American Economic Journal: Microeconomics 2009, 1:2, 60–67
http://www.aeaweb.org/articles.php?doi=10.1257/mic.1.2.60

Neuroeconomics: A Comment on Bernheim†

By JOEL SOBEL*

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

> I was thrown out of NYU my freshman year. I cheated on my metaphysics exam. I looked into the soul of the boy sitting next to me.[1]

If Woody Allen were in college today, perhaps he would be working in a neuroimaging lab, looking into brains for answers to questions that might appear on an economics examination. B. Douglas Bernheim's (2009) paper on neuroeconomics suggests while Allen should not be thrown out of the university for looking into brains, it is not yet clear that the answers he obtains would be sufficient to earn high grades.

Stated more directly, Bernheim views the potential contributions of neuroeconomics to economics with a mixture of skepticism and hope. I share these sentiments, although my skepticism and hope are distributed a bit differently.

This comment reconsiders and responds to some of Bernheim's arguments.

I. Appropriate Standards

> The persuasion that one's own infallibility is a myth leads by easy stages and with somewhat greater satisfaction to a refusal to ascribe infallibility to others.[2]

Bernheim describes several valid reasons why techniques from neuroeconomics will not be broadly accepted and shows us why we should be cautious about accepting the claims of researchers in the field. These observations are important methodological criticisms. Bernheim's comments are valuable because they point out weaknesses of individual studies. They also identify criteria that we should apply in evaluating research in mature areas. They do not identify an intrinsic weakness in neuroeconomics, however. They are reason to be skeptical about claims made in specific research papers without rejecting the potential value of neuroeconomics.

* Department of Economics, University of California, San Diego, La Jolla, CA 92093 (e-mail: jsobel@ucsd.edu). I thank Nageeb Ali, Julie Cullen, Steven Durlauf, Roger Gordon, Clive Granger, Antonio Rangel, and a referee for helpful comments and NSF (SBR-9514809) for financial support. I chose to write this note, but doing so did not make me happy.

† 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.1.2.60.

[1] Woody Allen, quoted in Eric Lax (2000, 133).

[2] Benjamin N. Cardozo (1949, 30).

In Section IE, Bernheim (2009) does the important work of identifying critical assumptions made by neuroeconomists. He carefully reviews the results of two recent neuroeconomic studies and effectively identifies a gap between what neuroecomic studies claim and what they deliver. Economists have much less experience identifying, interpreting, and evaluating the assumptions underlying neuroeconomic research than the assumptions we use to draw conclusions from empirical studies. For this reason, it is natural to be skeptical of conclusions from neuroeconomic studies. Neuroeconomists should make these assumptions explicit rather than waiting for their critics to do so.

To do any empirical work, we must impose assumptions on the data to reach conclusions. Often these assumptions are implicit. When evaluating a natural experiment, the researcher must assume that the experiment is truly natural. When evaluating a laboratory experiment, the researcher must assume that subjects understand the instructions.[3] There will be controversy over the interpretation of assumptions. Controversial and potentially important findings will generate more research. This is true in mainstream economics and will remain true for neuroeconomics.

Take the work of Samuel M. McClure et al. (2004) that Bernheim discusses on page 20. McClure and co-authors assert "that decisions activate distinct regions of the brain to differing degrees depending on whether they involve immediately available or delayed rewards" and interpret this finding as evidence in favor of hyperbolic discounting. Bernheim correctly states that the paper delivers less than this. In Bernheim's words, the paper shows that "*under the maintained hypothesis that choices conform to the $\beta-\delta$ model*," the authors "test the supplemental hypothesis that such choices are generated by a dual-system process," (Bernheim 2009, 21) but the real issue is whether the existence of different methods for processing immediate and delayed rewards is evidence that supports the hypothesis of hyperbolic discounting. From my naïve perspective, the dual-system seems like a plausible necessary condition for the brain to be implementing a $\beta-\delta$ model, but quite far from a sufficient condition.[4] Without further work, I find the research provocative without being convincing. I do not expect further research to provide conclusive evidence in support of the $\beta-\delta$ model, but it may raise my level of comfort with the assumptions underlying the assertion.

