CHAPTER

9

The Psychology and Neurobiology of Judgment and Decision Making: What's in it for Economists?

B. Douglas Bernheim

OUTLINE

INTRODUCTION

The last few years have witnessed impressive progress toward understanding the neurobiology of decision making. Many participants in this growing field, as well as interested observers, hope that neuroeconomics will eventually make foundational contributions to the various traditional fields from which it emerged, including economics, psychology, and artificial intelligence. My purpose here is to evaluate its potential contributions to economics.

Some would argue that any aspect of economic decision making is definitionally an aspect of economics. According to that view, neuroeconomics necessarily contributes to economics by expanding the set of empirical questions that economists can address. I will avoid such semantic quibbles. My interest here is in assessing whether, in time, neuroeconomics is likely to shed useful light on traditional economic questions. While the scope of traditional economics is difficult to define with precision, I am content with an operational definition, based on the collection of questions and issues currently discussed in standard economic textbooks and leading professional journals.

The potential importance of neuroeconomics for economic inquiry has already been the subject of much debate. For example, an optimistic assessment appeared in a paper titled "Neuroeconomics: Why Economics Needs Brains," by Colin Camerer, George Loewenstein, and Drazen Prelec (2004)[1]. Subsequently, Faruk Gul and Wolfgang Pesendorfer (2008) penned a broad critique of neuroeconomics, titled "The Case for Mindless Economics," which expressed deeply rooted skepticism. My assessment lies between those extremes. I caution against dismissing the entire field merely because current technology is limited, or because some of the early claims concerning its potential contributions to standard economics were excessive and/or poorly articulated. However, because I share many of the conceptual concerns raised by Gul and Pesendorfer, I also see a pressing need for a sober and systematic articulation of the field's relevance. Such an articulation would ideally identify standard economic questions of broad interest (e.g., how taxes affect saving), and outline conceivable research agendas based on actual or potential technologies that could lead to specific, useful insights of direct relevance to those questions. Vague assertions that a deeper understanding of decision-making processes will lead to better models of choice will not suffice to convince the skeptics.

In Bernheim (2008), I have attempted to identify and articulate the specific ways in which neuroeconomics might contribute to mainstream economics, as well as the limitations of those potential contributions. This chapter briefly summarizes both my reservations and my reasons for guarded optimism. Due to space constraints, it touches only lightly on many important issues; readers are referred to the longer version for a more comprehensive and detailed discussion. Perhaps most significantly, I focus here exclusively on positive economics, as does nearly all existing research on neuroeconomics[2]. For the reasons discussed in Bernheim (2008), the possible applications of neuroeconomics to normative economic analysis are intriguing and largely unexplored[3]; see also Bernheim and Rangel (2004, 2007a, 2007b, 2008).

As will be evident, my evaluation of neuroeconomics (as it pertains to standard ecoomics) is based in large part on the contemplation of research agendas that may or may not become technologically or practically feasible. My contention is only that there are conceivable paths to relevant and significant achievements, not that success is guaranteed. At this early stage in the evolution of neuroeconomics, the speculative visualization of such achievements is critical, both because it justifies the continuing interest and patience of mainstream economists, and because it helps neuroeconomists to hone more useful and relevant agendas.

A FRAMEWORK FOR DISCUSSION

While neuroeconomists are convinced that a better understanding of *how* decisions are made will lead to better predictions concerning *which* alternatives are chosen, many traditional economists greet that proposition with skepticism. Advocates and critics of neuroeconomics (as it pertains to standard economics) often appear to speak at cross-purposes, using similar language to discuss divergent matters, thereby rendering many exchanges largely unresponsive on both sides. In the earnest hope of avoiding such difficulties, I will first provide a framework for my discussion, so that I can articulate and address particular issues with precision.

Suppose our objective is to determine the causal effects of a set of environmental conditions, x, on a decision vector, y[4]. For the time being, we will take x to include only the types of variables normally considered by economists, such as income and taxes. We recognize nevertheless that y depends not only on x, but also on a set of unobservable conditions, ω, which may include variables of the type studied by neuroeconomists. We hypothesize that the causal relationship between y and the environmental conditions, (x, ω), is governed by some function f:

$$y = f(x, \omega) \tag{9.1}$$

It is important to emphasize that the function f could be either a simple reduced form (e.g., a demand

[1]See also Glimcher and Rustichini, 2004; Camerer *et al.*, 2005; Rustichini, 2005; Glimcher *et al.*, 2005; Rustichini, 2005; Camerer, 2007.

[2]The object of positive analysis is to make descriptive, testable statements concerning factual matters. In answering positive questions about decision making, an economist typically attempts to predict the alternatives an individual would choose under specified conditions.

[3]The objective of normative analysis is to make prescriptive statements – that is, statements concerning what *should* occur.

