that is, in providing guidance to the planner on how to make choices that only affect the well-being of others. In Gul and Pesendorfer (2007), we offer an alternative interpretation of work in welfare economics, one that does not define welfare/normative economics as a subfield of moral philosophy. Here, we add that nothing in the recent history of economics, or the career of the

[1] For an extensive list of systematic biases in predicting ones own tastes (i.e., hedonic utilities), see Loewenstein and David A. Schkade (1999).

typical research economist, suggests that the notion of welfare/normative economics as grounding policy assessment in moral philosophy plays an important role in the discipline. Whether or not the reader agrees with our position, she is unlikely to be satisfied with Bernheim's neuroscience augmented welfare analysis, an analysis that offers an ethical stance without an ethical argument to arrive at a criterion that places few, if any, constraints on what is deemed to be the right choice.

REFERENCES

Bernheim, B. Douglas. 2009. "On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal." *American Economic Journal: Microeconomics*, 1(2): 1–41.

Camerer, Colin F., George Loewenstein, and Drazen Prelec. 2004. "Neuroeconomics: Why Economics Needs Brains." *Scandinavian Journal of Economics*, 106(3): 555–79.

Gul, Faruk, and Wolfgang Pesendorfer. 2007. "Welfare without Happiness." *American Economic Review*, 97(2): 471–76.

Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." In *The Foundations of Positive and Normative Economics: A Handbook*, ed. Andrew Caplin and Andrew Schotter, 3–42. Oxford: Oxford University Press.

Loewenstein, George, and David A. Schkade. 1999. "Wouldn't it be Nice? Predicting Future Feelings." In *Well-Being: The Foundations of Hedonic Psychology*, ed. Daniel Kahneman, Edward Diener, and Norbert Schwarz, 85–105. New York: Russell Sage Foundation Press.

McClure, Samuel M., David I. Laibson, George Loewenstein, and Jonathan D. Cohen. 2004. "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science*, 306(5695): 503–07.