How (Not) to Do Decision Theory[1]

Eddie Dekel[2]   Barton L. Lipman[3]

First Draft
August 2009

[2]Economics Dept., Northwestern University, and School of Economics, Tel Aviv University E–mail: dekel@nwu.edu.

[3]Boston University. E–mail: blipman@bu.edu.

Abstract

We discuss the goals and means of positive decision theory and the implications for how to do decision theory. We argue that the goal of positive economic theory generally is to provide predictions and understanding and that representation theorems and other results of decision theory should be seen as ways to achieve these goals. We also argue that the "story" of a model is relevant to whether and how we use the model, that psychological considerations are not necessary for useful decision theory but can be helpful, and that nonchoice data, interpreted properly, can be valuable in predicting choice and so should not be ignored.

Whatever it is, I'm against it.

— Sung by Groucho Marx in *Horsefeathers.*

1 Introduction

In recent years, a basic building block of economics — the theory of individual decision making — has become the renewed focus of an enormous amount of research and reconsideration from many different perspectives. In decision theory, originally the center of such research, the insights of Allais (1953) and Ellsberg (1961, 2001), followed up by the work of Machina (1982), Schmeidler (1989), and Gilboa and Schmeidler (1989) led to a complete reconsideration of the classical notion of expected utility and subjective probability.[1] This research influenced and was influenced by the experimental and theoretical work of Kahneman and Tversky (1979) and many others in psychology who pushed for an even more fundamental reconsideration of how people make economic decisions. The latter research fed into work in economics by Laibson, Rabin, and others, who developed the behavioral approach to economics. In the same period, Rubinstein began the development of formal models which gave voice to the earlier insights of Simon (1955, 1982) regarding procedural rationality.

The ideas from outside of decision theory have reverberated back to decision theory again, leading to an explosion of research. Recent work has continued to develop our understanding of issues like ambiguity aversion (see Epstein and Schneider (2003), Klibanoff, Marinacci, and Mukerji (2005), Maccheroni, Marinacci, and Rustichini (2006), and Siniscalchi (2009), for example) and has gone further into a reconsideration of more fundamental notions such as optimization itself (see, for example, Kalai, Rubinstein, and Spiegler (2002), Manzini and Mariotti (2007), and Rubinstein and Salant (2008)), models of entirely new issues such as temptation (see Gul and Pesendorfer (2001)), and revisiting many older concerns such as incompleteness (Bewley (2002), Dubra, Maccheroni, and Ok (2004), and Eliaz and Ok (2006)) and regret (Hayashi (2008), Sarver (2008), and Stoye (2008)).

Given this large scale reconsideration, it is perhaps unsurprising that many economists, including but not limited to decision theorists, are rethinking what decision theory is, what it can do, and how we should think about it. For a sampling from the wide variety of critical perspectives, see Al–Najjar and Weinstein (forthcoming), Bernheim and

---

[1]For a detailed bibliography of the literature, see Wakker (2009). For an excellent introduction, see Kreps (1988).

Rangel (2007, 2009, forthcoming), Binmore (2008), Caplin and Schotter (2008), Fudenberg (2006), Gilboa (2009), Manski (2009), and the Northwestern workshop on "Decision Theory and its Discontents" (April 2009, proceedings forthcoming in *Economics and Philosophy*). Obviously, we cannot summarize such a vast set of ideas and arguments in a few pages. As should not be surprising given the range of opinions represented in these papers, we agree with some and strongly disagree with others of the views expressed therein.

Making an assessment of decision theory more difficult is the fact that it has never been an entirely unified field. Some work in decision theory has been motivated by essentially normative questions such as characterizing how a "rational" decision maker should act. Other work takes an entirely positive or predictive approach, seeking convenient modeling tools for summarizing real behavior, rational or not. Of course, there is an overlap with normative/prescriptive modeling — properties of behavior which we believe to be normatively desirable and relatively easy for people to do may also be sensible predictions.

In this paper, we discuss our views of decision theory and what it can contribute to economic theory. We almost entirely focus on positive theories, commenting briefly at the end on normative considerations.

In a nutshell, we argue that the goal of positive economic theory is to provide useful predictions and ways to understand the world, these goals being related but not identical. We distinguish between the "story" of a model and its predictions, arguing that the story need not be literally true for the model to be useful. However, confidence in the story of the model may give us more confidence in its predictions and so is not irrelevant. Also, the story affects our intuitions about the model and hence whether and how we use it.

Unfortunately, in seeking to develop a "good" model, one runs into some difficult tradeoffs, the resolution of which will typically depend on the specific goals of the modeler. For example, a theory which helps us understand one particular issue better will often necessarily oversimplify other issues. The net result may lower predictive power in the short run. On the other hand, as we come to understand the new issue better, we may be able to model it in ways which allow greater complexity on other dimensions, thus giving us greater predictive power than before.[2]

When it comes to normative theories, we need more than good prediction. We cannot avoid the much more difficult task of identifying, at least approximately, the motives for and effects of choice — a disturbing observation given how difficult prediction is.

---

[2]This is not to endorse a view that all problems should be dealt with in a single model, a view we argue against later.

In Section 2, we discuss the goals of positive economic theory generally and decision theory more specifically. Section 3 turns to a discussion of the kinds of results in decision theory and what they are intended to accomplish. In Section 4, we consider some of the implications of the perspective we offer. Section 5 concludes with a brief discussion of the difficulties faced by normative theories, particularly welfare questions.

We close this introduction with caveats. First, we make no claim that our comments reflect anyone's views other than our own. Second, in writing this paper and rereading what others have said on these topics, we've learned how difficult it is to make broad methodological statements clearly and succinctly. We apologize if we have unintentionally distorted anyone's views and hope that we have made our views clear enough to avoid misinterpretation. Third, many of the points we make have been made *many* times before, too many for us to provide anything close to complete citations. We err on the side of providing few or no specific cites on each of these points. Our hope is that our particular combination of these observations is sufficiently idiosyncratic to be worth the reader's time. Finally, our goal in this paper is to say something useful both to decision theorists and to those unfamiliar with decision theory. We hope each audience will be patient with our comments to the other.

2   Positive Decision Theory: Ends

Unfortunately, discussions of choices between models in economics often focus entirely on predictions and suggest that our choices between models should be based on refutation only. That is, we reject the model if and only if its predictions are falsified by the data, retaining it as "unrejected" in the meantime. We certainly agree that prediction is important, indeed central, to theory. On the other hand, we do not believe that the predictions are the only determinant of the usefulness of a model. In particular, the "story" the model tells about the world is also important.

As we see it, the purpose of positive economic theory is to provide tools and concepts, a *language*, which aids in understanding — that is, explaining what we observe — and prediction — that is, saying what we expect will happen at some future date or in some other situation.[3] To state our views more concretely, suppose empirical observations or our intuition about the world suggests the hypothesis that A leads to X where A and

---

[3]Much of the work in theory involves results which are primarily about understanding the language itself better with an indirect, long run view toward understanding the world with the language (e.g., equilibrium existence results). While the boundary between understanding the language and understanding the world is not always clear, we generally focus here on the latter. This should not be taken to imply a view that the latter is superior, only that we think its role is more central in the debate discussed in the introduction.

X are statements about observable variables.[4] We claim that having a model explaining why A leads to X would be useful above and beyond our conjecture or whatever empirical observations led to it. By a "model," we mean a precise, mathematical object which is our attempt to reflect an informal intuition we have about how the world works. Constructing such a model forces us to define a mathematical analog to A and X as well as to make various auxiliary hypotheses relating them. What do we learn from developing such a model?

First, we may learn things about whether our intuition is flawed or not. For example, we may find that our seemingly sensible intuition of why A leads to X actually requires an additional, unpalatable assumption of B, leading us to question whether our initial conjecture is sensible. Alternatively, perhaps the model also implies an unexpected and unpalatable conclusion Y or implies that there is an internal inconsistency in the logic that A would lead to X.[5] On the other hand, we may find that only relatively weak and plausible additional assumptions are required, reinforcing our initial intuition.

Second, we may flesh out our initial intuition in ways that enable additional or better predictions. For example, if we see that A also leads to an unexpected and plausible implication of Y, then we have a new prediction. Also, if we see that A only leads to X when some other condition B holds and that A has a different implication otherwise, then we can improve our ability to predict when A will lead to X since we now believe that B is required. If B is observable, the recognition of its role gives us a different prediction for those out of sample forecasts when A holds and B does not. For example, the initial discovery of revenue equivalence in independent private value auctions indicated that some other feature (the B) such as a common–value component is needed to explain the ubiquity of English auctions (the X).

Finally, even if the model doesn't immediately change or enlarge our set of predictions, it may yield a clearer understanding of why A might cause X. Why would such an understanding be useful? We see the primary value of such understanding in the way it may lead in the long run to more or better predictions. Lest this comment be misinterpreted, we emphasize that understanding may involve concepts whose translation into observables is not direct or immediate. Consequently, in the short run it may be difficult to know exactly what these predictions might be. Finally, we emphasize that "prediction" should be broadly understood to mean a perhaps imprecise statement about a relationship between observable variables. An imperfect example is Spence's (1973) paper on job market signaling. While he did provide the new prediction that, controlling

---

[4]Our arguments also naturally apply to other kinds of hypotheses. For example, we may instead hypothesize that A and X are correlated because of some (possibly unobservable) common cause Z.

[5]Paul Krugman gives a good illustration of this in his first Lionel Robbins lecture in June 2009: "What I did at the time, a very economist thing to do, was to build myself a little model to prove the point that I believed. So I built a little intertemporal optimizing whatever and to my shock — and this is the point, of course, of doing models — it actually gave me the opposite answer."

for human capital, workers with higher education would receive higher wages, the greater contribution of the paper seems to be in generating an understanding of signaling that ultimately led to many other predictions in other environments.

Thus in developing a model, we want it to have predictions about observable variables, either immediately or eventually. If we learn that these predictions are highly inaccurate, we should abandon the model. However, viewing model selection as only a matter of refutation leads to some incorrect conclusions. For one thing, "understanding" is a goal about eventual predictions, so we certainly cannot evaluate the accuracy of these predictions in the short run. But the problem with viewing refutation as our sole criteria for model selection has problems that go quite beyond this consideration.