[4]Sometimes, the objective of traditional positive economics is simply to forecast y given a set of observed conditions x, without interpreting the forecasting relation as causal. In some contexts, it may be helpful to condition such forecasts on neuroeconomic variables; see for example the discussion on p. 120.

function expressing purchases of a good as a function of its own price, the prices of other goods, and income), or a more elaborate structural economic model. For instance, f could identify choices that maximize some objective function given the available alternatives when the conditions x and ω prevail[5].

Economists typically treat the unobserved conditions, ω, as noise, and attempt to determine the causal effects of the observed environmental conditions, x, on the *distribution* of decisions, y. If the distribution of ω is governed by a probability measure μ, then the distribution of y will correspond to a probability measure $\eta(\cdot|x)$, where for any Borel set A, $\eta(A|x) = \mu(\{\omega|f(x,\omega) \in A\})$. For example, the standard linear model assumes that

$$f(x,\omega) = x\beta + \varepsilon(\omega),$$

where ε is an unspecified function. It follow that $\eta(A|x) = \mu(\{\omega|x\beta + \varepsilon(\omega) \in A\})$.

Generally, economists attempt to estimate η directly from data on observable conditions, x, and decisions, y. In the case of the linear model, they estimate the parameter vector β along with parameters governing the distribution of $\varepsilon(\omega)$. There is no opportunity to recover the form of the function ε or the distribution of ω. Nor is there an obvious need. For example, when studying the behavioral effect of a sales tax on consumption, a traditional economist would not be concerned with quantifying the variation in consumption attributable to specific genetic traits; rather, he would focus on the distribution of responses (most notably the average) without conditioning on genetics. Accordingly, the identification of the causal relation $\eta(\cdot|x)$, where x consists of standard economic variables such as income and taxes, is arguably the primary objective of traditional positive economics.

In contrast, the objective of positive neuroeconomics is, in effect, to get inside the function f by studying brain processes. To illustrate, let's suppose that neural activity, z (a vector), depends on observed and unobserved environmental conditions, through some function Z:

$$z = Z(x,\omega)$$

[5] In the latter case, an economist would typically interpret the free parameters of the objective function as aspects of preferences. However, modern choice theory teaches us that preferences and utility functions are merely constructs that economists invent to summarize systematic behavioral patterns. We are of course concerned with the accurate estimation of those parameters, but only because they allow us to recover the behavioral relation f.

Choices result from the interplay between cognitive activity the environmental conditions[6]:

$$y = Y(z,x,\omega)$$

It follows that

$$f(x,\omega) = Y(Z(x,\omega),x,\omega)$$

Positive neuroeconomics attempts to uncover the structure of the functions Z (the process that determines of neural activity) and Y (the neural process that determines decisions). Neuroeconomics necessarily treats the function f as a reduced form, even if it represents a structural economic model. Neuroeconomic research can also potentially shed light on the distribution of ω (the measure μ), which is the other component of η, the object of primary interest from the perspective of traditional positive economics.

The tasks of traditional positive economics and positive neuroeconomics are therefore plainly related. The question at hand is whether their interrelationships provide traditional positive economists with useful and significant opportunities to learn from neuroeconomics.

IS THE RELEVANCE OF NEUROECONOMICS SELF-EVIDENT?

Most members of the neuroeconomics community believe that the relevance of their field to economics is practically self-evident; consequently, they are puzzled by the persistent skepticism among mainstream economists. To motivate their agenda, they sometimes draw analogies to other subfields that have successfully opened "black boxes." For example, some liken neuroeconomics to the theory of the firm, which opened up the black box of production decisions (see Camerer *et al.*, 2004, 2005; Camerer, 2007).

From the perspective of a mainstream economist, analogies between neuroeconomics and the theory of the firm are misleading. 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

[6] The arguments of Y include x and ω in addition to z because the same neural activity could lead to different outcomes depending on the environmental conditions.

concern the nature of organized exchange between individuals. An economist who seeks to understand prices, wages, risk-sharing, and other traditional aspects of resource allocation has an undeniable stake in understanding how trade plays out within a range of institutions, including markets and firms, and how different types of exchange come to be governed by different types of institutions. In contrast, the mind is not an economic institution, and exchange between individuals does not take place within it.

Notably, economists have not materially benefited from a long-standing ability to open up other black boxes. For example, we could have spent the last hundred years developing highly nuanced theories of production processes through the study of physics and engineering, but did not. A skeptical mainstream economist might also note that models of neural processes are also black boxes. Indeed, the black-box analogy is itself false: we are dealing not with a single black box, but rather with a Russian doll. Do we truly believe that good economics requires a command of string theory?

