CHAPTER 3

WHAT'S SO INFORMATIVE ABOUT CHOICE?

ANDREW SCHOTTER

> Those who can do science; those who can't prattle about its methodology.
>
> —Paul Samuelson

This is a chapter on economic methodology written by someone who, I hope, contradicts Samuelson's dictum. It is hard to discuss scientific methods without first agreeing on the purpose of our science. It is my claim here that the goal of economics, like the goal of any science, is to explain phenomena. As Milton Friedman [1953: 7] suggested, "The ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed."

Note that Friedman expects us to make "predictions about phenomena not yet observed," so our theories should be predictive ex ante and not simply ex post rationalizations of events we know the outcome of. While he certainly permits us to look back in history and explain the past, he states that

> "predictions" by which the validity of a hypothesis is tested need not be about phenomena that have not yet occurred, that is, need not be forecasts of future

events; they may be about phenomena that have occurred but observations on which have yet to be made or are not known to the person making the prediction. [7]

Being an experimentalist, my objective is to test theory by matching the predictions it makes with choices made in the lab, so to some extent, much of what I say here I comes off sounding like old-style positivism, but I think my position is far more nuanced (or perhaps confused) than that. In fact, I am a very reluctant positivist, my position being more like that of Clark Glymour [1980: ix]: "If it is true that there are but two kinds of people in the world—the logical positivists and the god-damned English professors—then I suppose I am a logical positivist."

Like other sciences, the process of discovery used to explain phenomena is first to construct a theory, then to test it empirically (experimentally), and finally to modify it in the light of empirical (experimental) evidence.

A question has been raised recently about what types of data are admissible in this scientific quest. For conventional theory, choices and their direct observable consequences, such as prices, are considered the only truly reliable types of data. While choices give us insights into the preferences of individuals, prices give us insights into the allocative properties of institutions (mostly markets). Variables such as beliefs, emotions, intentions, verbal reports, and mental states are suspect because only in making choices are people "putting their money where their mouth is." Only choices are reliable data in this operationalist view of science. This is what economics is all about. Faruk Gul and Wolfgang Pesendorfer (chapter 1) put it this way:

> Standard economics focuses on revealed preference because economic data come in this form. . . . The standard approach provides no methods for utilizing nonchoice data to calibrate preference parameters. The individual's coefficient of risk aversion, for example, cannot be identified through a physiological examination; it can only be revealed through choice behavior. . . . In standard economics, the testable implications of a theory are its content; once they are identified, the nonchoice evidence that motivated a novel theory becomes irrelevant.

If our goal is theory falsification and prediction, however, it is not clear that we would want to limit ourselves only to choice variables. We should be able to use whatever helps us in that task. If a variable is useful in testing theory or in prediction, why not use it?

In this chapter, I discuss the consequences of placing choices and only choices on such a pedestal. While I certainly do not question their usefulness (no one trained as an economist could) or even their primacy, I do ask why choices alone should be placed in such an exalted position and why they are considered to be reliable indicators of such things as preferences.

I explain my position in five sections. In each section I rely on experimental evidence to make my point. I tend to use experiments that I have done simply because I am most familiar with them and know how they can be used to make my point. My argument can be summarized by the following five points, which are discussed in the remaining sections of this chapter:

Point 1: Making a Virtue Out of a Constraint

Sanctifying choices confuses a constraint with a virtue—we rely on choices because the variables we are most interested in many times are unobservable non-choice variables (preferences, costs, beliefs, etc.), and we use choice to infer them. We use choices, therefore, because we are constrained to and not because proper science dictates that we must. But using choice data as a proxy for unobservables involves the creation of a theory linking observable choice data to the unobservable variables we are interested in. If that theory is faulty, the choice data used are suspect, and nonchoice variables may be as (or even more) reliable. Think of the structural empirical literature on auctions. The object of analysis is the underlying distribution of costs (or values), and this is identified by assuming all bidders use a Nash bid function and inverting to find the revealed costs associated with the observed bids. What we'd like to know directly is what costs or values are, but that is not available, so we are forced to use bids (choices). Using bid data is therefore a solution to a constraint and not a free choice.

Point 2: Knowing Why Instead of Knowing That

Theory cannot be confirmed by merely observing *that* a choice has been made that is consistent with it. We must know *why* it was made and verify it was made for the reasons stipulated by the theory. It is only when choice is consistent not only with the predictions of a theory but also with the reasons stipulated for that choice that we can be confident that the predictions of the theory will remain valid when the parameters of the model change. In other words, we need to know why a choice was made, and not just that it was made, in order to explain the comparative statics of the theory. Knowing why a choice was made may require data other than choices. This may require us knowing something more deeply psychological about the person making the choice, about his or her mental states, wants, needs, thought process, desires, and so on. It is here that a door is opened for neuroscience.

Point 3: Incomplete Real-World Data

We must decide what subset of available choice data we will restrict ourselves to. Real-world choice data are incomplete and therefore cannot be used alone to test theory. The world offers us data that have not been generated to test our theories, so inferences from real-world choices to the validity of a theory are incomplete and therefore unreliable. Improper inferences about the validity of a theory can be made if the data used do not span the relevant sets of treatments that one would need in order to (tentatively) validate a theory. This suggests that controlled experiments are required, but if they are not feasible, we may need to supplement our available choice data with other types.

Friedman agrees with this point, or at least laments the fact that economists cannot perform controlled experiments. His famous 1953 essay "The Methodology of Positive Economics," however, was written before the advent of experimental economics. "Unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences. Generally we must rely on evidence cast up by "experiments" that happen to occur" [xx].

Point 4: Framing and the Reliability of Choices

There has been a long and robust literature documenting the fact that how one frames a choice problem to a person can change the way that person makes his or her choice. The problem we face as social scientists is that when we observe choice data we do not always observe the frame used to elicit it. If that is the case, then the inferences we can make from these choices are very likely to be limited and unreliable unless we are able to observe the actual frame used.

Point 5: Institutional Design Using Neuro Data

While much has been made about the usefulness of neuroscience (mental state) data and their failure to be of interest to economic theory and of use in designing institutions, I present experimental evidence that such information can, in fact, be useful in that enterprise. That is, fMRI data can be used to design revenue-enhancing auction institutions and can do so by capitalizing on knowledge of the mental states bidders experience when they contemplate winning and losing.

Getting Biblical: Making a Virtue Out of a Constraint

Let me start with a biblical reference. When God kicked Adam and Eve out of the Garden of Eden, the punishment he placed on Adam was to force him to earn his living by the sweat of his brow (perhaps the birth of labor economics). Eve's punishment was to give birth in pain (certainly the beginning of labor economics).

Finally, the punishment that He placed on the serpent was forbidding him the ability to walk upright.