For example, an exclusive focus on refutation leads to the conclusion that two models with the same set of predictions are "equivalent" and hence we cannot have any reason to prefer one over the other. By definition, two models with the same set of predictions are equivalent in terms of the conditions under which they are refuted. But we can still have strong reasons for favoring one model over the other. Many of these reasons have to do with the "story" of the model — the informal interpretation of the mathematics.

Consider, for example, the theory of subjective expected utility. The story of this model is that the agent forms subjective probability beliefs about whatever uncertainty he faces and chooses that action which maximizes expected utility as computed with these beliefs. To us, this story seems very idealized but intuitive at a certain basic level. While we don't believe this story is literally true, the idea that agents have beliefs of some kind which guide their choices, that they choose actions they consider are more likely to work out well, etc., seems very natural. Furthermore, even if we dismiss the story as plausible at all, it is a story that is very handy for our reasoning process. That is, it is useful to organize our thinking around ideas like beliefs, information, expectations, and the other concepts suggested by this model.

Since models don't simply sit on a shelf but are to be used, tractability is valuable. For example, a model may be taken to new domains to generate new predictions/understanding. While the story isn't the sole determinant of tractability, having a nice intuition about the story facilitates the use of the model. Similarly, a more intuitive model is likely to be more valuable in helping us understand the world.[6]

To see the point, suppose we had a black box which, when input a choice set, etc., would immediately tell us the agent's choice from that set. E.g., in a game, we could input the behavior of the opponents and it would give the response. Even if this predicts

---

[6]In line with our argument that understanding is about the eventual development of predictions, one could say that this second aspect is really a repetition of the first and/or referring to having a larger set of predictions in the long run.

perfectly, it's not clear how useful it is. For example, we couldn't use this to compute an optimal mechanism since without understanding the structure of the function describing this agent, we'd have to search over every possible mechanism. Hence even a flawed intuition might be more useful. In short, general principles, even ones which are not entirely accurate, may be more comprehensible and hence more useful than very detailed and accurate specifications. Thus we see that not only may there be reasons to select among models with the same predictions, these factors may lead one to prefer a model with *worse* predictions.

Once we consider comparisons of models with different predictions, we see another problem with viewing refutation as the only way to select among models. Economists have limited data and need to make predictions, not just to determine asymptotically whether the model is right, but because of the need to choose policies or other actions. When one recognizes the need to choose among models with different predictions before the relevant data arrives, one sees again the value of "intuition." As Kreps (1990) argues, consistency of the model with our intuition about how the world works can be thought of as another kind of consistency with data. If we had to choose between the predictions of two models that are equally consistent with the data we have, we would prefer to predict with the model that fits our intuition better.[7]

We note that this use of intuition is different from that emphasized earlier and may even conflict. Earlier, we argued that a more intuitive model would be easier to extend to new domains. If a model tells an intuitive story *which we know to be false*, the intuitiveness of the story can still be very valuable in this sense. For example, as argued above, a skeptic about the story of subjective expected utility may still find that the intuition of the story makes it easy to use the model. By contrast, here we refer to the case where a model tells a story which is intuitive in the sense that *it seems plausibly true*. While that story may be complicated in ways which make the model harder to work with, the plausibility of the story may enhance our confidence in the model's predictions.

To introduce a final concern with an exclusive focus on refutation in assessing models, we emphasize an obvious but important point: a model is *never* a completely accurate description of the world. Having a model which is exactly correct would require the model to be as complex as the world, something about as useful as a map with a scale of "1 inch = 1 inch."[8] Less facetiously, all understanding is generated by deciding which

---

[7]Gul and Pesendorfer (2008) argue forcefully that the implausibility of the story of a model cannot refute the model. We entirely agree. However, the implausibility may make us less confident in the predictions of the model.

[8]According to Wikipedia, this idea originated in Lewis Carroll's 1893 *Sylvie and Bruno Concluded*, though other readers may recognize it from Jorge Luis Borges' 1946 short story "Del Rigor en la Ciencia" ("The Rigors of Science"), Umberto Eco's essay "On the Impossibility of Drawing a Map of the Empire on a Scale of 1 to 1" in his 1995 book *How to Travel with a Salmon and Other Essays*, or a joke by Steven Wright.

aspects of reality to ignore. The world never repeats exactly the same situation twice, so if all aspects of reality are relevant, it is impossible to learn from one situation how to predict what will happen in another.

Thus almost every model is refuted if we define refutation in the strictest sense of the term. To see the implication of this, consider again our discussion of a model of above regarding how A might imply X. We noted that the model might tell us that this intuitive story, once formalized, has the additional, unexpected, prediction of Y. Suppose that Y is something absurd, violated either by repeated observation or simply intuitively implausible. Does this mean we reject the model?

Not necessarily. We know that our model is unrealistically simple. Hence we should not be surprised to find some odd implications popping up in the model — either odd predictions about other issues such as Y or excessively extreme predictions about the magnitude of the effect of A on X. The key question is whether we believe that the simplifications which led to the odd predictions are crucial to the model's explanation of why A leads to X. If so, we should conclude the model is inadequate in providing such an explanation, undermining our confidence in its predictions more broadly. Depending on whether we continue to believe our original hypothesis about the relationship between A and X, we would reconsider either the formalization of our idea or the idea itself. On the other hand, if the simplified aspects of the model which lead to implausible predictions are not important to our explanation for the relationship between A and X, then it seems reasonable to continue to use the model as a working hypothesis for formulating predictions about A and X, at least qualitatively.

In the end, choices between models will hinge on a number of considerations. In addition to consistency with data, we value intuitive appeal (in both senses discussed above), tractability, and the range of additional predictions/understanding the model generates, several of which depend at least in part on the story of the model. Hence the choice of a model will depend on the purpose the model is used for, the modeler's intuition, and the modeler's subjective judgement of plausibility. Since all of these things vary across economists and between economists and psychologists, it should not surprise us to see different models chosen by different social scientists, especially if they wish to focus on different aspects of the issue at hand.[9] To be clear, we agree with the importance of minimizing the role of subjective judgements to the extent possible and are not arguing that all intuitions are equally valid. One economist may reject another's intuition and, ultimately, the marketplace of ideas will make some judgements. (Given the amount of market failure in this context, perhaps a sociological notion is more apt than our economic metaphor.)

---

[9]We are not experts on the philosophy of science, but our understanding is that the unavoidability of subjective judgements in science is widely acknowledged there.

Now we specialize our discussion by turning to decision theory more specifically. Before explaining the motivation for the approach decision theory takes, we first explain what it does. Decision theory provides a formalization of certain intuitive ideas about decision making and relates these to potentially observable choice behavior. Typically, decision theory develops a "model" in the form of a utility function (or a class of utility functions, e.g., expected utility) which is used to predict choice behavior.[10]

For example, suppose we want to model agents who suffer from "temptation." The first step is to identify the behavior that we think corresponds to our intuitive notion of temptation. This step, giving "temptation" a behavioral definition, is in many ways the crux of what decision theory does — see Section 4.3 for more on this point. Since it would be too difficult to work with a model that assumes only this behavioral definition, we add some assumptions on behavior that simplify and either seem plausibly true or at least plausibly unrelated to the issue of temptation. We then relate these assumptions to a particular functional form for the utility function. This result thus identifies a functional form that we can think of as a useful formal embodiment of our intuitive notion of temptation.

Decision theory is used for a range of purposes, from predictions about individual choice (as in demand theory) to being an ingredient in a much larger model. For an example focused on individual choice, we might use the model of temptation to analyze how demand functions are affected by temptation. For an example of using decision theory as an ingredient, suppose we observe that the addition of dessert items by a restaurant leads to a short–run increase in sales, followed by a severe downturn. Suppose we hypothesize that people are surprised by the tempting items at first and give in to temptation, but afterward avoid these restaurants. Then we can use our model of temptation as an ingredient in analyzing this question.

To relate decision theory to our earlier discussion of prediction and our "A implies X" story most simply, think of the A as some statement of the choice situation an agent faces — for example, the set of available options. Think of the X as a statement about the choice made — for example, that the agent purchases insurance. Then much of our discussion of prediction above carries over directly to decision theory. Just as in the general case, we may wish to understand why the agent purchases insurance. A trivial explanation is that he just likes to buy insurance policies, but a fuller and therefore

---

[10]For simplicity, throughout this paper, we will describe a model in decision theory in terms of utility functions. However, classic models in decision theory also include other objects such as subjective probabilities. Recent models have introduced many other approaches to representing behavior. For example, the agent is represented with more than one utility function in the literature on temptation (Gul and Pesendorfer (2001), Dekel, Lipman, and Rustichini (2009)) and incomplete preferences (Dubra, Maccheroni, and Ok (2004)) and with a set of probabilities in the literature on ambiguity (Bewley (2002), Gilboa and Schmeidler (1989)). Another type of model involves criteria which eliminate certain options before applying utility maximization (Eliaz and Spiegler (2008), Manzini and Mariotti (2007)).

more appealing explanation is that insurance reduces risk and the agent values it for this reason. We then employ decision theory to define notions of risk and risk aversion in order to formalize this explanation. This can be useful for many reasons. For example, our intuition about why the agent buys insurance may lead us to make other predictions about his behavior — e.g., investment decisions. More generally, we take "axioms" or properties of his choice behavior in a certain domain, assume they hold more broadly, and use this to predict choices on a larger domain. As in our discussion of positive economics, the model may lead to predictions about other variables, show unexpected problems with our reasoning, etc.

As we discussed earlier, what data leads us to reject a model is a subtle question. To explain the point, return to the insurance example above. We suggested that an intuitive explanation of the observation of an agent purchasing insurance is that he is averse to risk and buys insurance to reduce his risk. We might then adopt the simplest approach to formalizing this explanation, assuming that we can write utility as a function of money only, assuming the agent maximizes expected utility, defining risk and risk aversion in this context, and assuming that the risk aversion property holds globally. This yields a simple, intuitive model with many predictions. Suppose that our goal with this model is to make predictions about the agent's other insurance purchases. For example, perhaps our initial observation concerned the purchase of house insurance and we wish to predict choices regarding car and boat insurance. If our predictions of these decisions are reasonably accurate, then we should not be too troubled by the observation that the agent occasionally plays poker with his friends, even though this contradicts the joint hypothesis that utility depends only on money and that the agent is globally risk averse. We would argue this is not troubling for two reasons. First, we know that our hypothesis that the agent cares only about the amount of money he has is a simplification and that we could easily change that assumption to incorporate the utility of games with friends to reconcile the model with this observation. Second, the poker games were not our focus in constructing the model, so inaccuracy in that dimension is not costly. Thus we would not bother to address this observation.[11]

Similarly, if we have data which refutes an auxiliary assumption of the model, this need not be a reason to abandon the model. For example, suppose we have evidence that the agents whose behavior we are studying do not maximize expected utility, but the model still does a good job of predicting insurance decisions. Since this data does not refute the predictions of the model, we see no reason to abandon it, though this may shake our confidence in its predictions. Of course, our confidence in the predictions would be enhanced if we knew it were possible to modify the expected utility model to generate the same predictions about insurance and avoid refutation by other data.