It is therefore understandable that so many economists are unmoved by the amorphous possibility that delving into the nuts and bolts of decision making will lead to better and more useful economic theories. To persuade them that a particular black box merits opening, one must at least provide a speculative roadmap, outlining reasonably specific potentialities which economists would recognize as both exciting and within the realm of possibility. What has been offered along these lines to date is far too vague and insubstantial to convert the skeptics.

SOME SPECIFIC SOURCES OF SKEPTICISM

Neuroeconomists have certainly attempted to offer economists a variety of affirmative motivations for opening the black box of the human mind. Many mainstream economists find those motivations unpersuasive because they see neuroeconomic inquiry as largely orthogonal to traditional economic analysis, a view that finds its most forceful articulation in the work of Gul and Pesendorfer (2008). To identify motivations that economists would generally find persuasive, one must first understand the logic of that view, and appreciate its appeal.

Much of the prevailing skepticism concerning the magnitude of the contribution that neuroeconomics can potentially make to standard positive economics arises from the following three considerations.

First, unless neuroeconomics helps us recover the behavioral relation η, its contributions will not advance the historical objectives of positive economics. Though the functions Y and Z are obviously interesting, the questions they address directly are not ones that mainstream economists traditionally examine.

Second, because the behavioral relation η involves no neural variables, traditional positive economists can divine its properties from standard economic data. Distinguishing between two neural processes, (Y, Z, μ) and (Y', Z', μ'), is helpful to such an economist only if the differences between those processes lead to significant differences between the corresponding reduced form representations, η and η'. But if the latter differences are indeed significant, then an economist can test between η and η' directly using standard economic data, without relying on neuroeconomic methods.

Third, while neuroeconomics potentially offers another route to uncovering the structure of the relation η, there is skepticism concerning the likelihood that it will actually improve upon traditional methods. The prospects for building up a *complete* model of complex economic decisions from neural foundations would appear remote at this time. Even if such a model were assembled, it might not be especially useful. Precise algorithmic models of decision making of the sort to which many neuroeconomists aspire would presumably map highly detailed descriptions of environmental and neurobiological conditions into choices. In constructing the distribution η from Y, Z, and μ, a microeconomist would treat vast amounts of this "micro-micro" information as noise. An economist might reasonably hope to apprehend the structure of η more readily by studying the relationship between y and x directly, particularly if the explanatory variables of interest (x) include a relatively small number of standard environmental conditions. As an example, suppose η is the household demand function for a good. What does a standard economist lose by subsuming all of the indiosyncratic, micro-micro factors that influence decisions, many of which change from moment to moment, within a statistical disturbance term? What can neuroeconomics teach us about the relationship between average purchases and the standard economic variables of interest (prices, income, and advertising) that we cannot discern by studying those relationships directly?

These considerations do not, however, rule out the possibility that neuroeconomics might make significant contributions to mainstream economics. With respect to the second consideration, even the most skeptical economist must acknowledge that the standard data required to address questions of interest are sometimes unavailable, and are rarely generated under ideal conditions. Surely we should explore the possibility that new types of data and methods of analysis might help us overcome those limitations. Thus, the third consideration emerges as the most central to my appraisal, and the rest of this chapter is devoted to its evaluation.

In principle, even without providing a complete neural model of complex economic decision making, neuroeconomics offers several potential routes to uncovering the structure of standard behavioral relationships. First, it will lead to the measurement of new variables, which may usefully find their way into otherwise standard economic analyses. I discuss that possibility in the next two sections. Second, detailed knowledge concerning the neural processes of decision making may help economists discriminate between theories and/or choose between models. As discussed on pp. 120–121, the formulation of rigorous tests may prove challenging. Standard economic theories of decision making concern choice patterns, and are therefore agnostic with respect to decision processes; hence, they may have few testable neural implications. The penultimate two sections examine the more modest possibility that understanding a neural process may provide economists with informal but nevertheless useful guidance with respect to model selection (specifically, explanatory variables and functional forms).

A skeptic might observe that the most promising routes to meaningful contributions are also the most limited. An economist who examines neural variables would not necessarily require extensive knowledge of neuroeconomic methods or a deep appreciation of neural processes; instead, he might simply rely on neuroeconomists to identify and collect the relevant data. Similarly, even if findings from neuroscience informally guide aspects of model selection (variables and/or functional forms), once a traditional positive economist knows the structure of the selected model, he can discard all information concerning neural processes without loss.

Many psychologists would view the positions outlined above as a form of radical behaviorism. They are surprised that economists still hew so rigidly to a perspective that psychology abandoned decades ago. Yet the different paths of psychology and economics are not so difficult to understand once we consider divergent objectives of those disciplines. I would point to two important differences. First, unlike economics, the field of psychology has traditionally subsumed questions about the mind. Thus, traditional psychological questions pertain to aspects of the functions Y and Z, whereas traditional economic questions do not. Second, questions in psychology often focus on the micro-micro determinants of behavior. A psychologist is potentially interested the particular factors that cause a single individual to behave in a certain way at a specific moment. In contrast, traditional economic analysis usually treats such idiosyncratic influences as background noise.