If economists had also been living in the Garden of Eden and had been kicked out, their punishment would have been to force them to make indirect inferences using only those variables that they could observe, namely, choices. In response, economists would need to infer what they, could see. The entire corpus of econometrics is involved in this enterprise, and the theory of revealed preference is another exquisite response to this constraint.

But our use of choices and only choices is merely a response to a constraint, and the fact that we as a profession have been so clever in constructing theories that explain the world, despite the paucity of the data available, does not mean that we should not search to expand the list of variables of interest and perhaps admit the use of what are presently unobservable but may, in the future, via new technologies, become observable. Stating, as Gul and Pesendorfer do, that "standard economics focuses on revealed preference because economic data come in this form" is equivalent to cursing the darkness rather than lighting a match.

In an influential article on the use of experiments in economics, Vernon Smith [1982: 929] writes, "Over twenty five years ago Guy Orcutt characterized the econometrician as being in the same predicament as that of an electrical engineer who has been charged with the task of deducing the laws of electricity by listening to the radio play." For many years, I failed to understand the significance of this quote, until I had to sit down and write this chapter. It then struck me that Orcutt is saying exactly what I am trying to say here—that the challenge of most sciences is precisely the fact that the world does not offer us data in the form we would like it to be, and the scientific process is therefore forced to be an inferential one, much like being constrained to try to understand the laws of electricity through inferences made only by listening to the radio.

The beauty of economics is how well we have done in the face of this constraint, but we should not conclude that that is our destiny. Obviously, Orcutt would have loved to have had more, better, and different types of data. If they were available, I would suspect that he would embrace them. No good applied economist throws away useful data of any type.

For example, what about such nonobservables as beliefs, intentions, and emotions? Clearly they affect behavior but are not directly observable and are not choice based. Are they ruled out as being useless to our scientific enterprise? To answer this question, let me talk about beliefs and learning.

Consider the theory of learning. Much of what is done in the lab is the comparison of various learning models, since one focus of attention in experimental economics is not only theory falsification but also theory comparison: which theory predicts better. When comparing theories or models, however, one must make sure that the data used for the comparisons are comparable. More precisely, every model has an ideal data set that, if available, would allow it to perform at its peak efficiency. When such a data set is not available, one is forced to use proxies for those variables that are unobservable, and one must provide a theory linking these proxies to the behavior of the unobservable variable that one is interested in. For example, take two models, A and B, and assume that the ideal data set for model A is observable to the analyst while this is not the case for model B. If we were then to compare the performance of these two models and conclude that model A was better, we would face the problem of not knowing how much of its performance to impute to the model and how much to impute to the fact that it had its ideal data set available while model B did not and had to use proxies. First-best comparisons of two models occur only when each has its ideal data set available.

The consequences of these ideas for the comparison of learning models are numerous and significant. For example, reinforcement models of learning are fortunate in often being able to have their ideal data sets observable. This is so because they typically rely only on the observable outcomes of play (actions and payoffs) as inputs into their models. In other words, in order to calculate a player's propensity to choose a given strategy, all that is needed is a time series of that player's own payoffs from all previous plays of the game. The Experience Weighted Attraction (EWA) model, as written, is a model where now attractive a strategy is depends on how it has been reinforced in the past when it was used and counterfactually when it was not. It also has its ideal data set available since it posits a model in which strategies are chosen in accordance with their attractions that, as defined by Camerer and Ho [1999] are functions of past observable outcomes.[1] In general, ideal data sets for belief learning models are not available since their basic components, beliefs, are private and hence not observable. To rectify this, experimentalists and theorists have used history-based proxies as substitutes for these beliefs, but as Nyarko and Schotter [2002] have demonstrated, these proxies are inadequate. Using a proper scoring rule to elicit truthful beliefs about the actions of their opponents, Nyarko and Schotter [2002] found that belief learning models that use elicited beliefs outperform all other belief learning models that employ historical proxies for their belief estimates.

To explain, consider a "belief learning" model such as fictitious play, where at every point in time people form beliefs about their opponent based on the observable frequency of his or her actions in the past. Once these beliefs are formed, agents best respond to these beliefs either deterministically or stochastically.

Note that if we are only allowed to use observables in our models, then their predictive value will only be as good as the theories we create connecting these observables to the unobservable variables we are most interested in. For example, fictitious play presents us with a rather rudimentary model connecting past choices with beliefs. If this model is faulty, which it most certainly is (see Nyarko and Schotter [2002] and Ehrblatt, Hyndman, Ozbay, and Schotter [2007] for two examples), then the predictions of the theory will fail. But if we could find ways to directly measure beliefs, then we may be able to produce a more reliable theory. Let me explain.

The model of fictitious play as specified is a fully complete model based only on observables—in fact, choices. Choices are the basis of beliefs, and beliefs are the basis of choice. Note that beliefs do not depend on the payoffs that one's opponent receives, just his or her choices in the past. Hence, in this theory, one does not need to even know the payoffs of the game to one's opponent, much less his or her unobservable utility function. If one fears that the fictitious play belief model is too primitive, one can improve it by adding a more flexible form belief equation such as that of Cheung and Friedman [1997], where past choices by one's opponent are geometrically discounted. Still, restricting us to using only choices as inputs ties the reliability of the theory to the theory relating previous choices to beliefs.

Under the constraint that beliefs are unobservable, these models, or models like them, might be all we could do. But what if we could observe beliefs or make them observable? Would the resulting model explain behavior better, and if so, might we be missing something by limiting our models only to data that are choice based?

Nyarko and Schotter [2002] and others [Huck and Weizsäcker, 2002; Costa-Gomes and Weizacker, 2006] have tested belief learning models using beliefs elicited with a proper scoring rule. In other words, people are paid for their beliefs in such a way that revealing one's true belief is a dominant strategy. This is one technology that laboratory experiments have provided us for making unobservables observable. Of course, one could say that since people are paid for their beliefs, this elicits choice data and hence its use does not violate the constraint, but we could have alternatively simply asked people to reveal to us their beliefs and use them, or waited until fMRI technology is developed to a point where it can locate where beliefs are created in the brain and use them. The relevant question is which learning models are better predictors of behavior: choice-based belief or elicited (fMRI measured or self-reported) belief learning models?

So ultimately, if we do better by simply asking people what they believe than by relying on historical models, then that is legitimate, especially if our goal is prediction. Why restrict our data to those that are choice based?

Nyarko and Schotter [2002] come to the conclusion that belief learning models based on elicited beliefs (using a proper scoring rule) do a far better job of describing behavior than do those based only on historical choice. While this is not the place to go into detail, let me provide two pieces of evidence to support their position. First, one could ask if the time series of beliefs for fictitious play and elicited beliefs are similar for individuals. In other words, do historical-choice-based fictitious-play beliefs look like elicited beliefs? The answer here is a resounding no. To understand why, consider figure 3.1, a and b, which shows a striking difference in two belief series: while the elicited belief series is very volatile and indicates that beliefs can take on very extreme values, as expected, the fictitious play beliefs are very stable and more centered. Obviously, this must be true by construction, so elicited beliefs look nothing like fictitious-play, choice-based beliefs.