---

[11]Of course, if the original model is consistent with the agent's poker playing, this is only mild support for the model for the same reasons.

A perhaps more controversial point is that even if we are initially making poor predictions about insurance choices, this might not lead us to abandon our general explanation based on risk aversion, though it forces us to revise our specific model. For example, if we used expected utility and we find that choices of insurance similar to those generating the Allais paradox are causing our model to mispredict, we may find that we can switch to a generalized expected utility model a la Machina (1982) with risk aversion and predict quite well. The observation of Allais type behavior leads us to drop the expected utility aspect of our model, but if our main idea was about risk aversion, not expected utility, this change is not saying the original insight was "wrong."

As this discussion highlights, part of the problem with a focus on refutation is that almost any phenomenon can be explained in almost any broadly defined class of models by "tweaking" the model appropriately. If we can define the set of factors which influence the utility of the agent any way we like, can define how the objects in the world map into these factors any way we like, and can choose any functional form we like, surely we can rationalize any set of data.

So when would we abandon our explanation of why A leads to X? The point is that if we have to do a huge number and variety of tweaks, then the resulting model loses all the properties we said were valuable. It ceases to be tractable. It ceases to be intuitive. Unless our new situation is very close to the data our model has been constructed to fit, we lose the ability to make out of sample predictions with any degree of confidence. At this point, we conclude that we need a better model, presumably one based on a completely different explanation of the behavior we observe.

To be clear, it is valuable when a single model is consistent with a wide variety of empirical observations. However, when a model achieves this consistency only by adding a tweak or an additional parameter for each data point, then the model adds no understanding and is unlikely to predict well (except in situations that are very close to our past observations). The only valuable version of a model with such breadth is one which explains a variety of observations with a small number of basic principles.[12]

3   Positive Decision Theory: Means

Broadly speaking, there are two branches of work in positive decision theory. One group of papers (see, for example, Gollier (2001) and the references therein) develops new tools, e.g., different measures of risk, for working with existing models. The other, which we focus on henceforth, develops new models to introduce new considerations to economics. For example, Schmeidler (1989) and Gilboa–Schmeidler (1989) introduced the notion of

---

[12]This is reminiscent of the way Copernicus' model of astronomy replaced the Ptolemaic model.

ambiguity aversion where agents prefer risks based on known probabilities to those based on unknown, a phenomenon impossible in the usual subjective expected utility model.

The usual approach in the development of new models is not to start from scratch but instead to generalize existing models to allow the new possibility under consideration. As discussed above, one identifies the behavior that is intuitively associated with the new issue to be introduced. Since the new behavior is, by hypothesis, inconsistent with standard models, this requires modifying some of the standard assumptions. Typically, one drops as little of the standard assumptions as possible, retaining the rest of the standard assumptions for the purposes of simplification as discussed earlier. This is also useful because it makes it easier to connect the new ideas to the existing literature to see how the model with the "new" ingredient changes our understanding from what we had before.

A related methodological consideration is that we do not consider multiple new ingredients at the same time. For example, we have mentioned temptation and ambiguity aversion as new ingredients that have been added in recent years. Yet the models which introduced temptation did not include the previously added ingredient of ambiguity aversion. Why not? The point is that these seem like conceptually distinct issues. If so, then the simplest way to achieve understanding of these two issues would be to first understand each in isolation. If we realize some connection between the two issues, perhaps because some phenomenon of interest stems from the interaction of them, then and only then would a model that combines these new ingredients be of interest.

Kreps (1990) provides a useful analogy (which he attributes to Jose Scheinkman). He notes the usual joke about how the economist looks for his keys under the streetlight not because that's where he lost them but because "the light is better." He suggests that we should think instead of progress in theory as putting together a series of lights from where the light is good already to where we think the keys are. We agree, but would extend the analogy. The point is that we also need to consider what direction to move the light. I.e., we begin by moving in the direction of adding temptation *or* in the direction of adding ambiguity aversion but not both together. If moving in one direction works, we'll "find the keys." If not, this means there is an interaction between these issues that is important and we'll have to learn to understand that interaction and move in that direction to find the keys.

Turning to specifics, most theorems in decision theory are about the following issues:

1. What is the behavior corresponding to a particular functional form?

2. To what extent does the behavior in question identify parameters in the function?

3. How do intuitive changes in the behavior correspond to changes in the parameters?

11

The first kind of result is called a representation theorem since it shows that a particular functional form represents (corresponds to) certain behavior. We refer to the second kind of result, for obvious reasons, as an identification theorem or a uniqueness result. Finally, we refer to the last kind of result as a comparative or comparative static.

3.1   Representation Theorems

A representation theorem relates a model of decision making (a class of utility functions) to properties of the implied choices. The latter is usually stated in one of two forms. First, it may be stated as a choice function or correspondence — that is, a specification of the item(s) the agent would choose as a function of the feasible set. Second, it may be stated as a preference relation — that is, a statement about which object (if either) the agent would prefer from any given pair of objects.[13] For our purposes, positive decision theory, the notion of preference should be interpreted as *revealed preference*. In other words, a is preferred to b is shorthand for a would be chosen if the available options are a and b. (Unfortunately, many of the motivations offered for axioms in the literature implicitly equate revealed preference with "real" preference.)

A representation theorem says that some utility function in a particular class predicts the agent's choices if and only if those choices satisfy certain properties. These properties are typically referred to as *axioms* and so representation theorems are often called *axiomatizations*.[14]

What is the purpose of a representation theorem? In a nutshell, the purpose is to clarify what the model — the functional form for a utility function — predicts. While we find the story of a representation very useful, it does not tell us what the model predicts. For example, as discussed in the last section, it is common in economic theory to assume that an agent has subjective probability beliefs and maximizes expected utility given them. One interpretation of the model takes this literally: agents have probability beliefs and compute expected utilities with them to make decisions. Since people often find this difficult to do and rarely seem to explicitly follow such a procedure, some skeptics argue that this theory is simply wrong.

However, as we discussed, the predictions of positive decision theory are predictions of choice behavior. While the story's plausibility (or lack thereof) may affect our confidence in the predictions of the model, it cannot refute or confirm those predictions.[15]   The

---

[13]This is simply a partial specification of a choice correspondence.

[14]The word "axiom" is somewhat inappropriate for positive models in that it suggests a universally true principle.

[15]There is an analogy here to our earlier argument that data refuting an auxiliary assumption in a model does not necessarily refute the predictions of the model and hence does not refute the model. If we

contribution of Savage (1954) was to identify the predictions by demonstrating that an agent whose behavior satisfies certain natural and easily interpretable properties would behave as if she has probability beliefs and computed expected utility.

One reason this is useful is that it can help us find and/or understand examples that show the limitations of the model. As Ellsberg (1961, page 646) put it, one effect of Savage's axioms is that they gave "a useful operational meaning to the proposition that people do *not* always assign, or act 'as though' they assigned, probabilities to uncertain events." In other words, the identification of the behavior that corresponds to subjective expected utility also identifies the behavior that does *not* correspond to the theory. By thinking about the limitations of Savage's axioms, Ellsberg (1961) was led to identify an example of natural and plausible behavior which was inconsistent with subjective expected utility.[16] By doing so, Ellsberg identified both a set of circumstances under which Savage's theory is unlikely to predict well and pointed to the behavior which alternative models would need to address.

Relatedly, if we want to test a model, it may be easier to test behavioral implications directly rather than to estimate a general functional form representation and see if the estimated function is in the class corresponding to the model. This is particularly true if we want to be sure to test the parts of the model we take seriously, not the parts we view as a convenient but inaccurate simplification. If we reject a functional form, it can be difficult to say whether this is because of the inaccuracies of the convenient simplification or the essence of the model. If we directly test the key behavioral predictions instead, we don't have this problem.

Another illustration of the value of a representation theorem is provided by Masatlioglu and Ok (2009). The classic loss aversion and reference dependence model discussed in Kahneman and Tversky (1979) and developed further in Tversky and Kahneman (1991, 1992) was motivated in part by the idea of the endowment effect where the agent is biased in favor of the status quo. Specifically, Tversky and Kahenman (1991) note the evidence of Samuelson and Zeckhauser (1988) and the property they term status quo bias: if an agent is indifferent between x and y when z is the status quo, then he should strictly prefer x to y when x is the status quo. However, Masatlioglu and Ok (2009) show by example that a commonly used version of the functional form proposed by Tversky and Kahenman (1992) used does not satisfy this property.

We do not mean to argue that this property is necessary for a model of reference

---

think of the procedural story of subjective expected utility as analogous to an "auxiliary assumption," we see that the failure of this assumption does not refute the theory. On the other hand, as we argued before, such refutation may lessen one's confidence in the predictions of the model.

[16]To be fair, Ellsberg did not immediately recognize the sure–thing principle as the key to the issue. Indeed, he comments (Ellsberg (2001), footnote 2, page 244) that he was "considerably surprised" to realize this was the axiom he was questioning.

dependence to be interesting. Instead, our point is simply that without a representation theorem that clearly identifies the behavior implied by the model and how it relates to the properties the model is intended to capture, a researcher who adopts the model with the goal of generating the relevant behavior would not know whether he succeeded.

The discussion so far only demonstrates the role of *necessity*, not *sufficiency*. A representation theorem proves that behavior satisfies certain properties if and only if it is the result of maximizing a utility function in a certain class. Thus these properties of behavior are necessary and sufficient for the representation. Typically, proving the behavioral properties are necessary (implied by the representation) is very easy — it is not unusual for this part of a proof to be omitted because it is trivial. On the other hand, the sufficiency part of the proof (showing that the behavioral properties imply the representation) is often quite difficult.