ARE THERE USES FOR EXOGENOUS NEUROECONOMIC VARIABLES?

The discussion above takes $\eta(\cdot|x)$, with x defined to include only traditional economic variables, as the object of interest for traditional positive economics. It therefore ignores the possibility that neuroeconomics might redraw the boundary between the set of variables that economists treat as observable (x), and those they treat as unobservable (ω). More formally, by measuring some vector of variables $\tilde{\omega}$, a neuroeconomist can repartition the environmental conditions (x, ω) into (x^0, ω^0), where $x^0 = (x, \tilde{\omega})$ and $\omega = (\omega^0, \tilde{\omega})$, and potentially allow economists to recover the causal relation $\eta^0(\cdot|x^0)$. It is important to acknowledge that the barriers to redrawing this boundary may be practical and political (e.g., privacy concerns), not merely technological. For the purpose of this discussion, let us suspend disbelief and consider the possibilities.

Why might the distribution $\eta^0(\cdot|x^0)$, which subsumes the behavioral effects of neural variables, as well as the effects of standard environmental factors conditional on neural variables, be of interest to mainstream economists? The answer is not obvious. Suppose a neuroeconomist discovers a genetic trait that helps predict saving (a "patience gene"). Should economists greet that discovery with enthusiasm? Economics has not concerned itself historically with the relationship between genetics and saving. An economist might question whether that knowledge is likely to improve his understanding of the effects of, say, capital income taxes (an element of x) on asset accumulation, averaged or aggregated over the elements of ω (including genetics).

Further reflection suggests, however, that exogenous neural variables may have a variety of uses. For a more complete discussion of possible uses, along with examples, see Bernheim (2008). First, neural proxies for tastes and talents may facilitate the detection of biases arising from omitted variables, and the inclusion of such proxies in econometric specifications may mitigate omitted variables bias. Second, when the decisions of several distinct individuals are co-determined (as in peer groups), we may be able to measure the causal effect of one individual's choice on another's decision by using the first individual's exogenous neural predispositions as instruments. Third, if an economist is narrowly concerned with forecasting behavior as of a particular moment in time, and if a time-varying neural condition is known to affect the behavior in question, then the use of information concerning that condition can improve the forecast. Fourth, causal relationships that are conditioned on neural characteristics may be useful when projecting the effects of a policy from one population to another, particularly if the two populations differ compositionally. Fifth, understanding the roles of genetic predispositions in decision making may shed light on the likely sensitivity of behavior to policy interventions. Sixth, if private firms begin to measure the neural characteristics of consumers or employees and use that information in the course business, economists will need to consider the roles of neural variables in resource allocation. Even if governments prevent such activities due to privacy concerns, economists will be unequipped to evaluate the effects of such policies unless they study the neural correlates of behavior.

ARE THERE USES FOR ENDOGENOUS NEUROECONOMIC VARIABLES?

As I explained earlier, one of the main objectives of neuroeconomics is to uncover the structure of the function Y, which maps endogenous neural activity, z, along with the environmental conditions x and ω, to decisions. Based on existing findings concerning Y, it is already possible to predict certain choices from particular types of endogenous neural activity with a high degree of accuracy. For examples, see Knutson *et al.* (2007), Kuhnen and Knutson, (2005), and Hsu *et al.* (2005). Because accurate behavioral prediction is a central goal of positive economics, many neuroeconomists have offered such findings as evidence of their field's relevance (see, for example, Camerer, 2007).

Why are mainstream economists unpersuaded by this evidence? In the context of most traditional economic questions, they see little value in predicting behavior based on its endogenous components (here, z). Consider the following stark example. Suppose our goal is to predict whether individual customers at a grocery store will purchase milk. After carefully studying a large sample of customers, a confused graduate student declares success, noting that it is possible to predict milk purchases accurately with a single variable: whether the customer reaches out to grab a carton of milk. The technology to collect this highly predictive data has long been available; economists have demurred not due to a lack of creativity, boldness, and vision, but rather because such predictions are of no value to them.

Mainstream economists should not, however, completely dismiss the possibility that endogenous neural variables will prove useful. In some situations, information concerning some aspect of the environmental conditions, x, or the decision, y, may not be available. Data on neural activity (z) along with knowledge of the functions Y and Z can then potentially permit us to impute the missing conventional variables, and use the imputed values in otherwise standard economic analyses. For example, the analysis of Wang *et al.* (2006) suggests that it may be possible to infer private information concerning standard economic variables from neural responses. See Bernheim (2008) for a more detailed discussion of the possibilities for imputing both exogenous variables and choices.