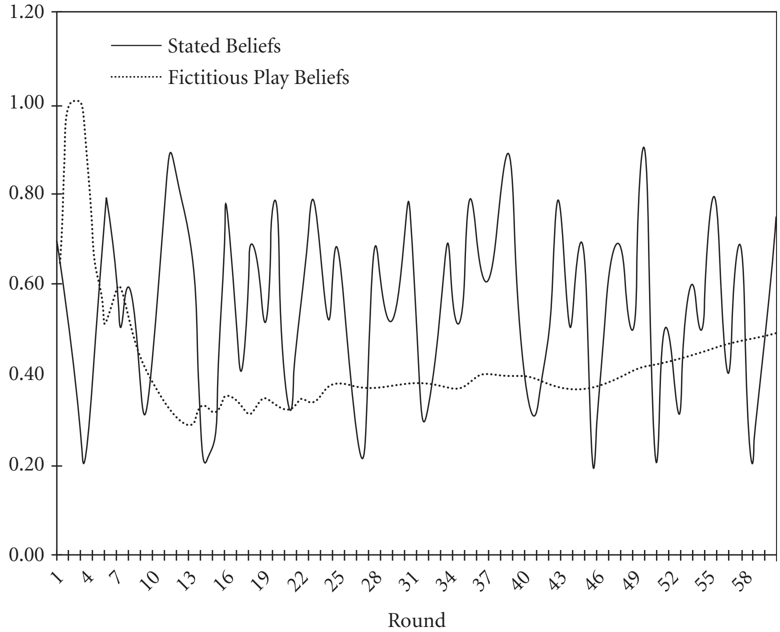

Figure 3.1. Time series of elicited beliefs and fictitious play beliefs of a subject who played the 2 × 2 constant sum games shown in table 3.1 repeatedly with the same partner for 60 periods. From Nyarko and Schotter [2002, figure 2a,b].

Table 3.1. Payoff Matrix for Nyarko and Schotter [2002]

Strategy	Green	Red
Green	6, 2	3, 5
Red	3, 5	5, 3

Rows indicate strategy for player 1; columns, strategy for player 2. Cells show the payoff for player 1 followed by the payoff for player 2.

But the proof of the pudding is in the eating. Which predicts choices better? In other words, which type of beliefs, when used in the logit choice model

$$\text{Probability of Red in period } t = \frac{e^{\beta_0+\beta_1\left[E(\pi_t^d)\right]}}{1+e^{\beta_0+\beta_1\left[E(\pi_t^d)\right]}}, \tag{3.1}$$

$$\text{Probability of Green in period } t = 1 - \frac{e^{\beta_0+\beta_1\left[E(\pi_t^d)\right]}}{1+e^{\beta_0+\beta_1\left[E(\pi_t^d)\right]}},$$

where $E(\pi_t^d)$ is the expected payoff difference to be derived from using the Red strategy instead of the Green strategy in period t given the beliefs that the subjects hold at that time, best fits the actual choices made by the subjects?

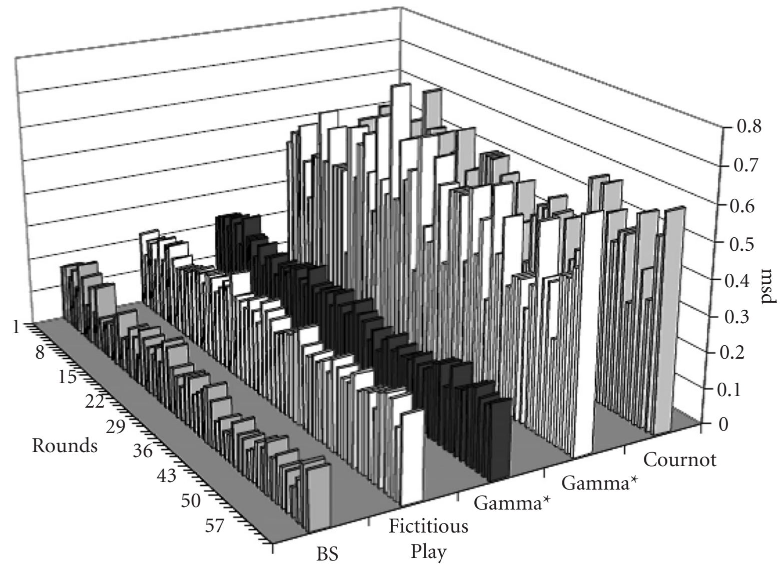

Figure 3.2. Mean cross-subject mean standard deviation scores for five different methods of belief estimation. From Nyarko and Schotter [2002].

To answer this question, Nyarko and Schotter [2002] constructed the following goodness-of-fit measure. For each individual and for each of our logit models (i.e., stated or elicited, fictitious play), we have an estimated β_0 and β_1 coefficient. Hence, for any round, if we were to plug one of our belief measures into the logit equation, we would get a predicted probability of Red (and Green) for that round. This predicted probability vector can be compared to the actual $\{0, 1\}$ choice vector made in that round to generate a standard deviation (SD) score for that subject in that round. If in any round we average these scores across the K subjects in the experiment, we get a mean cross-subject MSD score (MCSMSD) for round t defined as follows:

$$\text{MCSMSD}_t = \frac{1}{K}\sum_{i=1}^{K}(p_t^i - a_t^i)^2,$$

where p_t^i is the predicted probability of choosing Red for subject i in round t, and a_t^i is i's actual choice (equal to 1 if Red is chosen and 0 if Green is chosen). For any model, we estimated we can define 60 such MCSMSDs, one for each round of the experiment.

Figure 3.2 presents these scores for experiment 1. Our stated belief model clearly outperforms the fictitious play model. This is seen by comparison of the MCSMSD measure for the BS model using elicited beliefs and the fictitious play model. The MCSMSD is smaller for BS than for fictitious play in every round, suggesting clearly that elicited beliefs are better proxies for "true beliefs" or those beliefs that seem to be used to make decisions.

So what does this all mean? To my mind, it means that we can make good use in economics of unobservable nonchoice variables if we can find a reliable way to discover them. This may be done in several ways. First, it can be done by inferring them from choices. In Nyarko and Schotter [2002], it was done using a proper scoring rule. But the same exercise they did could be done by merely asking people their beliefs or running the experiment in an fMRI machine if the science progresses far enough to discover where in the brain beliefs are processed. Restricting ourselves to choice-based belief models such as fictitious play makes sense only if we are confident that the theory of belief formation we use, which relies only on previous historical choices, is more reliable in predicting beliefs than are similar theories we could construct that would relate beliefs to brain scans or verbal reports. This is an empirical question, not a matter of economic methodology or theology.