Yet the justifications above really only show value to the easy part. To see this, suppose we have a property which we only know is necessary for the functional form and don't know whether it is sufficient. If we observe behavior inconsistent with this necessary property, then we reject the model. Similarly, if we understand the intuitive limitations of a necessary property, again, this may enable us to identify the limitations of a model. That is, recognizing that the property conflicting with Ellsberg's example is the sure–thing principle requires only knowing that the sure–thing principle is necessary. Finally, it would be enough for our hypothetical researcher to know that a particular version of the Tversky–Kahneman model implied the status quo bias property.

The sufficiency part of a representation theorem is important for several reasons. First, obviously, if we don't know *all* of the implications of a model, we don't know if we are seeing behavior consistent with it. For example, if we only knew that expected utility requires that preferences be transitive and continuous but did not realize it requires the independence axiom, we might fail to realize that the Allais paradox behavior contradicts expected utility theory. Without knowing that the behavior we have identified is sufficient, we can never know whether there is a necessary property, perhaps a very undesirable one, that we have failed to identify.[17]

The importance of knowing we have all the implications is particularly clear when the story of the model is potentially misleading about its predictions. For example, in the multiple priors model of Gilboa and Schmeidler (1989), the agent has a set of probability beliefs instead of a single belief. In evaluating a given course of action, he uses the "worst

---

[17]On the other hand, it is worth noting that even if we know the necessary and sufficient conditions for a representation, we may miss important behavioral implications. There could easily be some subtle implication of a combination of two or more axioms that we do not recognize simply by examining the axioms in isolation. There are infinitely many ways to express the behavioral implications of any model and part of the art of axiomatization is bringing out the key implications in the clearest possible way. See Section 4.3.

case" expected utility across the beliefs in his set. Thus the model seems to describe an extraordinarily pessimistic agent. Yet the axioms that characterize behavior in this model do not have this feature. Instead, the key axiom just says that the agent only satisfies the independence axiom on a restricted domain, an axiom we find quite natural. The sufficiency theorem ensures that there is not some hidden pessimism requirement that we haven't identified yet. Similarly, the minimax regret model assumes that agents derive disutility from choosing an action and later realizing that a different action would have been better. Thus the story of the model seems to require the agent to know he will later learn which action would have been best. Yet the axiomatizations of Milnor (1954), Stoye (2008), or Hayashi (2008) identify very natural properties of choice which do not require such ex post observation in any way. See Schlag (2006) for further discussion.[18]

Second, recall our discussion of developing a model of temptation in Section 2. We noted that one proposes a behavioral definition of the phenomenon of interest (temptation in the example), adds some convenient simplifying assumptions, and derives a model which corresponds to that behavior. This development is entirely about the sufficiency part of a representation proof. It is the sufficiency argument which enables us to say that the representation is the more tractable embodiment of our notion of temptation.

Third, sufficiency clarifies the links between observables. To see the idea, consider the Koszegi and Rabin (2006) model of reference dependence which is based on the Kahneman–Tversky model of loss aversion. Koszegi and Rabin extend the model to lotteries and show that the properties identified by Bowman, Minehart, and Rabin (1999) imply a weak form of status quo bias (if x is preferred to y when y is the status quo, then x is preferred to y when x is the status quo). They also use the model in some economic contexts to demonstrate other results such as a novel explanation for sales. In other words, a particular representation implies the weak status quo bias and also implies the role for sales. Does this mean that weak status quo bias is a key element of their explanation for sales? Without a sufficiency result, it is impossible to say. We don't know whether the weak status quo bias (with some auxiliary assumptions) implies the role for sales or conversely or perhaps both follow from a third behaviorally identified property. By contrast, suppose there was a sufficiency theorem for this model. Then we would know that weak status quo bias (plus some other auxiliary assumptions) implies this representation of the agent which implies the role for sales. In this sense, we would know that weak status quo bias is the key to the result.[19]

---

[18]Interestingly, Sarver's (2008) axiomatization of a different regret model seems (to us) most naturally interpreted as assuming ex post observation.

[19]This argument does not imply that the sufficiency analysis must be in the form of a representation theorem. If one takes the basic loss aversion model as the auxiliary hypothesis and proves that weak status quo bias holds if (and not just only if as in Koszegi–Rabin) the functional form restrictions of Bowman, Minehart, and Rabin are satisfied, then one can make the same argument that a representation theorem would allow. On the other hand, we find auxiliary assumptions in the form of axioms easier to assess than auxiliary assumptions of a functional form.

Finally, since we do not have the data to distinguish models perfectly, one way to enhance our trust in a model is to see that its results are robust. That is, if we find that a particular conclusion proven for one class of models also holds when we enlarge the set of allowed models, then we have more confidence in this conclusion. Of course, our confidence is enlarged more if the enlarged set of models is "bigger." Conversely, if we find that a "small" change in a model leads some conclusions to fail, then we are less confident in those conclusions. Clearly, to consider such issues, we need some way to measure the size of changes in models. Since changes in models typically involve non–nested changes in functional forms, there is rarely a parameter whose change measures the change in the model. Even when models are nested, it is hard to know how behaviorally significant a given change in a parameter really is. If we know the behavior in question, though, we can assess this. For example, suppose we learn that a conclusion which holds when preferences satisfy the independence axiom continues to hold when we relax independence to betweenness (Chew (1983), Dekel (1986)). That is, the property does not require both parallel and linear indifference curves, just the linearity. This gives us a clear sense of the degree of robustness being demonstrated.[20]

We are not arguing that representation theorems are the only tool for achieving a useful understanding of our models. In our view, behavioral economics and decision theory are different approaches with similar but not identical purposes. The work we find most interesting in both areas notes some observations that are difficult or impossible to explain with existing models, proposes some alternative model which has an interesting interpretation in light of the observations, and then provides some means to understand the alternative. The means and ends differ and we see advantages to each approach.

Starting with what we prefer about the decision theorist's approach, it generally takes the form of a complete characterization of the model, clarifying both its general predictions and the nature of the simplifications used. By contrast, behavioral economists typically omit any analog of the sufficiency results we argued above are important. Indeed, many papers don't provide much in the way of general necessary conditions either, making it hard to see what the model says about individual choice aside from the stories one can tell about the representation. Decision theorists typically tell stories about their representations as well, but provide clarification via a representation theorem. There are notable exceptions to this critique. Koszegi–Rabin (2006)[21] and Eliaz–Spiegler (2008), for example, show several interesting implications (i.e., necessary conditions) of the general functional form. We wish more papers in behavioral economics did likewise.[22]

---

[20]To be fair, it is sometimes possible to study robustness via the functional form as in Machina (1982).

[21]Some of Koszegi and Rabin's implications are about observable variables only if we assume the reference point is observable. In their model, reference points are derived endogenously and equal the agent's expected choice, so they are potentially observable.

[22]See Spiegler (2008) for a similar argument in favor of a more systematic exploration of the implications of new models.

In favor of the behavioral approach, it clarifies our understanding of the new model via some exploration of what the model does in a simple economic environment. By contrast, the decision theorist's analysis is typically carried out in a simpler context than the setting in which we hope to use the model. For example, one normally considers axioms in a single–agent setting and often only in the static case, even though many of the applications of interest have many agents and important dynamic considerations. The behaviorial economist's use of less stark environments means that the economic implications of the model are often more clearly developed than in typical papers in decision theory.

Thus we reject the argument that either approach is inherently superior to the other. Indeed, we think these approaches should be seen as complements, not substitutes. While we have generally taken one of these approaches in our own work, this reflects a combination of historical accidents and comparative advantage (or comparative disadvantage as the case may be), not a view that only one approach is legitimate.

Relatedly, representation theorems are not always essential. While we see representation theorems as a valuable way to understand models better, we could find a model worth exploring further even before it has been axiomatized. Also, there are cases where the behavioral meaning of the model is sufficiently obvious that an explicit representation theorem adds little. For example, in a model of utility of money, the assumption that $u' > 0$ has a very obvious meaning without having to restate this in the language of preferences. Similarly, as discussed above, much of behavioral economics offers a functional form instead of axioms. In some cases, the translation of this functional form into the behavior it predicts is sufficiently obvious that this does not seem unreasonable. In such cases, we can think of the functional form itself as the axiom, so a representation theorem is not needed. In other cases, it is very difficult to see what this "axiom" says, so we would prefer "real" axioms.

Not only is axiomatization not necessary for a model to be interesting, it is certainly *not* sufficient. A model that is intractable, complex, unintuitive, and makes lousy predictions may have an axiomatization, but we wouldn't want to use the model anyway.

3.2   Identification Theorems and Comparatives

One point which seems not well understood outside the decision theory community (and perhaps not inside some parts of it either) is the importance of identification. A model is supposed to be an intuitive and simplified description of some aspect of reality. The purpose of decision theory is to understand this model and the behavior it predicts/describes. One key to doing so is to determine the extent to which the objects in the representation can be pinned down from the agent's behavior and, in this sense, given behavioral

meaning. If the objects in the representation cannot be given such meaning, then the model is, at best, loosely connected to what it is supposed to be describing and, at worst, misleading.

For example, consider state–dependent expected utility. This is a model where the utility function depends on what is consumed but also on the state of the world. It is well–known that we cannot identify subjective probabilities separately from utilities in such models. To see this, suppose there are two states, s_1 and s_2. The agent is represented via an evaluation of *acts*, functions which say what the agent gets in states s_1 and s_2. Suppose the agent's preferences over such acts can be represented by $(1/4)u(f(s_1), s_1) + (3/4)u(f(s_2), s_2)$. That is, if the agent chooses act f, he is interpreted as receiving outcome $f(s_1)$ in state s_1, generating utility $u(f(s_1), s_1)$ and analogously in state s_2. His subjective probabilities are that state s_1 has probability $1/4$, while s_2 has probability $3/4$. It is easy to see that the choices predicted by this model would be the same as those predicted by the model where the agent maximizes

$$\frac{1}{2}\left[\frac{1}{2}u(f(s_1), s_1)\right] + \frac{1}{2}\left[\frac{3}{2}u(f(s_2), s_2)\right]$$

since this is another way of writing the same function. Define $v(x, s_1) = (1/2)u(x, s_1)$ and $v(x, s_2) = (3/2)u(x, s_2)$. Then we are saying that the agent with utility function u with subjective probabilities $(1/4, 3/4)$ is indistinguishable from an agent with utility function v and subjective probabilities $(1/2, 1/2)$. Intuitively, if we see that the agent pays more attention to outcomes in one state than outcomes in the other, we can't tell if this is because his utility function is more sensitive in that state or because he considers that state more likely. Thus, as in the $u + r$ example, the issue is that we cannot behaviorally identify one object separately from the other.