DO ECONOMIC THEORIES HAVE TESTABLE IMPLICATIONS CONCERNING NEURAL PROCESSES?

Perhaps the most tantalizing claim concerning the potential prospects of neuroeconomics is that an understanding of neural processes may provide economists with new opportunities to formulate direct tests of both standard and nonstandard (behavioral) theories of decision making (see, e.g., Camerer, 2007)[7]. While such advances are conceivable, it is important for neuroeconomists to acknowledge the difficulty of that endeavor, and to avoid premature conceptual leaps, especially if they hope to be taken seriously by mainstream economists.

The central conceptual difficulty arises from the fact that standard economic theory (including neoclassical economics as well as much of modern behavioral economics) is agnostic with respect to the nature of decision *processes*. No explicit assumptions are made concerning the inner workings of the brain. For example, contrary to the apparent belief of many non-economists, economists do not proceed from the premise that an individual literally assigns utility values to alternatives, and from any opportunity set chooses the alternative with the highest assigned value. This disciplinary agnosticism with respect to process accounts for Gul and Pesendorfer's (2008) contention that neural evidence cannot shed light on standard economic hypotheses.

[7] This issue is distinct from the possibility that the measurement of neural variables may facilitate tests of conventional economic theories, e.g. by providing instruments or permitting reliable imputations for missing variables. The question here is whether one can test an economic theory of behavior by examining the *process* that governs decision making.

Foundational economic assumptions concern choice patterns, not processes. Neoclassical decision theory follows from a collection of choice axioms, the most critical of which is sometimes labeled *independence of irrelevant alternatives* (a generalization of the more familiar weak axiom of revealed preference). According to that axiom, if an individual chooses a particular alternative from an opportunity set, then he will also choose that alternative from any smaller set, provided the alternative remains available. When the independence axiom is satisfied, there exists an ordering (interpreted as preferences) that rationalizes all of the individual's choices, in the sense that he always chooses the most highly ranked alternative according to the ordering. With some additional (but largely technical) axioms, one can also represent his choices as maximizing a continuous utility function. Within this framework, preferences and utility are merely constructs, invented by the economist to provide a convenient representation of choice patterns. The theory does not assert that these constructs have counterparts within the brain. Consequently, those who would test the theory by searching for such counterparts have misunderstood the theory's foundations.

The preceding observations do not, however, imply that neural evidence is conceptually incapable of shedding light on standard economic hypotheses. Choice axioms cannot be valid unless the neural processes that govern choice are capable of delivering decisions that conform to the axioms; thus, a mainstream economist cannot remain *entirely* agnostic as to process. To take an extreme possibility, if neuroeconomists succeed in reducing all pertinent neural decision processes to a precise computational algorithm for some reasonably large class of decision problems, they will be able to determine whether the algorithm delivers choices that satisfy the independence axiom, and thereby test neoclassical decision theory. However, that potentiality does not convincingly establish the value of neuroeconomics, for two reasons.

First, assume we have reason to believe that the brain sometimes employs a particular decision algorithm, but have not yet established the scope of that algorithm's application. Suppose the algorithm's implications for choice within some domain of decision problems, A, would be inconsistent with some economic theory; moreover, there is no subset of A for which the same statement holds. We might hope to disprove the economic theory by demonstrating that the decision algorithm in fact governs choices throughout the domain A. However, a formal test of the latter hypothesis would presumably involve a comparison between the algorithm's behavioral predictions and actual choices throughout A. But if data on those decisions are available, we can test the economic theory directly, without concerning ourselves with the nuts and bolts of decision processes. Thus, the incremental contribution of neuroeconomics is not obvious.

Second, neuroeconomics is still a long way from reducing the neural processes that govern the complex decisions with which economists are conventionally concerned to precise algorithms, especially for broad classes of environments. Existing algorithmic representations of such processes pertain only to very simple tasks and functions. Much of what is known has a qualitative flavor – for example, that certain types of decisions involve elevated activity in particular regions of the brain, and that those regions tend to be associated with specific functions. While it is conceivable that we might be able to test economic theories using such information, the necessary conceptual groundwork for such a test has not yet been laid. See Bernheim (2008) for a more detailed dicsussion, including an analysis of what that groundwork would entail.

Unfortunately, the neuroeconomic community has not yet generally acknowledged the conceptual challenges that one necessarily confronts when attempting to derive testable implications of economic theories for neural processes. Instead, neuroeconomists have sometimes proceeded (at times implicitly) as if those implications are obvious or easily motivated. That practice leaves many mainstream economists with the regrettable (and often inaccurate) impression that neuroeconomists do not adequately understand the economic theories upon which they hope to shed light. I discuss three examples in Bernheim (2008): the contention that McClure *et al.* (2004) provided a neural test of quasihyperbolic discounting, the claim that Harbaugh *et al.* (2007; see also Chapter 20 of this volume) tested theories of altruism and "warm glow" giving, and the notion that the evidence in Platt and Glimcher (1999) supports expected utility theory. As I explain, none of those claims withstands scrutiny.

