Economics and Philosophy, **16** (2000), 99–115 Copyright © Cambridge University Press

REVEALED PREFERENCE, BELIEF, AND GAME THEORY

Daniel M. Hausman
University of Wisconsin-Madison and London School of Economics

The notion of 'revealed preference' is unclear and should be abandoned. Defenders of the theory of revealed preference have misinterpreted legitimate concerns about the testability of economics as the demand that economists eschew reference to (unobservable) subjective states. As attempts to apply revealed-preference theory to game theory illustrate with particular vividness, this demand is mistaken.

1. REVELATION

Until recently, 'revealed-preference theory' referred to efforts, originating with Paul Samuelson (1938), to define claims about subjective preferences in terms of claims about choices or to eliminate reference to subjective preferences altogether. In search of 'operationally meaningful theorems' (1947, p. 3), Samuelson argued that certain claims about consumer *choices* capture the whole 'empirical meaning' of utility theory (1947, p. 116). In Samuelson's project the economic task of determining the empirical consequences of utility theory was wedded to a philoso-

As the first Ludwig Lachmann Fellow at the London School of Economics, I would like to thank the Charlottenburg Trust for its support of my research and to dedicate this paper to Ludwig Lachmann's memory. He would appreciate its central argument, which supports his emphasis on the importance of subjective expectations to economic phenomena. I would like to acknowledge helpful comments from Max Albert, Sebastiano Bavetta, Pierluigi Barotta, Richard Bradley, Nicolas Gravel, Maarten Janssen, Mark Machina, Philippe Mongin, Marco del Seta, Max Steuer, John Sutton, audiences at the University of Cergy-Pontoise, the Swedish Collegium for Advanced Studies in the Social Sciences, Erasmus University, and the London School of Economics, and anonymous referees, although no doubt some of my errors survived all this criticism. Some parts of Section 5 derive from Section 1.3 of Hausman, 1992.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

phical project of eschewing all reference to subjective (and hence unobservable) states.

In some contemporary work, the term, 'revealed-preference' is used in a different way. Green and Osbard (1991) and Border (1992) take revealed-preference theory as addressing the question of when preferences can be inferred from choices *given premises concerning belief*. This sort of 'revealed-preference theory' should be distinguished from Samuelson's. One reason to abandon talk of 'revealed preference' is to avoid confusing these different projects.

I shall use the term 'revealed-preference theory' to refer to the original project. It is an interpretation of formal results explored initially by Paul Samuelson (1938), generalized and developed by many others (especially Houtthakker (1950)), and elegantly summarized a generation ago by Amartya Sen (1971). It is important to scrutinize these formal results, since many economists have drawn the mistaken conclusion that the proofs in this literature demonstrate that choice reveals preference. Thus, for example, Henderson and Quandt take it as proven that 'the existence and nature of her [an agent's] utility function can be deduced from her observed choices among commodity bundles' (1980, p. 45) and Gravelle and Rees take it as demonstrable that 'the utility maximizing theory of the consumer and the revealed preference theory are equivalent' (1981, p. 115). If Henderson and Quandt or Gravelle and Rees were correct, critics of revealed-preference would be conclusively refuted before they began, unless they could show some mistake in the proofs.

Here is a simple version of the fundamental technical result: Let X be a non-empty finite set, S a subset of X and K the set of all non-empty subsets of X. Let $C(\cdot)$ be a set-valued function from each $S \in K$ to a non-empty subset $C(S)$ of S. Define a two-place relation R as xRy if and only if $\exists S$ such that $x, y \in S$ and $x \in C(S)$. Define $C^R(S)$ as $\{x \in S \mid \forall y \in S, xRy\}$. One axiom is needed (WARP): If $\exists S \in K$ s.t. $x \in C(S)$ and $y \in S$ but $y \notin C(S)$, then $\nexists S$ s.t. x and $y \in S$ and $y \in C(S)$. Given the definition of R, WARP may be restated: if $\exists S \in K$ s.t. $x \in C(S)$ and $y \in S$ but $y \notin C(S)$, then $\neg yRx$. WARP implies that R is complete and transitive and that $C^R(S) = C(S)$. I shall call this theorem, 'the revelation theorem'. Its proof can be sketched simply. Since $\forall x,y$, $C(\{x,y\})$ is not empty, R is complete. Suppose that R were not transitive and that xRy, yRz, but not xRz. Then by WARP $C(\{x,y,z\})$ would have to be empty. So R must be transitive. $\forall S$, if $x \in C(S)$, then $\forall y \in S$, xRy, and so $x \in C^R(S)$. Conversely, if for any S, $x \notin C(S)$, then for some $y \in S$, yRx and $\neg xRy$, and so $x \notin C^R(S)$. For elaborations and proofs, see Sen (1971).