The usual approach for dealing with objects that cannot be behaviorally separated is to adopt a normalization for one object that yields (conditional) identification for the other. In the case of state–dependent utilities, we could normalize the probabilities to a uniform distribution or, equivalently, simply omit the probabilities altogether and work with the sum of the utilities. Note that this normalization means that the probabilities then do not have any meaning — it cannot be significant to say that the probabilities are uniform when we have to normalized them to be so. Other normalizations leave some flexibility for both objects, so both can have some meaning.

For an example of such a normalization, one which is very useful, consider the state–independent model. This is a case where an additional behavioral property, called state independence by Anscombe and Aumann (1963), implies that there is a normalization available where the utility function u is independent of the state. This is still a normalization in the sense that there will be state–dependent models that are equivalent. In particular, we have not proved that the utilities are state–independent; we have normal-

ized to make them so.[23] However, this is a normalization which has a nice intuition and is useful.

To see how this can be useful, suppose we have a representation with a state independent u and we know that the agent is weakly risk averse. One can define this from the preferences but for brevity, we focus on an equivalent statement in terms of u — namely, that u is weakly concave.[24] A standard result for this kind of model implies the following. If we compare two acts f and g such that the distribution of outcomes under f (given the subjective probabilities) is a mean–preserving spread of the distribution under g, then it must be true that g is weakly preferred to f, denoted $g \succeq f$.

Suppose, however, that we take a different normalization. To see the idea, suppose our initial state independent model has two states, a utility function of $u(x) = x$, and probability beliefs $(1/4, 3/4)$. Exactly the argument we gave earlier says that this agent's behavior would be the same as the behavior of an agent with state–dependent utility function $v(x, s_1) = (1/2)x$ and $v(x, s_2) = (3/2)x$ and probabilities $(1/2, 1/2)$. Consider the acts $f = (1, 1)$ and $g = (0, 2)$. Given the belief $(1/2, 1/2)$ that we attribute to the agent in this state–dependent representation, g is a mean–preserving spread of f — it has the same expected value but is not constant and hence is riskier. However, both representations agree on the conclusion that g has strictly higher expected utility than f. Thus the useful property of the state–independent model doesn't hold for an arbitrarily selected but equivalent state–dependent model. Hence the "normalization" of state independence strikes us as a natural one to adopt. As we argue later, even if there was some sense in which the "true" representation was not the state independent one, we'd still prefer this normalization.[25]

Probably the most familiar illustration of identification is when a utility function is said to be unique up to a monotonic transformation. To see how this relates to our explanation of identification, note that a utility function can only behaviorally identify when one object is ranked higher than another. Thus the determination of which item has higher utility is identified, but the magnitude of the utilities is not. Since a monotonic transformation preserves comparisons but not magnitudes, the "up to a monotonic transformation" identification tells us exactly what is behaviorally meaningful. Many

---

[23]This point is made forcefully by Karni (2007), though he takes a very different perspective on the issue.

[24]The equivalent statement in terms of preferences that we have in mind is for the Anscombe–Aumann model where the set of "prizes" is $\mathbf{R}$, interpreted as money. We can write any given act f as (f_s, f_{-s}) where f_s is the lottery over money in state s and f_{-s} is the profile of such lotteries for other states. Suppose that for every s, $(f_s, f_{-s}) \preceq (\mathrm{E}f_s, f_{-s})$ for every nondegenerate lottery f_s.

[25]It is worth noting that one can have models where the Anscombe–Aumann state independence property fails and there is no normalization of utilities for which this result holds. In this sense, even though we can identify probabilities conditional on a normalization, the identification is not as useful in inherently state–dependent models.

identification statements use this approach of characterizing uniqueness up to an appropriate transformation.

One of the key reasons we care about identification is that this is crucial to making comparatives possible. A comparative or comparative static is a result connecting a change in the representation to a change in the behavior it represents. To see the link to identification, note that identification results tell us the extent to which objects in the representation are behaviorally meaningful. Once we understand this, we can see how changes in those objects translate into changes in behavior.

Comparative statics enable us to obtain results connecting behavioral properties (e.g., attitudes to risk or ambiguity) to observables of interest (e.g., investment or insurance decisions). Much of the use of decision theory in economics is either directly of this form or an indirect use where one shows that a particular phenomenon implies a conclusion which does not hold in its absence.[26] A well–known example is the Arrow–Pratt (Arrow (1971), Pratt (1964)) measure of risk aversion.

One reason we think people sometimes overlook or underrate the importance of identification is that achieving identification is not costless. It often requires additional structure and/or additional assumptions. For example, if the domain a particular utility function is defined over is too small, then one may not have enough information to identify aspects of the utility function. Thus it is not unusual for decision theory papers to consider preferences over large and sometimes complex sets of objects. Relatedly, to develop the kind of structure needed for identification often requires stronger assumptions. Naturally, it seems better to have fewer assumptions and/or to focus on preferences over simpler and presumably more realistic domains.

As we see it, the key question in whether the use of a particular domain is appropriate comes down to whether the objects in the domain can be thought of as reasonable approximations of some actual objects in the world that the agent may consider. For example, in the Savage model, the objects of choice are functions from an infinite set of states of the world to an infinite set of consequences. We don't see real people often making choices between such functions. On the other hand, we do see people viewing options they have in the form of "if x happens, then I'll end up with a, but if y happens, I'll get b." The functions in Savage are just a more formal version of this and so seem reasonable to us. In proposing a large space for use in identification, we think the burden is on the decision theorist to point to such a correspondence.

When multiple models can be used to represent the same behavior, we have a different kind of identification issue.[27] When two different models correspond to intersecting sets

---

[26]The indirect use is also a comparative static in that moving from the absence of the phenomenon to its presence is equivalent to moving from a "zero level" to a positive one.

[27]On the other hand, multiple models can be thought of as a special case of an unidentified parameter

20

of behaviors, this should not be a surprise. Intuitively, a given model is like a particular explanation of the behavior it generates. It is unsurprising that at least some choice behaviors may have multiple explanations.

Furthermore, multiple explanations may be useful. Different explanations will suggest different intuitions, different questions to consider, different comparatives that might be useful. For example, we have a model of temptation (Dekel and Lipman (2007)) where the agent is uncertain about how/whether he will be tempted in the future and believes his tempted future self will always give in to temptation. We show that the set of preferences for partial commitments that this model can generate include all preferences generated by Gul and Pesendorfer's (2001) temptation model — where the agent knows his future self's level of temptation and may expect his future self to exert costly self control.[28]

To see the idea, consider a dieting agent who can commit himself to a healthy dish h, commit himself to a fattening dish f, or leave open the choice between the two, denoted $\{h, f\}$. Then it seems natural that we might have $\{h\} \succ \{h, f\} \succ \{f\}$ where $\{h\}$ denotes committing to h and $\{f\}$ is defined similarly. In Gul and Pesendorfer, this preference is interpreted as saying that the agent knows that if he does not commit to one of the two dishes, then he will be tempted by f but will manage to consume h at the cost of exerting self–control, thus leaving him in between the two commitments in utility terms. In Dekel and Lipman, this preference is interpreted as saying that in the absence of commitment, the agent gives some probability to sticking with h and some probability of being tempted away to f, thus ending up in between the other two options in terms of expected utility.

That different stories about temptation can overlap on at least some predictions should surely not surprise us. Yet these two stories are quite different and so suggest different applications or questions. For example, Gul and Pesendorfer focus on how we can identify the self–control costs and how higher costs would affect behavior, while Dekel and Lipman focus on the degree of uncertainty about future temptations. Thus we do not see this overlap as saying anything about which model is more useful or "better."

4   Implications for How to Do Decision Theory

The perspective on decision theory we have offered leads to a number of conclusions about how we should do decision theory, some of which are not entirely standard. In this

---

in a single model. We can always put two different models into a single class by defining the functional form of α times a function in one class plus $1 - \alpha$ times a function in the other. In cases where there is a function in each class representing the behavior, the α is unidentified.

[28]In a richer behavioral domain involving choices *under* partial commitment instead of choices *of* such commitments, the predictions of the models would differ.

section, we discuss several such issues.

4.1 The Role of Psychology

Some economists seem to reject the notion that psychology could be useful to economics and decision theory. In many cases, this view revolves around the distinction between studying the choices people make versus the way people make them. Many economists, notably Gul and Pesendorfer, have argued that we are only interested in the former. Others, such as Rubinstein (1998), suggest that the study of choice procedures may yield better understanding of choice behavior. Yet others, such as Camerer (2008), argue that while choice, not procedure, is the economist's traditional focus, this was an unfortunate compromise forced on the profession in an era when there was no way to observe how people make choices. These economists argue that with the advent of neuroeconomics, we can study the brain and its processing of choice problems.

We think that what psychologists are interested in is not always the same as what economists are interested in, so that it is not necessarily useful to blend the two perspectives. On the other hand, in our view, the real question is whether psychology can help us generate insights that enable more or better predictions (about variables of interest to economists) or to do so more easily.[29]

To begin thinking about this issue, consider again the theory of risk aversion. As discussed above, the basic idea of taking the concavity of u to represent a property we call risk aversion seems extremely useful in economic theory. Yet this is very different from the way a psychologist might think about risk aversion. As discussed by Gul and Pesendorfer (2008), a psychologist (or a neuroeconomist) might focus more on cognitive or emotional factors, "fear responses," etc. Thus a psychologist might say that the curvature of u is unrelated to "true" risk aversion but is instead related to diminishing marginal utility of money. In our view the fact that the psychologist sees the economist's model as wrong hardly trumps all other considerations by itself. Instead, the key questions are whether the psychologist's model makes more or better predictions (about the variables of interest to economists), whether it is easier to work with, etc.