CAN AN UNDERSTANDING OF NEURAL PROCESSES USEFULLY GUIDE MODEL SELECTION?

The number of empirical models that an economist could construct to describe any particular decision as a function of conventional explanatory variables is vast. Even if neuroeconomics does not provide new variables of interest or an independent foundation for testing one model against another, it could conceivably generate suggestive findings that informally guide the search for an appropriate empirical model in useful directions, leading to more rapid and effective identification of the best predictive relationship. Here, the two main aspects of model selection are discussed: variable selection and the choice of functional form.

First, consider variable selection. Neuroeconomic evidence could in principle motivate the inclusion of particular conventional variables in specific behavioral models. Suppose, for example, that mandated transfers to others influence brain activity in centers linked to reward-processing (Harbaugh *et al.*, 2007; see also Chapter 20 of this volume). While such evidence would not prove that altruism motivates behavior, it might well *suggest* such a hypothesis to an empirical economist, who might then investigate behavioral models that incorporate related variables (e.g., measures of potential external effects). Similarly, an examination of neural evidence concerning the processes that govern attention might suggest that consumers are potentially susceptible to tax illusion, and that they will respond differently depending on whether a product is tagged with tax-inclusive or tax-exclusive prices. Such evidence might lead an empirical economist to examine empirical models that separately include explanatory variables measuring posted prices and hidden taxes.

While acknowledging the possibility described in the preceding paragraph, a skeptic might nevertheless question whether neuroeconomics is likely to make such contributions in practice. Empirical economists have other sources of guidance and inspiration, such as introspection and research from psychology. Indeed, neural studies such as that by Harbaugh *et al.* (2007; see also Chapter 20) are themselves motivated by hypotheses imported from other fields. Likewise, economists formulated and tested conjectures concerning tax illusion based on a common-sense understanding of attention, without the benefit of neuroeconomic evidence; see in particular Chetty et al. (2007), and Finkelstein (2007). Empirical economists who are not persuaded to investigate the roles of pertinent variables in behavioral relationships on the basis of other considerations are unlikely to find neural evidence convincing. To uniquely motivate the inclusion of a potential explanatory variable that empirical economists have ignored, a neuroeconomist would literally have to stumble across some unexpected environmental correlate of brain activity. I do not dismiss that possibility, but neither does it convince me that the field holds great potential for conventional positive economics.

Even if research on the neurobiology of decision making had provided the impetus for investigating altruism, tax illusion, or some other phenomenon, it seems unlikely that an empirical strategy for estimating the function η would have been influenced by the details of the neurobiological evidence. Rather, that evidence would have merely *motivated* (to use Gul and Pesendorfer's term) an examination of functional forms that include the pertinent variables. It is not at all obvious that an economist who possesses a deep understanding of the motivating scientific evidence would be any better equipped to estimate η than one who simply apprehends the pertinent psychological principles intuitively.

In addition to suggesting that certain variables may play roles in particular behavioral relationships, neuroeconomic evidence may also indicate that others play no role. Such evidence could motivate exclusion restrictions. Indeed, formal neural tests of exclusion restrictions are conceivable in principle, even without precise knowledge of the computational algorithms that govern decision making. We can frame the issue as a computer-programming task. To implement a choice mapping that depends on a particular variable, computer code must reference that variable. For any neural process that implements the same computational algorithm, there must presumably be some neural response to the variable's value. Consequently, the absence of any response would formally justify an exclusion restriction in the behavioral relationship.

Next, consider the choice of functional form. In principle, the nature of neurobiological response mechanisms may suggest particular empirical specifications. For example, there is some evidence that temporal difference reinforcement learning (TDRL) models accurately describe the operation of neural systems governing dopamine learning (Schultz *et al.*, 1997; Schultz, 1998, 2000). These parsimonious, tightly parameterized learning models could guide the formulation of empirical behavioral relationships in settings that involve the accumulation of experience. Because other learning processes may also influence choices, the neural evidence cannot *prove* that one functional form is better than another for the purpose of predicting behavior. However, it could lead economists to examine particular parsimonious specifications that they might not otherwise consider, and some of these may outperform more conventional alternatives.

A mere catalog of such possibilities will never suffice to convince the skeptics, nor should it. Mainstream economists should acknowledge the conceptual possibilities discussed above, and exercise intellectual tolerance and patience while neuroeconomists explore them. Neuroeconomists in turn should recognize that the burden of proof is squarely on their shoulders. Skeptical reactions define a specific challenge: *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 environmental condition – one with which economists have been historically concerned – and a standard economic choice*. Unless the neuroeconomics community eventually rises to that challenge, the possibilities discussed in this section will eventually be dismissed as unfounded speculation.