Knowing "Why" Instead of Knowing "That"

To return to the theme I started with, if our objective is to test theory, then we must ask what constitutes a proper test. According to standard economic theory, proof is finding choices that are consistent with predictions, so choices validate theory (or falsify it). But can we stop there? Don't we, perhaps, need to know that the choices are made for the reasons stipulated by the theory? This is important for a number of obvious reasons, but mostly because it is only when we know that a theory works for the right reasons that we can be confident that it will continue to work when the parameters of the world it is applied to change, that is, only then will we have confidence in the theory's comparative statistics.

This fact has been demonstrated in a very clever way by Holt and Goeree [2001]. They present subjects with a number of games that they play once and only once. The clever thing they do, however, is to present them with several versions of the same game, one of which they call a "treasure" treatment, since it is their expectation that in that treatment the Nash predictions will hold. In the other versions, they change one of the parameters defining the game and investigate whether, when they change this parameter, they will be able to predict behavior.

For example, consider the 2×2 matching pennies game in table 3.2. When this game is offered to subjects, they choose top left 48% of the time and bottom right 52% of the time. Since the Nash prediction is 50–50, this appears to be strong support for the theory. But the theory suggests that these choices are being made in an effort to make one's opponent indifferent to his or her choice and hence are invariant to changes in one's own payoff. Hence, when the payoff of 80 is changed to 320 in the

Table 3.2. Payoff Matrix for Symmetric Matching Pennies

Strategy	Left	Right
Top	80, 40	40, 80
Bottom	40, 80	80, 40

Rows indicate strategy for player 1; columns, strategy for player 2. Cells show the payoff for player 1, followed by the payoff for player 2.

Table 3.3. Payoff Matrix for Strongly Asymmetric Matching Pennies

Strategy	Left	Right
Top	320, 40	40, 80
Bottom	40, 80	80, 40

Rows indicate strategy for player 1; columns, strategy for player 2. Cells show the payoff for player 1 followed by the payoff for player 2.

top left cell, as is true in table 3.3, Nash predicts that this will not change the behavior of the row chooser at all. This is far from the truth. When that payoff is changed, row players choose too 96% of the time, in strong contradiction to the predictions of the theory. The choices made in the symmetric game were not sufficient to confirm the theory. While choices were theoretically consistent, the motives behind them were not—the subjects were not making Nash choices for Nash reasons. So while choices are our bread and butter, they alone are not sufficient confirmatory data for theory. We need to know why, and finding out why cannot easily be done by simply looking at choices. In this case, one might as well ask subjects what they were doing, and while such data is not on par with choices, it certainly could come in handy if subjects offered a coherent heuristic that explained their behavior.

This sentiment is common among experimentalists. For example, Johnson, Camerer, Sen, and Rymon [2002] use Mouselab in an experiment about backward induction to try to infer the thought processes that subjects went through in order to assess whether they were using backward induction, one of the cornerstones of game-theoretic logic. Broseta, Costa-Gomes, and Crawford [2001] used similar techniques in trying to assess how subjects reasoned in normal form games. In a paper dealing with signaling games, Partow and Schotter [1993] (henceforth PS) investigate whether subjects choose to behave according to various refinements of Nash equilibrium for the reasons stipulated by the theory.

To understand what they did, consider the signaling game described in table 3.4, first used by Banks, Camerer, and Porter [1994] (henceforth BCP) and Brandts and Holt [1993] (henceforth BH). In this game, there are two players, a sender, player S, and a receiver, player R. The sender can before of two types, type H and type L,

Table 3.4. Payoff Matrix for Signaling Game 1 of Brandts and Holt [1993]

	Message N			Message S		
Type	C	D[a]	E	C[b]	D	E
H	15, 30	30, 30	0, 45	45, 15	0, 0	30, 15
L	30, 30	15, 60	45, 30	30, 30	0, 0	30, 15

Columns show the message of the sender and the strategy of receiver. Rows show different types for the sender. Cells show the payoff for the sender followed by the payoff for the receiver.
[a] Message N and strategy D form the nonsequential Nash equilibrium.
[b] Message S and strategy C form the sequential equilibrium.

which is chosen with equal probability before the game starts. The type is revealed to player S but not player R, although player R knows the probability distribution determining types. Once player S knows his type he must send a message of either N or S. The message determines which of the two payoff matrices shown above is relevant for the game. At this point it is player R's turn to move, and she chooses an action of C, D, or E. Payoffs are then determined by the message sent by S, the action chosen by R, and the partially known type of player S.

In this game, there are two Nash equilibria or two Bayesian-Nash equilibria, only one of which is sequential. In the nonsequential equilibrium player S sends message N no matter which type he is, while the receiver takes action D no matter what message is sent. In the sequential equilibrium, player S sends message S no matter what type she is, and a message of S is responded to by action C while a message N is responded to by action D. This game, therefore, has two equilibria, one of which is more refined than the other.

Let us then see how the refinement literature expects us to rule out the nonsequential Nash equilibrium from consideration. Consider either type of subject at the Nash equilibrium. Note that if player S is of type H, she will receive a payoff of 30 at the equilibrium, while if she is of type L she will get 15. According to the equilibrium, if the out-of-equilibrium message S is ever sent, it should be met with a reply of D leading to a payoff of 0 for each player. However, such a response on the part of player R is irrational since it requires that he would choose his dominated strategy D instead of his weakly dominant strategy C, leading to a payoff of 45 for type H and 30 for type L. As a result, both types have an incentive to deviate from the Nash equilibrium and can be expected to do so. It is this process that game theory expects subjects to work their way through.

Clearly, this is a very complex thought process, one that might well exceed the reasoning capabilities of mere mortals. In fact, it is this complexity that has led us to question whether game-theoretic logic could be responsible for the observed behavior of subjects in the laboratory experiments run on such games even when that behavior is consistent with the predictions of the theory.

In response to the same set of concerns, BH posit the following naive decision model to explain people's behavior in such games. Assume that player S thinks of his opponent not as a player but as a naive decision machine that chooses actions C, D, and E with equal probabilities. In such a case, when player S is of type H he will find it beneficial to choose message S since that message will maximize his expected utility. If player S is of type L, message N will be sent. Likewise, if player R believes that messages N and S are equally likely to be sent by either type, she will choose action C when S is sent and D when N is sent. Hence, these choices may be what we expect in the early rounds of the experiments. As time goes on, however, type L subjects will have an incentive to deviate since they are better off sending message S if player R is going to choose action C than they are sending message N and having D chosen. As a result, the play of the game converges to sequential equilibrium but for a very different reason than suggested by the refinement literature.

Now consider game 2, described in table 3.5. This game has exactly the same equilibria as does game 1 except that here, because of the payoff changes, the BH heuristic does not select the sequential equilibrium but rather the Nash equilibrium. So if choices converge to the sequential equilibrium in game 1 but the Nash equilibrium in game 2, such convergence cannot be the result of game-theoretic logic since in both cases the refinement predicts the sequential equilibrium.