For a more concrete example, recall the second example in Section 3.2 where a preference could be represented using a state independent utility function and probability beliefs of $(1/4, 3/4)$ or a state dependent utility function and probability beliefs of $(1/2, 1/2)$. As we noted, the state independent representation has the nice feature that we can relate the curvature of the utility function to a theorem on mean–preserving spreads and we do

---

[29]We suspect that most economists agree with this point, though they may disagree in evaluating whether a given paper is successful at this.

not get this theorem for the equivalent state–dependent model. Suppose a psychologist could measure "utils" and told us that the person's utility function was different in the two states, that he genuinely valued consumption in one state three times as much as consumption in the other state, implying that the person's "true" probability beliefs were the ones in the state dependent representation, namely $(1/2, 1/2)$. We would argue that we should ignore the psychological data here since making our model more psychologically realistic comes at the cost of losing convenient tools. This example illustrates that psychological realism is not costless, so the relevant consideration should be whether the costs exceed the benefits.

So what are the potential benefits of psychology? As we have argued throughout, all models try to take some observations or intuitions about reality, provide an intuitive explanation for them, and use that explanation to make additional predictions about observable variables of interest. We see no reason to rule out such explanations simply because they come from psychology.

Indeed, there is an argument for favoring such explanations, all else equal. Suppose we have two models, both consistent with all the data we have and which make different out of sample predictions. Suppose the two models are equally tractable in terms of generating these out of sample predictions. Suppose one is consistent with psychologist's views of human decision–making and the other is not. Which would we prefer to use to make predictions?

We have already argued that one may well prefer to use the more "intuitive" or the more "plausible" model in such a situation, noting Kreps' (1990) comment that this is like a kind of data. The consistency with psychology can be another way of generating such plausibility. It seems reasonable to believe that a psychotherapist who has seen thousands of patients over a long period of time might develop a real understanding or knowledge of things that are not easily quantified. If so, we can think of consistency with the kinds of explanations such a person gives as a different kind of consistency with data.

In addition to affecting the plausibility of competing predictions, psychological considerations can suggest — indeed, have suggested — altogether new predictions. In some cases, such as models of ambiguity aversion, the models economists have developed make little direct use of psychology in motivating the formulation or assumptions. In other cases, though, the "story" of the model explicitly incorporates ideas from psychology.

For example, consider Laibson's (2001) notion of cue–based consumption. He observes that psychologists have identified situations where external influences act as cues for various desires or cravings. He formalizes this idea through the existence of exogenous, random observables which affect the agent's utility function and hence the agent's choices. He notes various new predictions that this model yields. For example, observing a billboard advertising cigarettes may lead someone trying to quit smoking to crave

cigarettes and therefore to smoke. Note that Laibson's model is based on a utility func-
tion and, in that sense, is an economist's model. The point is that the restrictions on
the model — the effect of the random observables — and the resulting predictions about
choice are motivated by psychology and would not have been considered otherwise.[30]

As evidence that psychology has been useful, we note the rapidly growing body of
research in economics and decision theory along these lines. While the extent of the
influence of psychology varies from the minimal to the extreme, the influence seems (to us)
clearly present in such papers as Eliaz and Spiegler (2008), Epstein (2008), Epstein and
Kopylov (2008), Kopylov and Noor (2009), Masatlioglu, Nakajima, and Ozbay (2009),
and Rubinstein and Salant (2008).

4.2   Nonchoice Data

In the previous subsection, we alluded to the possibility of using data which was not
generated by the choices of agents. While this seems anathema to the traditional notion
of revealed preference, why wouldn't an economist want more data if it might help in
predicting choices? One reason some economists reject use of such data, we believe, is
that it is sometimes not clear how to interpret it.

For example, some recent empirical work (see Manski (2004)) asks agents what they
think the probability of some occurrence is. Some researchers (e.g., Blackburn, et al
(2008)) then assume that these agents maximize expected utility using the reported
probabilities. Even if (big if!) the agent can be predicted by subjective expected utility
and reports his true subjective probabilities, it is not clear that we should use his self–
reported beliefs to predict his behavior. As discussed in the previous subsection, it
is possible that the agent's true utilities are state–dependent and hence that his true
subjective probabilities are not the ones we'd use for a state–independent representation.
On the other hand, the fact that certain data must be used with caution doesn't imply
that it shouldn't be used at all.

Beyond this, we conjecture that the economists who reject the use of nonchoice data
do so for two reasons. First, they focus on refutation as the sole means of model selection.
Second, they focus on predictions which only involve statements about behavior. Clearly,
if one adopts these two principles, there is no purpose in considering data which is not
about behavior. We argued at length above against an excessive focus on refutation. As
to the second principle, we do agree that the predictions of interest should be predictions
of behavior. However, a prediction of the form "if certain nonchoice variables take on

---

[30]Gul and Pesendorfer (2008) note that psychologists may desire break behavior into that which is
related to cues and that which isn't in a way which is different from what would be most useful for
economists. We see this as a valid concern, but essentially orthogonal to our point.

certain values, then certain choices will be made" strikes us as a prediction of behavior, one for which nonchoice data would be crucial.[31]

The reality is that economists are typically faced with situations where many different models fit the available data and these models have different out of sample predictions. In some cases, nonchoice data may help distinguish among these models.

For example, suppose we observe a child who chooses 6 pieces of chocolate over what is described to him as "3 times 3" pieces and suppose we are trying to predict the child's choice between six and nine pieces. It seems natural to think that we would find information about the child's last multiplication test useful. Similarly, Camerer, Loewenstein, and Prelec (2004), henceforth CLP, note that observing that an agent does not buy peanuts does not allow us to distinguish a mild dislike for peanuts from a life–threatening allergy. Yet the difference would be important for predicting a willingness to accept money to eat peanuts or to consume food produced in the same factory as foods containing peanuts. Nonchoice data, this time in the form of medical tests, would help us predict their behavior.[32]

If we take nonchoice data seriously as relevant for predicting behavior, we'll need models that predict the relationship between behavior and these other variables and hence we'll want axioms that do the same.[33] We already discussed the idea that psychological considerations might motivate issues we consider and the models we use, in that indirect sense, affecting our axioms. Clearly, if we use psychological nonchoice data, the role of psychology in suggesting hypotheses will be stronger.

4.3   What is a Good Axiomatization?

In the preceding sections, we commented extensively on what makes a good model or, in the context of decision theory, a good representation of behavior. In this section, we take a representation as given and ask what makes a good axiomatization for this representation. Many of the same criteria we use for judging a model are again critical. Recall that the purpose of a representation theorem is to add to our understanding of and ability to test the model of the agent. In particular, the representation theorem gives our statement of the predictions of the model about observable variables.

---

[31]Indeed, most empirical work does precisely this. If one is estimating the demand for ice cream on a particular day in a particular place, surely one uses temperature as an explanatory variable. We don't think this point is controversial — if it is, we can't see why.

[32]We interpret Gul and Pesendorfer's (2008) criticique of CLP as arguing that the medical tests cannot be used to refute or confirm choice theory, a point we agree with. We do not believe (and are not sure if Gul and Pesendorfer believe) that this implies nonchoice data is never helpful.

[33]See Caplin and Dean (2009) for some steps in this direction.

Hence the first thing a representation theorem must do is to identify the key behavior which corresponds to the phenomenon being modeled, which sometimes requires a new domain of behavior. While this step is very basic in a sense, many of the most fundamental developments in decision theory stem from an insightful approach to it. A classic example in the theory of subjective expected utility is Ramsey's (1926) observation that if agents are subjective expected utility maximizers, then their preferences over bets will reflect the subjective probabilities (for the state–independent "normalization"). Specifically, if the agent prefers betting $1 if A occurs and 0 otherwise to $1 if B occurs and 0 otherwise, then the subjective probability of A is higher than that of B. This simple insight was a key step in the development of Savage's axiomatization.

As another example, Kreps' (1979) study of the demand for flexibility introduced a new domain and the key property on that domain. Kreps recognized that decisions about flexibility could be thought of in terms of choice of a *menu* — that is, a partial commitment regarding one's future choices. The key axiom is then the formal statement of a desire for flexibility, specifically, that if one menu is a subset of another, then the agent prefers the larger menu.[34] Gul and Pesendorfer (2001) recognized that temptation would induce the opposite — a demand for commitment instead of flexibility — and thus introduced the use of preferences over menus to study temptation. Their key behavioral property is that agents prefer smaller menus in certain situations. Recognizing that non–Bayesian updating induces an intertemporal inconsistency akin to temptation, Epstein (2006) extended the domain to menus over (Anscombe–Aumann) acts and used this richer domain to give a behavioral characterization of such updating.

Finally, we note that a creative development of a domain can itself be the key step. For example, prior to the work of Kreps and Porteus (1978) and Segal (1990), every model involving uncertainty and time implicitly assumed that only the probability distribution over the information received at the time of an action choice was relevant. In particular, as Kreps and Porteus emphasized, this assumes that the timing of the resolution of uncertainty is irrelevant — if an agent can't act till tomorrow afternoon regardless, then it would not matter to him whether the uncertainty he faces is resolved tonight or tomorrow morning. Similarly, as Segal emphasized, this assumes that a sequence of lotteries which determine the agent's consumption are equivalent to the agent to the overall lottery over consumption they imply. Kreps–Porteus and Segal introduce models where one can make these distinctions and hence illustrate the behavioral impact they have.[35]

Identifying the key behavior and the domain is the most essential step, but also the step which is closest to an art. Thus we find it difficult to tell the reader how to do it or

---

[34]Dekel, Lipman, and Rustichini (2001) extend the domain considered by Kreps to allow for menus of lotteries. In contrast to the other examples, this extension does not enable studying a novel form of behavior but instead serves the role of enabling better identification of the parameters in Kreps' model.

[35]These models underlie the framework developed by Epstein and Zin (1989) to study asset demands where risk aversion and consumption smoothing are identified separately.

how to distinguish "good" choices from "bad." In the rest of this section, we discuss a number of simpler issues where it is much easier to offer some advice.