Although intuitive notions of choice and preference guided the formulation of WARP (the 'weak axiom of revealed preference'), I stated the theorem without using the words 'choice' or 'preference'. To see how one can consistently deny that choices reveal preferences without

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

challenging the revelation theorem, one must distinguish carefully between the theorem and its interpretations. It helps to notice that one can interpret *C(S)*, *xRy*, and the revelation theorem in many ways. For example, *K* might be all sets of participants in athletic competitions and *C(S)* the set of winners of competitions among competitors in *S* (see Mackay 1980). '*xRy*' could be read as '*x* has at least as much athletic ability as *y*'. Given this interpretation, the revelation theorem says that if the results of competitions satisfy WARP, then 'has at least as much athletic ability' is complete and transitive, and athlete *x* beats or ties athlete *y* if and only if *x* has no less ability than *y*. It is obvious that one can question whether consistent winners never have less ability than losers without questioning the theorem.

The intended interpretation of the revelation theorem is of course that *C(S)* consists of the alternatives an agent chooses from *S*. Even though *R* is defined in terms of choice, '*xRy*' is supposed to be read as 'an agent weakly prefers *x* to *y*' Given these interpretations of *C(S)* and *R*, the revelation theorem shows that if an agent's choices satisfy WARP, then the agent's preference relation can be deduced from the agent's choices.

Revealed-preference theory consists of the revelation theorem plus these interpretations of C(S) and xRy. Two conclusions have been drawn:

1. The revelation theorem makes it possible for economists to dispense with the notion of preference. There is nothing to be said employing the notion of preference that cannot be said employing only the language of choice. In this way, economists with strict empiricist scruples, such as Samuelson or Little (1949), defended the scientific credentials of economics. Disreputable references to unobservable preferences are eliminable. Economics can proceed entirely on the observable basis of choices.
2. The revelation theorem shows how choices reveal preferences and how to test claims about preferences. Instead of taking the one-to-one correspondence between preferences and choices as showing that talk of subjective preferences is dispensable, the correspondence between choice and preference legitimates talk of subjective states. In Sen's words, 'The rationale of the revealed-preference approach lies in the assumption of revelation and not in doing away with the notion of underlying preferences . . . '. (1973, p. 244)[1]

The theorem by itself does not, of course, establish these interpretations.

Remember that revealed-preference theory is *not* the claim that, given claims about the agent's beliefs, one can infer the agent's

[1] On the alternative interpretation, the analogous conclusions are (1) that the notion of 'athletic ability' is dispensable and (2) that the notion can be constructed out of consistent winning relations.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

preferences from the agent's choices. The latter view does not avoid referring to 'unobservable' subjective states. It does not show that the notion of preference can be defined in terms of choices or that it can be eliminated altogether. Revealed-preference theory permits belief to play no role, and that is part of its extreme empiricist point. Revealed-preference theory holds that preference can be defined in terms of choice, regardless of belief. This essay is a critique of revealed-preference theory, not of the uncontroversial claims that choice (that is, action – including gestures and verbal behaviour) is virtually the only evidence of preference and that preferences can be inferred from choices, given information about beliefs. Although it would be charitable to interpret Samuelson and Little as arguing only that there is a correspondence between preferences and choices in contexts in which unproblematic a priori information about beliefs is available, this interpretation is insufficient to support their empiricist commitments; and it prevents one from appreciating the confusion that talk of revealed preference has caused.