PS replicate the BH experiment with one significant difference: subjects were only allowed to see their own and not their opponent's payoffs. If this is the case, then game-theoretic logic is ruled out since such logic is impossible without knowledge of one's opponent's payoffs. Hence, if we see behavior or choices that are identical to those predicted by the theory, it is evidence that such behavior is not occurring because of the theory and that something else is underlying it. More important, if behavior changes in a manner that is consistent with the BH heuristic when we move from game 1 to game 2, then that indicates that their heuristic is a better approximation to the thinking process of subjects than that of the Nash refinements. So again, simply seeing choices consistent with Nash in game 1 is not a good confirmation of the theory—we need to know why.

Table 3.5. Payoff Matrix for Signaling Game 2 of Brandts and Holt [1993]

	Message N			Message S		
Type	C	D[a]	E	C[b]	D	E
H	75, 30	45, 30	75, 45	60, 15	0, 0	0, 45
L	75, 30	15, 75	75, 30	45, 60	0, 0	45, 15

Columns show message of the sender and strategy of the receiver. Rows show different types for the sender. Cells show the payoff for the sender followed by the payoff for the receiver.

[a] Message N and strategy D form the nonsequential Nash equilibrium.

[b] Message S and strategy C form the sequential equilibrium.

Table 3.6. Proportion of Outcomes by Refinement

Game	Nonequilibrium	Nash equilibrium	Sequential equilibrium	Sample size
1[a]	.25	.12	.63	60
1[b]	.10	.32	.58	72
1[c]	.05	.31	.64	71
2[b]	.28	.50	.22	72
2[c]	.22	.68	.13	72

[a] Results from Banks, Camerer, and Porter [1994] in the first five rounds.
[b] Results from Brandts and Holt [1993] for the first six periods of their experiment.
[c] Results from Partow and Schotter [1993].

The results obtained by PS, as well as those reported in BH and BCP, are summarized in table 3.6. This table presents the results of the PS game 1 and game 2 experiments along with the results of BH and BCP, who ran identical games. Note that in the PS version of game 1, only 5% of the outcomes were nonequilibrium (vs. 10% found by BH and 25% for BCP), while 64% were sequential (vs. 58% for BH and 63% for BCP). In game 2, 22% of the PS outcomes are nonequilibrium (vs. 28% in BH); PS also found a bigger shift toward the Nash outcome and away from the sequential outcome as the payoff structure was modified.

We consider the data presented in table 3.6 to support for our hypothesis that game-theoretic logic is not the determinant of the results obtained by either BCP or BH. Hence, merely observing that strategic choices are consistent with a theoretical refinement, as was true in BCP, is not sufficient grounds to conclude that a theory is supported—one needs to know why.

Incomplete Real-World Data

If we are to restrict ourselves to choice data in testing theory, the question becomes whether or not the choice data existing in the world present a sufficiently rich environment upon which to test theory. More precisely, to fully put a theory through its paces, we need to see it function in a variety of parametric situations. If the world presents us only with data taken from a small subset of the potential parameter space upon which the theory is defined, then even if the choices observed are consistent with the predictions of the theory, we have no idea if the theory would continue to be supported if the world changed sufficiently to present another part of the parameter space. As a result, if the choice data we face are not sufficiently rich to fully test a theory, then supplementing the data we look at with nonchoice data may be advisable.

Let me explain this point by referring to an experiment conducted by Kfir Eliaz and myself [Eliaz and Schotter, 2006]. In this experiment, subjects were faced with the following simple decision problem: A monetary prize X is hidden in one of two boxes, labeled A and B: The probability that each box contains the prize depends on the state of nature. There are two possible states: high and low. The probability that box A contains the prize is h in the high state and $l < h$ in the low state, where h and l are both strictly above 1/2. The subject's task is to choose a box. If she chooses correctly, she wins the prize. Before she makes her choice, the subject can pay a fee to learn the state of nature. If she chooses not to pay, the subject then must choose a box without knowing the state of nature. Whether or not the subject pays the fee, she receives a payment immediately after she makes his choice. Since a choice of A first-order stochastically dominates a choice of B, knowing the state of nature should not affect one's choice. This is true under any model of decision making under risk that respects first-order stochastic dominance. It is also independent of the subject's attitude toward risk. Thus, a subject who pays the fee exhibits an intrinsic preference for noninstrumental information.

In this experiment, we ran three treatments, a baseline and two variants, although I only discuss two of them here: the baseline and treatment 2. In the baseline, subjects are first asked if they want to pay a fee to learn the state of nature *before* making their choice. If they answer yes, they are shown the true state and are then asked to choose a box. If they answer no, they are asked to choose a box without any further information. In treatment 2, subjects are first asked to choose a box without knowing the state of nature. After they make their choice, but before being paid, they are given the opportunity to pay a fee to learn the state of nature.

In Eliaz and Schotter [2006] we posited two theories that would explain that despite the fact that the information offered had no instrumental value, subjects overwhelmingly were willing to pay for it. One is the theory of Kreps and Porteus [1978], which indicates that subjects are willing to pay the fee because they have a preference for the early resolution of uncertainty. The other is a hypothesis that indicates that people pay the fee because they get utility from being confident about the decisions they make at the time they are making them and are willing to pay just to feel good about their choices. We call this effect the "confidence effect" since we posit that people are willing to pay a fee in order to be confident that what they are doing is correct.

Note that these two explanations have very different implications for choice in the baseline and treatment 2. According to the Kreps-Porteus theory, if a subject is willing to pay a fee in the baseline, he should be willing to also pay it in treatment 2 since both situations offer him identical date 1 lotteries over lotteries at date 2 (see Eliaz and Schotter [2006] for a demonstration of this point). In contrast, the confidence effect predicts that subjects should be willing to pay for noninstrumental information only before they make their decision (as in the baseline) but not in treatment 2 since there the decision has already been made and the information comes too late to make them feel confident while they make their decision.

Table 3.7. Choice Situations—Baseline and Treatment 2

Situation	α	β	Fee ($)
3	1.0	0.60	0.50
6	1.0	0.51	2.00
12	1.0	0.51	0.50

This table refers to experiments in Eliaz and Schotter [2006]. α is the probability that box A contains $20 in the high treatment, while β is the probability that A contains $20 in the low treatment.

Table 3.8. Fraction of Subjects Paying Fee [Eliaz and Schotter, 2006]

Situation	Baseline	Treatment 2
3	0.783	0.063
6	0.565	0.125
12	0.739	0.125

To make a long story short, we found a stunning rejection of the Kreps-Porteus explanation and strong support for the confidence effect. More precisely, consider three situations (called situations 3, 6, and 12 in Eliaz and Schotter [2006]) presented in table 3.7. In these situations, the probability that box A contains the $20 in the high state, h, is α, the probability that box A contains the $20 in the low state is β, and f is the fee that needs to be paid in order to find out what the state is.