While it is important to understand and separate the assumptions underlying the claims in McClure et al. (2004) and other neuroeconomic studies, it is equally important to remember that other areas draw conclusions from studies that rely on implicit assumptions. An empirical example is the literature on natural experiments designed, for example, to estimate the effect that education has on earnings. This research starts from a simple model that assumes that

$$y = \alpha + \beta \mathbf{X} + \varepsilon, \tag{1}$$

[3] It is not possible to guarantee that the structure of a laboratory game is common knowledge. Surely subjects believe that it is possible that other subjects did not understand the rules of the game, and so a necessary condition for common knowledge will fail in practice.

[4] Neuroeconomic research by Joseph W. Kable and Paul W. Glimcher (2007) and Sabrina M. Tom et al. (2007) casts doubt on the hypothesis that dual valuation systems are necessary for hyperbolic discounting.

where y is (log) earnings, and $\mathbf{X}$ is a vector of independent variables that may include education or investment after school. The empirical goal is to estimate the coefficients β. The standard problem is that the error term ε may be correlated with observed values of $\mathbf{X}$. The conventional estimation technique looks for instrumental variables Z that covary with $\mathbf{X}$ but are independent of ε. In natural experiments, researchers select these variables and attempt to justify the independence assumption. These identification assumptions are necessary to organize the data but are typically controversial. A seductive way to avoid this problem is to examine so-called "natural" natural experiments in which the instrumental variable is a naturally occurring random event (like the date of birth) rather than a policy change.[5] Even for these studies, the estimation strategy requires a detailed model of the error structure, something that is often unstated in empirical work. Mark R. Rosenzweig and Kenneth I. Wolpin (2000) review this literature, point out reasons for apparently incompatible estimates, and describe a constructive strategy for providing and assessing behavioral models that justify underlying models. In this work, as in the research in neuroeconomics, the issue is not whether conclusions require auxiliary assumptions. This is inevitable. It is whether one can identify and test the validity of the additional assumptions. Mature subfields may approach a consensus about which assumptions need to be discussed and which can be left unstated. Neuroeconomics has not, but I see no reason why it cannot.

As another example, consider a recent paper (James Andreoni and Bernheim forthcoming), in which Bernheim and his co-author present a new theoretical model that explores fairness norms in dictator games and supplements the theory with experimental evidence. The paper reports experiments in which a dictator selects how to divide a surplus. With positive probability, however, nature imposes the split of the surplus. The audience knows the probability that nature will intervene and the split it will impose, but observes only the split itself. When nature will give most of the surplus to the dictator, increasing the probability that nature will impose a split increases the probability that the dictator will imitate the unfair split.[6] These experimental results are an excellent example of research progress in an established field. Earlier experiments tested the joint hypothesis that dictators sought to maximize their monetary payoff and did not care about audience effects (being seen to violate fairness norms). A richer experimental design helped to separate the two hypotheses. At the same time, Andreoni and Bernheim's paper organizes its experimental finding by relating it to one of many equilibria in its theoretical model. This selection (which I find reasonable) leads to a theory that makes a prediction that is consistent with the experimental results. The connection between theory and experiments requires a secondary hypothesis (on equilibrium selection) that goes beyond those normally made in equilibrium analysis and may be difficult to test independently.

[5] For example, in response to different age requirements for starting or leaving school, there is variation in the length of school attendance that varies with date of birth. One can obtain estimates of the returns to schooling based on these differences, assuming that ability does not vary with the date of birth.

[6] For related studies, see Tomas Broberg, Tore Ellingsen, and Magnus Johannesson (2007); Jason Dana, Daylian M. Cain, and Robyn M. Dawes (2006); Edward Lazear, Ulrike Malmendier, and Roberto Weber (2006); and Steven Tadelis (2007).