Since axiomatizations are supposed to state the predictions of the model about observables, the first guideline is easy: axioms should be about variables of interest which are at least potentially observable. Unfortunately, it is not always obvious which variables are observable and many disagreements about axiomatizations hinge on such controversies. For example, consider Caplin and Leahy (2001), whose axioms concern preferences over lotteries over "psychological states" with a function which relates lotteries over physical outcomes to lotteries over psychological outcomes. Our first instinct is to argue that psychological outcomes are not directly observed and the function relating these to physical outcomes cannot be identified. On the other hand, we acknowledge that psychologists and the authors may well disagree. Perhaps measurements of psychological outcomes can be examples of the kind of nonchoice data we suggested earlier could be valuable. By contrast, Caplin and Dean (2009) study nonchoice variables of interest that are more naturally taken to be observable.

The second point is equally immediate: An axiomatization which does not say more than the representation is not helpful. In some cases, the behavioral meaning of the representation is obvious. We noted earlier that we don't need an axiomatization to tell us that $u' > 0$ represents "more is better." In other cases, the behavioral meaning is not obvious, but the key axioms are little more than a restatement of the representation. For example, we mentioned our 2007 representation of temptation where an agent has some probability beliefs regarding whether his future self will succumb to temptation. Formally, this model represents the agent's evaluation of a menu by the expected utility of a certain probability distribution over items in the menu. We could have axiomatized this model by making our main axiom the statement that for any menu, there is a lottery over the items in the menu such that the agent is indifferent between the menu and the lottery. This is indeed an implication of the model. On the other hand, this axiom just says the representation in different words and so doesn't tell us anything new.

Also, obviously, axioms should be simple and clearly interpretable. Otherwise, again, the axioms have added nothing to our understanding. While it is difficult to define simplicity, we note a few obvious guidelines. First, it is generally better to state axioms in terms of the preferences, not a series of relations derived from the preference. For example, a key in Savage's representation theorem is the "more–likely–than" relation, which is constructed from the preference relation. Yet Savage states his axioms in terms of the preference, not in terms of the derived relation, since the preference is what we are making predictions about. Second, axioms involving existential quantifiers are often too complex to be interpreted in this sense. It is hard to get an intuition about an axiom that says that an object exists with certain properties.[36] Finally, as noted in Section 2, it

---

[36]Mixture continuity is a notable (though untestable) exception.

is desirable to have axioms that are familiar, either because they are standard or because they are intuitive variations on standard axioms. Obviously, it is easier to understand the familiar. Also, this approach makes it easier to compare the new model to existing work and see what it adds of value to our understanding.

Since one role of axioms is to enable testing of the model, clearly, it is valuable to state axioms in a way which make such testing easier. Many of our comments about clarity of the axioms applies directly here. For example, this is another reason to avoid existential quantifiers since it can be very difficult to test a statement of the form "there exists an object with the following properties." Also, statements of axioms which suggest how to formulate experiments or otherwise identify key observations are particularly helpful.

Finally, we turn to several issues regarding the set of axioms as a whole. Just as individual axioms should be as interpretable as possible, it is important to convey the meaning of the group as well. An obvious point in this regard is that a representation theorem with a very large number of axioms is not going to be as clear as one with a small number. (Also, we suspect that if a model really requires a large number of axioms, there are probably many issues being combined that should be treated separately.) Second, we often find that axioms stated in a conditional form (e.g., if a is preferred to b and b is preferred to c, then a is preferred to c) are crucial to to keeping the stated properties clear. By contrast, if all axioms take the form of universal statements (e.g., more is better), the individual axioms may be clear but, except in trivial cases, we cannot imagine avoiding a large and complex set of such axioms. Intuitively, conditional axioms allow us to make many inferences from a few observations. Without such axioms, we have to effectively list all preference statements.

A more difficult principle to adhere to is to state the set of axioms in a way which enables the reader to see the role of each piece in generating the representation. This makes the analysis more informative. However, there is typically a great deal of interaction among axioms, making this hard to do. Indeed, if there were no interaction, arguably, this would indicate that this representation is mixing multiple issues together. An additional tension is that theorems with interactions among the axioms are often more elegant and mathematically interesting.

An easier but still nontrivial principle is to clearly separate the axioms intended to be relatively realistic and/or statements of the main focus from axioms intended primarily as useful simplifications. While this sounds easy enough, in practice, the "innocuous" simplifications can have unexpected substantive implications. For example, in Gul and Pesendorfer's (2001) paper on temptation, it is natural to read their "behavioral" axiom as the property they call set betweenness and to view their other axioms as useful but inessential simplifications. However, they show in Gul and Pesendorfer (2005) that independence buys more than the simplification of a linear structure. Indeed, Dekel, Lipman,

and Rustichini (2009) give an example of intuitive temptation behavior that is ruled out in their model not by set betweenness, but by the combination of set betweenness and independence.[37]

Finally, a tricky issue is whether the axiom should be stated in a strong form or a weak one. Suppose an axiom can be divided into two parts, one part of which is already implied by the other axioms and one of which isn't. Is it better to only state the nonredundant part or to state the whole axiom? Often, the stronger form of the axiom is much more intuitive, while, by definition, only the nonredundant part is really relevant. Usually, we find it's best to state both and explain the point. We find this particularly true of continuity axioms where the weakest possible form of the axiom is often mathematically very complex. Since such axioms are not testable, we think it's usually best to state such axioms in the simplest form, not the weakest.

On the same subject, we note that weaker axioms are not *per se* better axioms. If one uses very weak assumptions to get a weak result, it's not clear the exercise is useful. On the other hand, if one uses weak assumptions to get a strong result, this is surely either because the strong result is weaker than it appears or the assumptions are stronger than they appear.[38]

5   Conclusion: Welfare Economics

We conclude this paper with a brief discussion of a topic we are reluctant to omit but have relatively little to say about. When it comes to prescriptive analysis, prediction of choices is insufficient. We must also have some principled way to identify the welfare implications of choices. Traditionally, economists have assumed that choice reveals utility or welfare directly. That is, if an agent chooses A over B, this implies he is better off consuming A than consuming B. Of course, as soon as we consider models where agents have difficulties in making good decisions and hence may make mistakes, such an assumption seems rather inappropriate.

Gul and Pesendorfer (2008) take the position that even welfare economics is only predictive, not prescriptive. For example, they suggest that results about whether a particular institution leads to Pareto optimal outcomes should not be seen as determining whether the institution is socially valuable. Instead, this should be seen as a charac-

---

[37]Dekel, Lipman, and Rustichini maintain an independence axiom, but have some discussion of why this is not an innocuous simplification. For related discussions, see also Fudenberg and Levine (2006) and Noor and Takeoka (2009).

[38]On the other hand, if the misperception of the strength of the assumption or conclusion is widespread, such a theorem could be very instructive.

terization of whether the institution is likely to be long–lived. If there is some other institution which leads to Pareto preferred outcomes, then this statement, they argue, should be interpreted as saying that agents are likely to eventually switch to such an alternative.

While their predictive point is quite reasonable, it seems clear that many economists want to know what useful guidance welfare economics can give to policymakers. It seems plausible that at least some policymakers in at least some settings have some perhaps limited desire to choose policies that will benefit citizens. If we take this perhaps insufficiently cynical point of view seriously, what advice can we give such policymakers?[39]

Unfortunately, it is vastly more difficult to prescribe than to predict. Given how poorly economists do at prediction, this should be troubling! Prescription is more difficult than prediction because we need to know more than what people will do; we also need to know what really determines their well–being. If we do not have the usual revealed preference to identify this, what do we fall back on?

In some sense, this question is surely unsolvable. We cannot imagine how one could ever conclusively prove that a particular outcome is better than another for a particular person. On the other hand, reacting to this difficulty by giving up on the problem instead of accepting less than conclusive proof seems to give up too quickly. For example, Koszegi and Rabin (2008) discuss observations of a gambler betting on tosses of a coin. Suppose we saw this gambler betting more heavily on tails after a string of heads. Such behavior would strongly suggest that the gambler is subject to the gambler's fallacy, that is, the belief that a string of heads is likely to be followed by a tail. As Koszegi and Rabin note, our observation cannot prove this. In principle, it could simply be that the agent values income in a state of the world where tails follows multiple heads more than in a state of the world where there has been only a single toss which came up tails. Yet this alternative explanation seems highly implausible. Thus it seems most natural to explain the behavior as a mistake and to consider policies that benefit the gambler based on this hypothesis. On the other hand, if the gambler's "mistake" is because he enjoys believing he will win, would policies that interfere with his gambling necessarily improve his welfare?

We find it difficult to offer guidelines for when normative conclusions should be based on revealed preference and when they should not or for identifying useful policies when revealed preference fails. For example, if a person tries to find ways to commit himself to a particular course of action (losing weight, saving money) and expresses frustration over his lack of success at this commitment, then it seems plausible that social interference

---

[39]As Gul and Pesendorfer (2007) note, if policymakers have well–defined objectives over outcomes, they only need our predictions. Our comments here are focused on the possibility that their objective is to make citizens "better off" and the difficulty in defining what that means.

with this person might improve his well being. But it is not always obvious which actions express this person's "true" welfare. Is it the attempts to commit or the fact that these attempts were unsuccessful (and therefore possibly half–hearted)? While we don't argue against all policies that interfere with individual choices, our inclination in the face of doubt is to avoid such interference.

We note that numerous authors have made interesting attempts to resolve this problem and we refer the interested reader to their papers. See, for example, Bernheim and Rangel (2007, 2009, forthcoming), Chambers and Hayashi (2008), and Noor (2009).

References

[1] Al–Najjar, Nabil I. and Jonathan Weinstein, "The Ambiguity Aversion Literature: A Critical Assessment," *Economics and Philosophy*, forthcoming.

[2] Allais, Maurice, "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'École Américaine," *Econometrica*, **21**, October 1953, 503–546.

[3] Anscombe, F. J., and Robert J. Aumann, "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, **34**, March 1963, 199–205.

[4] Arrow, Kenneth, *Essays in the Theory of Risk Bearing*, Chicago: Markham, 1971.

[5] Bernheim, Douglas, and Antonio Rangel, "Toward Choice–Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review Papers and Proceedings*, **97**, May 2007, 464–470.

[6] Bernheim, Douglas, and Antonio Rangel, "Beyond Revealed Preference: Choice–Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, **124**, February 2009, 51–104.

[7] Bernheim, Douglas, and Antonio Rangel, "Neuroeconomics: A Sober (but Hopeful) Appraisal," *AEJ: Microeconomics*, forthcoming.