Subjects in the baseline and in treatment 2 were asked if they would be willing to pay the fee in one of these situations. The results are shown in table 3.8. Note that these data are a stunning disconfirmation of the Kreps-Porteus theory in this particular instance. While the fraction of subjects willing to pay for noninstrumental information is extremely high in the baseline, in treatment 2, where the propositions should be the same according to Kreps-Porteus, it is dramatically lower. In other words, after a choice has been made (as in treatment 2) subjects are no longer willing to pay a fee to reveal the state.

My point here is not to refute the Kreps-Porteus theory since this experiment falls far short of a general refutation. Rather, it is to demonstrate that if we relied on real-world choice data and the world was arranged such that people were never put in a position where they were asked to pay a fee for information after they had made a decision, then we would never have been able discover the weakness in the Kreps-Porteus theory for this particular choice situation. In other words, say the world was never in treatment 2 but always existed in the baseline condition. If that were the case, then by observing choices, we would infer strong support for behavior consistent with Kreps-Porteus, yet that inference might be wrong specifically because the world was not constructed to test that theory. Observed real-world choices alone seem inadequate to confirm a theory when naturally occurring data do not offer the proper tests.

Clearly, this is the justification for experimental economics in general, but if such experiments are not feasible, it still might be beneficial to simply ask people why they are doing what they are doing. If they say, "I am choosing to pay the fee because I prefer to get information earlier rather than later," then such statements support the Kreps-Porteus theory. If they say, as one of our subjects did, "I did understand the chance that the money was in box A was greater than 50%; however, I wanted to be positive of my choice," then they are refuting it. If they say nonsense, then who knows. Still, choices alone are inadequate to separate our conflicting explanations.

Framing and the Reliability of Choices

As stated in the introduction to this chapter, it is well known that the framing of problems can affect the choices made when those problems are presented to people. The question here is how this fact impinges on the reliability of choice data. The answer is not at all, if we as scientists know the frame used to generate the data we are using. But in many situations, this is not known. All we have are problems described by feasible sets, actions, and consequences and must infer some unobservable (preferences or risk attitudes) from the choice data we have. To illustrate how this may lead us astray, consider the famous framing example of Kahneman and Tversky [1984], modified here to represent a choice between pure money gambles (as opposed to choices about policy measures that would save or cost lives).

Consider the following two problems:

Problem 3.1: Imagine that you are about to lose $600 unless you take an action (choose a risky gamble) to prevent it. If gamble A is chosen, $200 of your $600 loss will be saved for sure. If gamble B is chosen, there is a one-third probability that all $600 will be saved and a two-thirds probability that you will not save any of the $600. Which of the two action would you favor?

Problem 3.2: Imagine that you are about to lose $600 unless you take an action (choose a risky gamble) to prevent it. If gamble A is chosen, you will wind up losing $400 for sure. If gamble B is chosen, there is a one-third probability that nothing will be lost and a two-thirds probability that you will lose the entire $600.

As we know, people when confronted with these two problems typically choose the risky action when the problem is framed negatively (stressing losses) as in problem 3.2. When the problem is framed positively (money saved), just the opposite is the case. Now say that your task was to estimate the risk aversion of people using the choices they made but you know only that they faced a problem where they could choose one gamble {1, −\$400} that would give them a sure loss of \$400 and another {1/3, \$0; 2/3, −\$600} that gave them a lottery with a 1/3 chance of losing nothing and a 2/3 chance if losing \$600 (both problems present these identical lotteries, just framed differently), but you do not know the frame. Obviously, if the frame of problem 3.1 is used we will get a very different answer than if the frame of problem 3.2 were used. So any inference we make is inherently unreliable. Of course we might try to infer what the frame was or, if we suspect different frames were used for different people, what the distributions of frames were, but this is making the situation worse, not easier. In other words, the reliability of choice data is suspect if such choices are not robust to such things as framing or other possible interferences, such as emotional states.

Institutional Design Using Neuroscience Data

Two claims have been made to discount the use of neuroscience data in economics. The first is that, while neuroscience data may be of interest to neuroscientists, it is of no interest or use to economists since we simply don't care why people do what they do as long as we can predict that they do it—the "who cares what's under the hood" argument. The second is that since we care about designing the proper set of institutions for use in society, neuroscience can hardly be expected to be of use in that endeavor since we do not care where a person's preferences come from. This argument is summarized by Gul and Pesendorfer in chapter 1 as follows: "The purpose of economics is to analyze institutions, such as trading mechanisms and organization structures, and to ask how those institutions mediate the interests of different economic agents. This analysis is useful irrespective of the causes of individuals' preferences." In this chapter I hope to demonstrate how these claims may be misguided. In other words, neuroscience data may be of use in the design of economic institutions precisely because it gives us insights into mental states that affect behavior.

To explain all of this, consider the experimental literature on auctions. There is an often perceived experimental regularity that laboratory subjects bidding in first-price sealed-bid auctions tend to bid above the risk-neutral Nash equilibrium bid function (see, e.g., Cox, Robertson, and Smith [1982]). One of the standard explanations of this phenomenon is risk aversion, since in these auctions risk aversion increases bids. But those same authors who suggest risk aversion invariably make reference to another explanation that they call the "joy of winning." Under this explanation, there is something inherently satisfying about winning that makes it worth bidding above one's Nash bid simply to experience the satisfaction of winning. Winning represents a positive utility shock—an enjoyable mental state that people are willing to pay for.

If choice data were all we could use in disentangling these two influences, then we might proceed as Cox et al. [1982] and Goeree, Holt, and Palfrey [2002] and try to infer from bid data whether a "joy of winning" parameter can be estimated and can have a significant and positive sign. This is indeed what Goeree et al. [2002] did.

In their study, they presented subjects with an auction where there are a finite (in fact six) values that a subject could have for the good being sold and only seven bids allowable to be made. They posit that subjects make "noisy" bids for any value they receive as dictated by the following power function:

$$P(b \mid \nu) = \frac{U^e(b \mid \nu)^{\frac{1}{\mu}}}{\sum_{b'=0}^{\nu-1} U^e(b' \mid \nu)^{\frac{1}{\mu}}}, \quad b = 0, 1, 2, \ldots \nu,$$

where $U^e(b \mid \nu)$ is the expected payoff of a subject given her conjecture about the probabilities that her opponent will make each of their bids. Using a constant relative risk aversion utility function of the form $U(b \mid \nu) = \frac{(\nu-b)^{1-r}}{1-r}$, Goeree et al. [2002] estimate μ and r after solving for the fixed point, implied by the quantal response equilibrium.