Some of the difficulties Bernheim has with neural welfare measures are really general criticisms of welfare measures. He worries (Bernheim 2009, 31) that we may not be able to interpret neurological evidence even if the brain codes for an overall sense of well-being at each moment in time. Specifically, he asks what we should do if u_t^i, the coded level of well-being for activity i at time t, satisfies $u_0^A > u_0^B$ and $u_1^B > u_1^A$. He could just as well ask: What would an economist do if she found that an agent selected meat over fish at time zero and fish over meat at time one? The economist would come up with many potential reasons for the apparent switch: the choice situations were not identical (different prices, different portion size, different quality), the agent has a taste for diversity, the choice depends on other factors (the choices of dining companions), or—presumably the last resort—preferences have changed. Revealed preference analysis cannot, in practice, distinguish whether a decision is made only to satisfy instantaneous preferences or extended interests are relevant. More generally, the appropriate domain of preferences is an important issue for conventional revealed-preference analysis.[7]

Bernheim argues that neuroeconomics is likely to be a useful way to reconcile inconsistent choice behavior. He gives examples from his own research program to illustrate this notion on page 34 (Bernheim 2009). This work is promising and puts physiological data to good use, but depends on assumptions on the extent to which cues influence behavior that may be hard to test and evaluate independently.

Some of Bernheim's comments suggest that the relevance of neuroeconomics may be limited without casting doubt on the general approach. These observations also apply more broadly. I agree with Bernheim that it is unlikely that mainstream economists will need to master techniques of neural science to do their research, but this observation hardly limits the potential impact neural evidence may have on the field. An economist who incorporates data obtained by neuroscientists into a study without "extensive knowledge of neuroeconomic methods or a deep appreciation of neural processes" (Bernheim 2009, 8) is no different from an applied economist who uses statistical tests or a theorist who uses a fixed-point theorem without understanding the proofs of these results.

Bernheim is right to point out that neureconomics may add to the understanding of choice processes initiated by researchers not associated with neuroeconomics. Mouselab experiments (see Vincent P. Crawford 2008 for a survey) and reaction time studies (Ariel Rubinstein 2006) gather data about the way in which subjects make decisions. These techniques are less invasive and less expensive than neural imaging, but have the common objective of providing more information about decision-making processes. There is good reason to view these developments as part of a trend where economists look for different ways to gain information about choice behavior. The theoretical motivation for this work is dissatisfaction with the descriptive power of classical equilibrium notions. The practical motivation is that the profession is taking notice of the fact that the way decisions are framed matters.

[7] Faruk Gul and Wolfgang Pesendorfer, visible critics of neuroeconomics, have advanced an important literature on self control by describing the axiomatic foundations of preferences over menus (for example, see Gul and Pesendorfer 2001). An observer who failed to take into account seemingly irrelevant alternatives would view an agent who obeys Gul and Pesendorfer's assumptions as inconsistent. These inconsistencies disappear once one allows a set of unchosen options to influence choice.

Understanding how changes in the choice environment influence decisions has positive and normative implications. In view of the growing interest in studying decision-making processes using techniques that do not directly monitor brain activity, skepticism about research into processes must extend beyond neuroeconomics.

II. Normative Economics

> In mathematics, names are free. It is perfectly allowable to call a self-adjoint operator an elephant, and a spectral resolution a trunk. One can then prove a theorem, whereby all elephants have trunks. What is not allowable is to pretend that this result has anything to do with certain large gray animals.[8]

The potential contribution of neural research to the development or evaluation of alternate welfare measures is, perhaps, the most intriguing application of the new techniques. Although contemporary economic theory focuses—with good reason—on observable choices, classical welfare economists expressed an interest in measuring happiness. Bernheim doubts that neuroeconomics will provide a foundation for alternative welfare measurements. He argues, however, that neuroeconomics is likely to be a useful way to obtain additional data when traditional methods are unavailable.

Any discussion of normative economics must fix a standard to be used to evaluate outcomes. Economists use Pareto efficiency, which respects individual preferences and provides a weak ordering of outcomes. We further assume that choice behavior reveals preferences.