CAN NEUROECONOMICS IMPROVE OUT-OF-SAMPLE PREDICTIONS?

Sometimes, economists wish to predict behavior under completely novel conditions (for example, a new and untried public policy). There is no assurance that reduced form behavioral models will perform well in such contexts, especially if the novel conditions are qualitatively distinct from any that have preceded them. In contrast, a good structural model, based on a deeper understanding of behavior, may permit reasonable projections even when fundamental environmental changes occur. Many neuroeconomists believe that their field will provide such models.

By way of analogy, suppose a computer has been programmed to make selections for choice problems falling into a number of distinct categories, but the tasks for which we have observed its choices belong to a subset of those categories. We could potentially develop a good positive model that predicts the computer's choices for problems within the categories for which we have data. However, based on that limited data, projecting choices for problems within the remaining categories is guesswork. Now suppose we obtain the computer code. In that case, even without additional choice data, we could accurately predict the computer's decisions in *all* circumstances. When neuroeconomists suggest that an understanding of the brain's computational algorithms will permit more reliable out-of-sample behavioral predictions, they are making an analogous claim.

Unfortunately, the issue is not quite so straightforward. If neuroeconomists only succeed in mapping a subset of the brain's neural circuitry to computational algorithms, out-of-sample prediction will remain problematic. To pursue the analogy to a computer program a bit further, suppose we obtain the code only for certain subroutines that are activated when the computer solves problems falling within the categories for which we have data. There is no guarantee that it will activate the same subroutines for related purposes when confronting problems within the remaining categories, particularly if those problems are qualitatively different from the ones previously encountered. Without knowing how the entire program operates, including the full array of subroutines upon which it can call, as well as the conditions under which it activates each of them, one cannot simulate its operation in fundamentally new environments.

Of course, one can proceed based on the *assumption* that the brain will continue to use the same neural circuitry in the same way when confronting new classes of decision problems. But there is no guarantee of greater out-of-sample stability at the neural level than at the behavioral level[8]. Whether we would be better off making out-of-sample predictions from structural neural models rather than structural behavioral models is therefore a factual question that can only be settled through experience, and not through logical arguments.

Still, there are reasons to hope that consideration of evidence on neural processes might at least help us select *economic* models that are more reliable for the purpose of making out-of-sample projections. Imagine, for example, that an estimated within-sample behavioral relationship is equally consistent with several distinct structural economic models, each of which has a different out-of-sample behavioral implication. Suppose the available neural evidence informally persuades us (but does not prove) that one of those models is more likely to match reality. Then we might reasonably hope to obtain more accurate out-of-sample predictions from the preferred model.

Consider the following example. Currently, tens of millions of people lack health insurance coverage. One theory holds that those households have carefully assessed the costs and benefits of insurance, and concluded that it is too costly; another holds that they are inattentive to their health-care needs, and hence unresponsive to costs and benefits. Both hypotheses are equally consistent with observed choices, but they have starkly different out-of-sample implications concerning the fraction who would purchase insurance if the cost of coverage were reduced well below historical levels. Can neuroeconomics help us judge between their divergent predictions? Suppose we use

[8]Just as a structural economic model can be viewed as a reduced form for a structural neural model, any structural neural model can also be viewed as a reduced form for some deeper structure, and the stability of the neural reduced form over classes of environments will depend on how that deeper structure operates. If, for example, secondary neural systems are designed to override a primary system whenever the latter would generate behavior too far from some norm, then an incomplete neural model of choice might be less stable out of sample than a behavioral model.

neural methods to measure attentiveness to health-care needs, as well as value assessments for insurance coverage. The first theory informally predicts high attentiveness and high-value assessments; the second has the opposite prediction. Neither finding would *prove* that the uninsured are more likely to behave one way or the other out of sample. For example, the uninsured might *start* attending to health-care issues and contemplating the benefits of insurance if they thought health care was affordable. Even so, the neural evidence would presumably influence our comfort with and degree of confidence in each model.

These possibilities are of course speculative. Mainstream economists will relinquish their skepticism only when confronted with examples of superior out-of-sample prediction in contexts involving the types of environmental conditions and behaviors that they ordinarily study.

CONCLUSIONS

The potential for the emerging field of neuroeconomics to shed light on traditional economic questions has been overstated by some, unappreciated by others, and misunderstood by many. With respect to positive economics, the case for studying the neural foundations of decision making is hardly self-evident. Nevertheless, neuroeconomics could in principle contribute to conventional positive economics in a number of ways, which I have attempted to catalog.