To investigate the "joy of winning," they modified their original utility function to include a joy of winning term as follows: $U(b \mid \nu) = (\frac{(\nu-b)^{1-r}}{1-r} + w)$. They then recalculated their estimation and compared the results. They concluded that there is little support for the joy of winning:

> The coefficient on the joy of winning is low in magnitude and has the wrong sign. The constrained model ($w = 0$) cannot be rejected since the inclusion of w adds little to the likelihood. Moreover, the estimates of μ and r are unaffected by the inclusion of w. Thus the "joy of winning" at least as we have formulated it here, does not add anything to the explanation of overbidding. [266]

Similar conclusions had previously been reached by most investigators (see Cox et al. [1982]).

Now the question whether we can do better than this by supplementing bid data with fMRI data and, more important, whether we can use these data to help us design better first-price auctions.

Delgado, Ozbay, Schotter, and Phelps [2007] tried to do exactly this: subjects bid while we observed brain activity using fMRI. As with Goeree et al. [2002], subjects could only receive a finite number of values (in our case, one of four values) in each round {12, 10, 8, 6} and make only one of four bids {8, 7, 5, 2}. The symmetric risk-neutral Nash equilibrium is for subjects to bid 8, 7, 5, and 2 for values 12, 10, 8, and 6, respectively. Brain responses were observed both while a subject was contemplating his bid and also when he received feedback as to whether he had won or lost in a given round.

What we found is that the striatum, a region of the brain that is known to be involved in processing reward, showed a differential response when subjects won or lost. Activation in this region increased, relative to a resting baseline, when subjects were informed that they won the auction and decreased when subjects were informed that they lost. An exploration of the relation of this brain signal to the bid placed by the subject revealed that only the loss signal significantly correlated with the bid amount, suggesting that subjects' anticipated "fear of losing" may be more strongly linked with an increase of their bids in a given round. In other words, instead of an anticipated "joy of winning" affecting bidding behavior, we seem to have uncovered a significant impact of the anticipated "fear of losing" (see Delgado et al. [2007] for details).

This result is established as follows. For each subject, we have data on the bid she made given the value she received in each round. We also have data on the strength of the striatal signal when the subject was told that she either won or lost a given profit (or profit opportunity). This strength of signal is measured by the blood oxygenation-dependent (BOLD) response, represented as a β, which assesses the hemodynamic changes in the striatum when the subject is informed of a win or loss. This allows us to run two separate regressions of the following type, one for win situations and one for loss situations:

$$\text{bid}^j = \alpha + \text{b}_1 \text{ value} + b_2(\beta) + \varepsilon,$$
$$j = \text{win, lose},$$

where ε is a normally distributed error term. Our results are presented in tables 3.9 and 3.10. While the striatal signal in loss situations is significantly related in a positive way to bids, the same is not true for striatal signals in win situations. We interpret this to mean that when subjects anticipate winning and losing before they make their bids, it is the anticipated fear of losing that has a significant impact on the

Table 3.9. Relationship Between Bid, Value, and Striatum Signal for Losses [Delgado et al., 2007]

Variable	Coefficient	Standard error	Probability = 0
β	0.509	0.1599	0.002
Value	0.307	0.0587	0.000
Constant	3.04	0.4574	0.000

$N = 163$, $R^2 = .166$, $F(2.160) = 15.98$, Prob $> F = 0.0000$.

Table 3.10. Relationship Between Bid, Value, and Striatum Signal for Wins [Delgado et al., 2007]

Variable	Coefficient	Standard error	Probability = 0
β	0.116	0.1256	0.357
Value	0.508	0.0438	0.000
Constant	1.75	0.4583	0.000

$N = 166$, $R^2 = .14527$, $F(2.163) = 67.40$, Prob $> F = 0.00000$.

bids that are made and not the anticipated joy of winning. This is true despite the fact that the activation in the striatum is significantly different from the baseline (at rest) activation for both win and loss situations.

Taking this result as informative led Delgado et al. [2007] to run a behavioral experiment to see if this hypothesis, derived from brain imaging, could be exploited by a mechanism designer in his design of an auction. More precisely, if the fMRI data implied that subjects' anticipated fear of losing caused them to increase their bid in order to avoid that unpleasant state, then that result implies that if we were to take a first-price auction and in one experiment make the losing state more unpleasant while in another make winning more enjoyable (while still keeping the equilibrium bids of subjects in these two auctions equal), we would expect, if our inference is correct, that the loss-emphasized auction would not only increase bids conditional on value, but also raise more revenue than either the control auction, where there was no manipulation of the win or loss outcome, or the treatment where the benefits of winning were enhanced.

To do this, Delgado et al. [2007] ran three first-price auctions, which can be called here the baseline, "win-frame," and "loss-frame" auctions. In each experiment, subjects engaged in a two-bidder auction with a random opponent for 30 rounds.

The baseline was a typical first-price auction where subjects were given values drawn uniformly from the interval [0, 100] and made bids accordingly. The loss-frame auction was identical to the baseline except that subjects were told the following:

> At the beginning of each round you will be given a sum of 15 experimental dollars, which are yours to keep if you win the auction. This will be your "initial endowment." If you lose the auction, you will have to give this initial endowment of experimental dollars back. Only the person who wins will be able to keep them. In other words, your payoff in this auction will be equal to your value minus your bid plus your initial endowment of 15 experimental dollars if you win the auction and zero if you lose, since by losing you must give back your initial endowment.

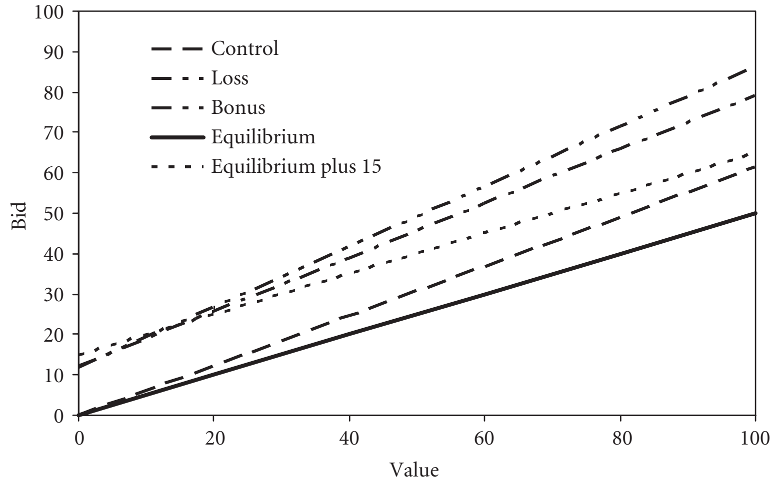

Figure 3.3. Estimated value functions for five different experimental conditions. See text for details.

The win-frame auction was again identical to the baseline except they were told that "the payoffs for the auction you engage in will differ slightly from what is described above since in addition to receiving your value minus your bid if you win, you will also be given an additional sum of 15 experimental dollars. Only the winner will receive this sum so if you lose your payoff is still zero."