There are three good reasons why economic notions of efficiency may be difficult to apply in practice. Insufficient data is one of them. Contemporary psychology and behavioral economics use surveys and low-stakes experiments to collect data. Supplementary information will be useful, provided that we know how to interpret it. Bernheim's research program illustrates a useful way in which neuroeconomics can become a reliable source of data. "Asking the brain" (as Colin F. Camerer, George Loewenstein, and Drazen Prelec 2004 describe it) may be most useful when we are confident that the brain's answer will agree with choice data. In this way, neuroeconomic data supplements traditional empirical evidence. When activation patterns do not track choices, however, we are left without a consistent welfare measure making it difficult to know how to reconcile conflicting data sources.

Another reason we cannot base all welfare arguments on choice behavior is that sometimes there is a consensus that favors paternalism. We do not honor the revealed preferences of the very young or of those judged incompetent to make their own decisions. On the other hand, the boundaries of these categories vary across cultures (and have varied over time). Traditional psychological evidence has a role in determining when society should ignore individual preferences. Neural evidence is likely

[8] Hector Sussman, quoted in Ivar Ekeland (Ekeland 1990, 102). Thirty years ago several distinguished researchers claimed that a mathematical theory that characterized singularities of mappings could have far-reaching implications for social science. Sussman, a mathematician, expressed skepticism about the applications of catastrophe theory and cautioned us to separate the theory from interpretations of the theory.

to be relevant in this case also, perhaps, by identifying situations in which decision makers are likely to make mistakes (see, for example, Bernheim and Antonio Rangel 2009).

Third, behavioral evidence in favor of other-regarding preferences, time consistency, and irrationality suggest circumstances in which it is either hard to define or hard to defend revealed-preference arguments. Richard H. Thaler and Cass R. Sunstein (2008) argue that awareness of framing effects suggests ways to influence outcomes without limiting choices. Inevitably, advocates of particular policies will reach for any evidence that supports their point of view. Neuroeconomic studies of happiness will be invoked (and possibly commissioned) with policy recommendations in mind. Academic economists should pay serious attention to the field if only to provide unbiased and intellectually grounded ways to evaluate the claims of neuroeconomic studies.

While there are practical reasons to encourage attempts by neuroeconomists to connect brain activation to welfare measures, there are also conceptual reasons. The tools of economics do not require that levels of consumption determine choice. There is intuitive evidence, supported by physiological studies that rates of change and reference points matter. Theories of choice that are sensitive to these considerations were constructed without motivation from neuroeconomic studies, but physiological data have a potentially useful role in suggesting additional models. Imaging studies can certainly identify whether reference points influences activation patterns. Evidence of this kind could provide guidance about what is an appropriate domain of preferences.

Finally, while I share Bernheim's skepticism of studies designed to measure happiness (or what would be more accurately described as reported happiness), I may be more willing to accept that it is possible to obtain consistent and meaningful measurements of some intrinsic indicator of well-being. I am conscious of biometric signals of pain, anxiety, and arousal. If someone could convincingly connect these sensations to brain activation, then I would find it plausible to associate heightened activity with welfare.[9]

III. Modesty

> If economists could manage to get themselves thought of as humble, competent people, on the level with dentists, that would be splendid![10]

Section IF speculates on whether research in neuroeconomics can guide economists to novel modeling assumptions. I fully agree with the challenge Bernheim makes at the end of the section. Neuroeconomic research should provide an example of a "*novel economic model derived originally from neuroeconomic research that improves our measurement of the causal relationship between a standard exogenous*

[9] The commercial ventures that argue fMRI technology can detect lies (http://www.noliemri.com/) are evidence that some believe that it is possible to associate brain activation with an important behavioral variable. Leaving aside the validity of these claims, I am not aware of economic uses of conventional lie detectors.

[10] John Maynard Keynes (1963, 373).

environmental condition... and a standard economic choice" (Bernheim 2009, 27). One can meet this challenge with an assumption that can be understood without neuroeconomic techniques and be explained in economic terms. Neuroeconomics can contribute by alerting traditional economists to assumptions that they may have overlooked. In fact, Camerer, Loewenstein, and Prelec (2004, 575) propose that including a new behavioral variable, the "degree of automaticity," into contracting models would lead to a better understanding when moral hazard or adverse selection problems arise.[11]

The most effective influence neuroeconomic research will have on formulating assumptions will be subtle. The alternative assumptions suggested by neuroeconomic experiments will be more readily accepted if they are so intuitively compelling that researchers take them seriously without evaluating the underlying research and, in fact, can find independent justifications using more traditional tools. Here neuroeconomists must learn to be modest, because while their research may have an important influence on the kinds of questions asked by the rest of the profession, mainstream researchers will claim that their assumptions would be fully justified without reference to the neuroeconomic literature.