[8] Bewley, Truman, "Knightian Decision Theory: Part I," *Decisions in Economics and Finance*, **25**, November 2002, 79–110.

[9] Binmore, Ken, *Rational Decision*, Princeton: Princeton University Press, 2008.

[10] Blackburn, Brian, Aprajit Mahajan, Alessandro Tarozzi, and Joanne Yoong, "Bednets, Information and Malaria in Orissa," Stanford University, unpublished, 2008.

[11] Bowman, David, Deborah Minehart, and Matthew Rabin, "Loss Aversion in a Consumption–Savings Model," *Journal of Economic Behavior and Organization*, **38**, February 1999, 155–178.

[12] Camerer, Colin, "The Case for Mindful Economics," in Caplin and Schotter (2008), 43–69.

[13] Camerer, Colin, George Loewenstein, and Drazen Prelec, "Neuroeconomics: Why Economics Needs Brains," *Scandinavian Journal of Economics*, **106**, #3 2004, 555–579.

[14] Caplin, Andrew, and Andrew Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*, Oxford: Oxford University Press, 2008.

[15] Caplin, Andrew, and Mark Dean, "Choice Anomalies, Search, and Revealed Preference," New York University, unpublished, 2009.

[16] Caplin, Andrew, and John Leahy, "Psychological Expected Utility Theory And Anticipatory Feelings," *Quarterly Journal of Economics*, **116**, February 2001, 55–79.

[17] Chambers, Christopher, and Takashi Hayashi, "Choice and Individual Welfare," Cal Tech, unpublished, 2008.

[18] Chew, Soo Hong, "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica*, **51**, July 1983, 1065–1092.

[19] Dekel, Eddie, "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory*, **40**, December 1986, 304–318.

[20] Dekel, Eddie, and Barton L. Lipman, "Self–Control and Random Strotz Representations," Boston University, unpublished, 2007.

[21] Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini, "Representing Preferences with a Unique Subjective State Space," *Econometrica*, **69**, July 2001, 891–934.

[22] Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini, "Temptation–Driven Preferences," *Review of Economic Studies*, **76**, July 2009, 937–971.

[23] Dubra, Joan, Fabio Maccheroni, and Efe Ok, "Expected Utility Theory without the Completeness Axiom," *Journal of Economic Theory*, **115**, March 2004, 118–133.

[24] Ellsberg, Daniel, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, **75**, November 1961, 643–669.

[25] Ellsberg, Daniel, *Risk, Ambiguity and Decision*, New York: Garland, 2001.

[26] Eliaz, Kfir, and Efe Ok, "Indifference or Indecisiveness? Choice Theoretic Foundations of Incomplete Preferences," *Games and Economic Behavior*, **56**, July 2006, 61–86.

[27] Eliaz, Kfir, and Ran Spiegler, "Consideration Sets and Competitive Marketing," University College London, unpublished, 2008.

[28] Epstein, Larry, "An Axiomatic Model of Non–Bayesian Updating," *Review of Economic Studies*, **73**, April 2006, 413–436.

[29] Epstein, Larry, "Living with Risk," *Review of Economic Studies*, **75**, October 2008, 1121–1141.

[30] Epstein, Larry, and Igor Kopylov, "Cold Feet," *Theoretical Economics*, **2**, September 2007, 231–259.

[31] Epstein, Larry, and Martin Schneider, "Recursive Multiple Priors," *Journal of Economic Theory*, **113**, November 2003, 1–31.

[32] Epstein, Larry, and Stanley Zin, "Substitution, Risk Aversion and the Temporal Behaviour of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, **57**, July 1989, 937–969.

[33] Fudenberg, Drew, "Advancing Beyond Advances in Behavioral Economics," *Journal of Economic Literature*, **44**, September 2006, 694–711.

[34] Fudenberg, Drew, and David Levine, "A Dual Self Model of Impulse Control," *American Economic Review*, **96**, December 2006, 1449–1476.

[35] Gilboa, Itzhak, and David Schmeidler, "Maxmin Expected Utility with Non–Unique Prior," *Journal of Mathematical Economics*, **18**, #2 1989, 141–153.

[36] Gilboa, Itzhak, *Rational Choice*, Cambridge, MA: MIT Press, 2009.

[37] Gollier, Christian, *The Economics of Risk and Time*, Cambridge, MA: MIT Press, 2001.

[38] Gul, Faruk, and Wolfgang Pesendorfer, "Temptation and Self–Control," *Econometrica*, **69**, November 2001, 1403–1435.

[39] Gul, Faruk, and Wolfgang Pesendorfer, "A Simple Theory of Temptation and Self–Control," Princeton University, unpublished, 2005.

[40] Gul, Faruk, and Wolfgang Pesendorfer, "Welfare without Happiness," *American Economic Review Papers and Proceedings*, **97**, May 2007, 471–476.

[41] Gul, Faruk, and Wolfgang Pesendorfer, "The Case for Mindless Economics," in Caplin and Schotter (2008), 3–42.

[42] Hayashi, Takahashi, "Regret Aversion and Opportunity–Dependence," *Journal of Economic Theory*, **139**, March 2008, 242–268.

[43] Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, **47**, March 1979, 263–291.

[44] Kalai, Gil, Ariel Rubinstein, and Ran Spiegler, "Rationalizing Choice Functions By Multiple Rationales," *Econometrica*, **70**, November 2002, 2481–2488.

[45] Karni, Edi, "Foundations of Bayesian Theory," *Journal of Economic Theory*, **132**, January 2007, 167–188.

[46] Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji, "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, **73**, November 2005, 1849–1892.

[47] Kopylov, Igor, and Jawwad Noor, "Self–Deception and Choice," Boston University, unpublished, 2009.

[48] Koszegi, Botond, and Matthew Rabin, "A Model of Reference–Dependent Preferences," *Quarterly Journal of Economics*, **121**, November 2006, 1135–1165.

[49] Koszegi, Botond, and Matthew Rabin, "Revealed Mistakes and Revealed Preferences," in Caplin and Schotter (2008), 193–209.

[50] Kreps, David, "A Representation Theorem for 'Preference for Flexibility,'" *Econometrica*, **47**, May 1979, 565–576.

[51] Kreps, David, *Notes on the Theory of Choice*, Boulder: Westview Press, 1988.

[52] Kreps, David, *A Course in Microeconomic Theory*, Princeton: Princeton University Press, 1990.

[53] Kreps, David, and E. L. Porteus, "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, **46**, January 1978, 185–200.

[54] Laibson, David, "A Cue–Theory of Consumption," *Quarterly Journal of Economics*, **116**, February 2001, 81–119.

[55] Machina, Mark, "'Expected Utility' Analysis without the Independence Axiom," *Econometrica*, **50**, March 1982, 277–323.

[56] Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini, "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, **74**, November 2006, 1447–1498.

[57] Manski, Charles, "Measuring Expectations," *Econometrica*, **72**, September 2004, 1329–1376.

[58] Manski, Charles, "Actualist Rationality," Northwestern University, unpublished, 2009.

[59] Manzini, Paola, and Marco Mariotti, "Sequentially Rationalizable Choice," *American Economic Review*, **97**, December 2007, 1824–1839.

[60] Masatlioglu, Yusufcan, and Efe Ok, "A Canonical Model of Choice with Initial Endowments," University of Michigan, unpublished, 2009.

[61] Masatlioglu, Yusufan, Daisuke Nakajima, and Erkut Ozbay, "Revealed Attention," University of Michigan, unpublished, 2009.

[62] Milnor, J., "Games Against Nature," in R. M. Thrall, C. H. Coombs, and R. L. Davis, eds., *Decision Processes*, New York: Wiley, 1954.

[63] Noor, Jawwad, "Subjective Welfare," Boston University, unpublished, 2009.

[64] Noor, Jawwad, and Norio Takeoka, "Uphill Self Control," Boston University, unpublished, 2009.

[65] Pratt, John, "Risk Aversion in the Small and in the Large," *Econometrica*, **32**, January–April 1964, 122–136.

[66] Ramsey, Frank P., "Truth and Probability," 1926, in Ramsey, F., *The Foundations of Mathematics and other Logical Essays*, ed. by R. B. Braithwaite,London: Kegan, Paul, Trench, Trubner & Co., 1931.

[67] Rubinstein, Ariel, *Modeling Bounded Rationality*, Cambridge, MA: MIT Press, 1998.

[68] Rubinstein, Ariel, and Yuval Salant, "(A, f): Choice with Frames," *Review of Economic Studies*, **75**, October 2008, 1287–1296.

[69] Samuelson, William, and Richard Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, **I**, March 1988, 7–59.

[70] Sarver, Todd, "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, **76**, March 2008, 263–305.

[71] Savage, Leonard J., *The Foundations of Statistics*, New York: Dover Publications, 1954.

[72] Schlag, Karl, "Why Minmax Regret," UPF, unpublished, 2006.

[73] Schmeidler, David, "Subjective Probability and Expected Utility without Additivity," *Econometrica*, **57**, May 1989, 571–587.

[74] Segal, Uzi, "Two–Stage Lotteries without the Reduction Axiom," *Econometrica*, **58**, March 1990, 349–377.

[75] Simon, Herbert, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, **69**, February 1955, 99–118.

[76] Simon, Herbert, *Models of Bounded Rationality, Volume 2*, Cambridge, MA: MIT Press, 1982.

[77] Siniscalchi, Marciano, "Vector Expected Utility and Attitudes toward Variation," *Econometrica*, **77**, May 2009, 801–855.

[78] Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, **87**, August 1973, 355–374.

[79] Spiegler, Ran, "On Two Points of View regarding Revealed Preferences and Behavioral Economics," in Caplin and Schotter (2008), 95–115.

[80] Stoye, Joerg, "Axioms for Minmax Regret Choice Correspondences," New York University, unpublished, 2008.

[81] Tversky, Amos, and Daniel Kahneman, "Loss Aversion in Riskless Choice: A Reference–Dependent Model," *Quarterly Journal of Economics*, **106**, November 1991, 1039–1061.

[82] Tversky, Amos, and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, **5**, October 1992, 297–323.

[83] Wakker, Peter, "Annotated References on Decisions and Uncertainty," Erasmus University, unpublished, 2009.