Note that while the risk neutral Nash equilibrium of the two nonbaseline auctions are identical, subjects should bid what they bid in the baseline plus 15 experimental dollars, and that risk aversion affects these bid functions identically, so there should be no difference in the bids made in each auction no matter how it is framed. Our fMRI data suggest that because we have highlighted losing in the loss-frame auction, we should observe higher bids there than in the win-frame auction.

Our fMRI conjectures were supported by the data of our behavioral experiment, as shown in figure 3.3 and tables 3.11 and 3.12, both of which report the results of our estimation of the bid functions of subjects (a quadratic specification was rejected in preference to a linear one). Note in figure 3.3 that the bid function for the loss-frame treatment is almost everywhere above that of all other treatments. This is true despite the fact that both the win- and loss-frame treatments have identical equilibrium bid functions.

In terms of expected revenue, we can calculate what the expected revenue to the auctioneer would be if subjects bid as they did in our three treatments and received values drawn from a uniform distribution defined over [0, 100]. Given our regression results, we found that the expected revenue from our auctions were 46.70, 41.89, and 40.67 for our loss-frame, win-frame, and baseline treatments,

Table 3.11. Estimated Bid Function for Loss Frame [Delgado et al., 2007]

Variable	Coefficient	Standard error	Probability = 0
Value	0.748	0.0136	0.000
Constant	11.70	2.130	0.000

$N = 540$, $R^2 = .765$.

Table 3.12. Estimated Bid Function for Bonus Frame [Delgado et al., 2007]

Variable	Coefficient	Standard error	Probability = 0
Value	0.671	0.0142	0.000
Constant	12.08	2.130	0.000

$N = 660$, $R^2 = .805$.

respectively, and 33.33 for the risk-neutral prediction. In other words, we have been able to leverage a neuroeconomic finding into a result that allows us to increase the expected revenue received from a first-price auction.

The point I am making here is that it may be of use for us as economists to know the "causes of individuals' preferences," or at least what motivates people, since that may make a difference in how we design auctions. In addition, it is unlikely that we could have designed our loss-frame experiment without the aid of some nonchoice data. Knowing why our subjects bid above Nash equilibrium was important to our ultimate mechanism design, and while this insight might have been arrived at in other ways, we have at least demonstrated that it may be possible to arrive at it using fMRI data. Finally note that risk aversion cannot explain our results here since no matter what the level of a bidder's risk aversion, his behavior is expected to be the same in both the loss-frame and the win-frame treatments. In short, it appears as if nonchoice data may be useful for mechanism design.

Conclusion

In this chapter I question why we, as a profession, are so reluctant to allow data other than choice data into our analysis. My point is that if the object of our enterprise is to test the properties and accuracy of our theories, then choices alone are many times not sufficient. This is true for several reasons. First, the use of choices as proxies for unobservables is only as good as the theory we have relating those choices to what we would like to observe. Second, to fully test theory, we need to know why choices are made, and this might require nonchoice data. Third, real-world data are incomplete and therefore cannot alone be used to test theory. If a laboratory experiment cannot be run to fill in the missing observations, then we may need to rely on nonchoice data. Finally, nonchoice neuroscience data may give us insights useful for intelligent mechanism design.

NOTES

1. The fact that Camerer and Ho [1999] suggest that EWA nests belief learning models is more a matter of interpretation and not essence. The two types of models are quite different in spirit, and EWA only nests belief learning models of a particular fictitious-play-like type. Other belief learning models that define beliefs differently are not nested in this model.

REFERENCES

Banks, J., Colin F. Camerer, and D. Porter. 1994. Experimental Tests of Nash Refinements in Signaling Games. *Games and Economic Behavior* 6: 1–31.

Brandts, J., and Charles Holt. 1993. Adjustment Patterns and Equilibrium Selection in Experimental Signaling Games. *International Journal of Game Theory* 22: 279–302.

Broseta, B., Miguel Costa-Gomes, and Vincent Crawford. 2001. Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica* 69: 1193–1235.

Camerer, C., and Teck H. Ho. 1999. Experience-Weighted Attraction Learning in Normal Form Games. *Econometrica* 67(4): 827–874.

Cheung, Y.-W., and Daniel Friedman. 1997. Individual Learning in Normal Form Games. *Games and Economic Behavior* 19: 46–76.

Costa-Gomes, M., and Georg Weizsäcker. 2006. Stated Beliefs and Play in Normal-Form Games. Mimeo, London School of Economics.

Cox, J. C., Bruce Robertson, and Vernon L. Smith. 1982. Theory and Behavior of Single Price Auctions. In *Research in Experimental Economics*, ed. Vernon L. Smith, 1–43. Greenwich, CT: JAI Press.

Delgado, M., Erkut Ozbay, Andrew Schotter, and Elizabeth Phelps. 2007. The Neuroeconomics of Overbidding: The Neural Circuitry of Losses and Gains in Auctions. Unpublished manuscript. 2007.

Ehrblatt, W., Kyle Hyndman, Erkut Ozbay, and Andrew Schotter. 2007. Convergence: An Experimental Study of Teaching and Learning in Repeated Games. Mimeo, New York University.

Eliaz, K., and Andrew Schotter. Paying for Confidence: An Experimental Study of the Demand for non-Instrumental Information. Mimeo, New York University, June 2007.

Friedman, M. 1953. *Essays in Positive Economics.* Chicago: University of Chicago Press.

Glymour, Clark. 1980. *Theory and Evidence.* Princeton, NJ: Princeton University Press.

Goeree, J., Charles Holt, and Thomas Palfrey. 2002. Quantal Response Equilibrium and Overbidding in Private-Value Auctions. *Journal of Economic Theory* 104(1): 247–272.

Holt, C., and Jacob Goeree. 2001. Ten Little Treasures of Game Theory, and Ten Intuitive Contradictions. *American Economic Review December* 91(5): 1402–1422.

Huck, S., and Georg Weizsäcker. 2002. Do Players Correctly Estimate What Others Do? Evidence of Conservatism in Beliefs. *Journal of Economic Behavior and Organization* 47(1): 71–85.

Johnson, E., Colin Camerer, Sankar Sen, and Talia Rymon. 2002. Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory* 104(1): 16–47.

Kahneman, D., and Amos Tversky, A. 1984. Choices, Values and Frames. *American Psychologist* 39: 341–350.

Kreps, D., and E. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice Theory. *Econometrica* 46: 185–200.

Nyarko, Y., and Andrew Schotter. 2002. An Experimental Study of Belief Learning Using Elicited Beliefs. *Econometrica* 70: 971–1005.

Partow, Z., and Andrew Schotter. 1993. Does Game Theory Predict Well for the Wrong Reasons? An Experimental Investigation. *Working Paper* 93–46, New York University.

Smith, Vernon. 1982. Microeconomic Systems as an Experimental Science. *American Economic Review* 72(5): 923–955.