At the same time, a number of the potential contributions discussed in this paper strike me as somewhat modest, rather special, and/or somewhat peripheral. While there is good reason to hope that some of the contributions will prove noteworthy, I have considerably more difficulty convincing myself 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. Whether that assessment reflects the field's actual limitations or the deficient imagination of a relatively mild skeptic remains to be seen.

Due to space constraints, I have not evaluated potential contributions to normative economics. I doubt that neuroeconomics will provide a technology for measuring utility directly, or that it will replace choice-based welfare analysis with a new utilitarian paradigm. However, it may hold the potential to improve choice-based welfare analysis; see Bernheim (2008) for a detailed discussion.

Many neuroeconomists have been surprised and frustrated to learn that skepticism concerning their field's potential among mainstream economists runs deep. How can they combat that skepticism? First, neuroeconomists need to do a better job of articulating specific visions of the field's potential contributions to mainstream economics. Such an articulation would ideally identify a standard economic question of broad interest (e.g., how taxes affect saving), and outline a conceivable research agenda that could lead to specific, useful insights of direct relevance to that question. Vague assertions that a deeper understanding of decision-making processes will lead to better models of choice do not suffice. Second, it is essential to avoid hyperbole. Exaggerated claims simply fuel skepticism. Sober appraisals of the field's potential, including its limitations, will promote its acceptance more effectively than aggressive speculation that involves loose reasoning or otherwise strains credibility. Third, the ultimate proof is in the pudding. To convert the skeptics, neuroeconomists need to accumulate the right type of success stories – ones that illuminate conventional economic questions that attracted wide interest among economists prior to the advent of neuroeconomic research.

Acknowledgments

This paper was prepared for a symposium titled "Neuroeconomics: Decision Making and the Brain," held at NYU on January 11–13, 2008. I am grateful to Antonio Rangel and Colin Camerer for stimulating discussions and comments. I also acknowledge financial support from the National Science Foundation through grant number SES-0452300.

References

Bernheim, B.D. (2008). *Neuroeconomics: a sober (but hopeful) appraisal*. AEJ: Microeconomics, in press.

Bernheim, B.D. and Rangel, A. (2004). Addiction and cue-triggered decision processes. *Am. Econ. Rev.* 94, 1558–1590.

Bernheim, B.D. and Rangel, A. (2007a). *Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics*. Mimeo, Stanford University.

Bernheim, B.D. and Rangel, A. (2007b). Toward choice-theoretic foundations for behavioral welfare economics. *Am. Econ. Rev. Papers Proc.* 97, 464–470.

Bernheim, B.D. and Rangel, A. (2008). Choice-theoretic foundations for behavioral welfare economics. In: A. Caplin and A. Schotter (eds), *The Methodologies of Modern Economics*. Oxford: Oxford University Press in press.

Camerer, C.F. (2007). Neuroeconomics: using neuroscience to make economic predictions. *Economic J.* 117, C26–C42.

Camerer, C.F., Loewenstein, G., and Prelec, D. (2004). Neuroeconomics: why economics needs brains. *Scand. J. Economics* 106, 555–579.

Camerer, C.F., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: how neuroscience can inform economics. *J. Econ. Lit.* 43, 9–64.

Chetty, R., Looney, A., and Kroft, K. (2007). *Salience and taxation: theory and evidence*. Mimeo, University of California, Berkeley.

Finkelstein, A. (2007). *EZ-tax: tax salience and tax rates*. Mimeo, MIT.

Glimcher, P.W., Dorris, M.C., and Bayer, H.M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games Econ. Behav.* 52, 213–256.

Glimcher, P.W. and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452.

Gul, F. and Pesendorfer, W. (2008). In: A. Caplin and A. Schotter (eds.), *The Methodologies of Modern Economics*. Oxford, Oxford University Press. In press.

Harbaugh, W.T., Mayr, U., and Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625.

Hsu, M.M., Bhatt, M., Adolphs, R. *et al*. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683.

Knutson, B., Rick, S., Elliott Wimmer, G. *et al*. (2007). Neural predictors of purchases. *Neuron* 53, 147–156.

Kuhnen, C.M. and Knutson, B. (2005). The neural basis of financial risk taking. *Neuron* 47, 763–770.

McClure, S.M., Laibson, D.I., Loewenstein, G., and Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* 306, 503–507.

Platt, M.L. and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238.

Rustichini, A. (2005). Neuroeconomics: present and future. *Games Econ. Behav.* 52, 201–212.

Savage, L. (1954). *The Foundation of Statistics*. New York, NY: John Wiley & Sons.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.

Wang, J.T.-Y., Spezio, M., and Camerer, C.F. (2006). *Pinocchio's pupil: using eyetracking and pupil dilation to understand truth-telling and deception in biased transmission games*. Mimeo, Caltech.