IV. Conclusion

> What can be said at all can be said clearly; and whereof one cannot speak thereof one must be silent.[12]

Neuroeconomics has received attention in mass-market media. Some of the brightest stars in the profession now specialize in designing, running, and interpreting experiments based on neuro-imaging technology. A few departments have invested in setting up capital-intensive research centers. At the same time, several equally bright stars have expressed strong skepticism about the contributions of the area and its potential to make future contributions.

Lacking expertise and wisdom, I have tried to heed Wittgenstein's advice by organizing my comments around Bernheim's essay rather than broader issues. I applaud Bernheim for his insightful and evenhanded comments. I have resisted the temptation to issue an immediate judgment on the field of neuroeconomics and, instead, wait patiently to see if it delivers results that have a lasting impact on the profession.

REFERENCES

Andreoni, James, and B. Douglas Bernheim. Forthcoming. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica.*

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

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

[11] Isabelle Brocas and Juan D. Carrillo (Brocas and Carrillo 2004), Drew Fudenberg and David K. Levine (Fudenberg and Levine 2006), and Bernheim's work with Rangel on addiction (Bernheim and Rangel 2004) are examples of theoretical work that motivates novel assumptions by using research on the brain.

[12] Ludwig Wittgenstein (1998, 27).

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics*, 124(1): 57–104.

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson. 2007. "Is Generosity Involuntary?" *Economics Letters*, 94(1): 32–37.

Brocas, Isabelle, and Juan D. Carrillo. 2004. "Biases in Perceptions, Beliefs, and Behavior." University of Southern California Center in Law, Economics, and Organization Research Paper C04-19.

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

Cardozo, Benjamin N. 1949. *The Nature of the Judicial Process.* New Haven: Yale University Press.

Crawford, Vincent P. 2008. "Look-ups as the Windows of the Strategic Soul." In *The Foundations of Positive and Normative Economics: A Handbook*, ed. Andrew Caplin and Andrew Schotter, 249–80. New York: Oxford University Press.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. "What You Don't Know Won't Hurt Me: Costly (But Quiet) Exit in Dictator Games." *Organizational Behavior and Human Decision Processes*, 100(2): 193–201.

Ekeland, Ivar. 1990. *Mathematics and the Unexpected.* Chicago: University of Chicago Press.

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

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica*, 69(6): 1403–35.

Kable, Joseph W., and Paul W. Glimcher. 2007. "The Neural Correlates of Subjective Value During Intertemporal Choice." *Nature Neuroscience*, 10(12): 1625–33.

Keynes, John Maynard. 1963. "Economic Possibilities for our Grandchildren." In *Essays in Persuasion*, 358–73. New York: W. W. Norton & Company.

Lax, Eric. 2000. *Woody Allen: A Biography.* New York: Da Capo Press.

Lazear, Edward, Ulrike Malmendier, and Roberto Weber. 2006. "Sorting in Experiments with Application to Social Preferences." National Bureau of Economic Research Working Paper 12041.

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.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature*, 38(4): 827–74.

Rubinstein, Ariel. 2006. "Instinctive and Cognitive Reasoning: A Study of Response Times." Fondazione Eni Enrico Mattei Working Paper 2006-36.

Tadelis, Steven. 2007. "The Power of Shame and the Rationality of Trust." Unpublished.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness.* New Haven: Yale University Press.

Tom, Sabrina M., Craig R. Fox, Christopher Trepel, and Russell A. Poldrack. 2007. "The Neural Basis of Loss Aversion in Decision-Making Under Risk." *Science*, 315(5811): 515–18.

Wittgenstein, Ludwig. 1998. *Tractatus Logico-Philosophicus.* Mineola, New York: Dover Publications.