2. WHY CHOICE ALONE DOES NOT REVEAL PREFERENCE

The alternative athletic-competition interpretation of the revelation theorem shows how the revelation theorem can be a genuine theorem yet fail to establish that choice reveals preference. Consistent winners may sometimes have less athletic ability than losers. Winners might cheat, they might be consistently lucky, or they might try harder. Winning does not depend solely on athletic ability, and it is a mistake to interpret the relation among athletes that is implicit in the consistency of the results as 'has no less athletic ability'.

The same points apply to the choice-preference interpretation. What agents choose is not necessarily what they prefer, because agents may make mistakes, because agents may act when they cannot rank some of the alternatives, and because their choices may be dictated by factors ordinarily distinguished from preference, such as moral principle.

Amartya Sen made these criticisms a generation ago, but not everyone accepts them. The last point, for example, commands broad scepticism among economists, who question the pre-theoretical notion of 'preference' that it assumes. So let us dig deeper.

It is false to maintain that if B prefers x to y, then B will never choose y when x is available. B may prefer x to y, yet choose y because B falsely believes that x is not available.[2] A revealed-preference theorist might reply that it does not matter whether B 'really prefers' x to y and fails to choose x because of mistaken beliefs. The choice of y is sufficient to show

[2] In such a case, one of the conditions of the revelation theorem will be ambiguous. Is $x \in S$ when the agent does not know that x is available?

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

that either B does not prefer x to y or that B violates WARP. This response would be unfortunate, because different predictions are in order depending on the reason why B fails to choose x. Some economists would, however, disavow any interest in predicting the behaviour of individuals, and they might be content, despite the costs, with the claim that if B does not choose x when x and y are available, then B does not prefer x to y. Later we will see that such complacence is not an option for game theorists and many other economists.

According to 'folk psychology' (the theory people employ in everyday life to predict and explain actions), an agent's desires or preferences, like an agent's beliefs and expectations, are mental states that constitute reasons for their actions and causes of their actions. This is a claim about a theory, not about the dictionary meanings of the words 'prefers' and 'preference', which may be used in English as synonyms for 'chooses' and 'choice'.[3] Beliefs and preferences, unlike actions, are subjective, and they are distinct from the actions they give rise to, explain, and justify. Within folk psychology, one cannot infer preferences from choices alone, because choices depend on both belief and preference. To explain why an investor purchases one stock rather than another, one needs to know not only her preferences (for larger returns), but also her beliefs (that the returns on the first were higher). The inverse inference from choice to preference depends equally on premises concerning beliefs. Indeed, opposite beliefs and preferences may lead to exactly the same choice. Even though choice is determined by preference, choice could not by itself reveal preference, because, given the right set of beliefs, any set of choices is consistent with any set of preferences. Neither beliefs nor preferences can be identified from choice data.

How then can a folk psychologist (that is, each of us) determine the beliefs and preferences of others? Since one cannot infer beliefs or preferences from actions alone, and actions are the only source of evidence, it would seem that one could never reach any conclusions about beliefs or preferences. This conundrum is resolved by 'bootstrapping'. By assuming one knows some beliefs and preferences, one can make inferences from actions about other beliefs and preferences. Since initial assumptions can lead to trouble later, this bootstrapping is subject to empirical control. When economists observe someone's purchases at the grocery store or their actions in a laboratory for experimental economics, they assume that the subjects they observe share the economist's understanding of the alternatives, that the subjects

[3] 'To set or hold (one thing) before another in favour or esteem; to choose rather; to like better. Now the chief sense'. *Oxford Universal Dictionary* (Little *et al*. 1955, p. 1568). One might for example say, 'She preferred the brown ones' in answer to the question 'Which shoes did she choose?'.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

are not attempting to make fools of the observers or experimenters, and so forth. Against a background of shared beliefs and shared preferences, people can discover details of the idiosyncratic beliefs and preferences of others.

Economists are presumably well aware of the truism that choice depends on both belief and preference.[4] Yet there appears to be only one discussion in the literature of how this truism bears on revealed-preference theory.[5] Choice cannot reveal preference, since one cannot infer preferences from choices without premises concerning beliefs.

Revealed-preference theorists have not discussed this simple problem, because, explicitly or implicitly, they have supposed that one could hold beliefs 'fixed'.[6] Consider the following remarks of Kim Border's,

> This paper characterizes choice *behaviour* that is consistent with EUH [the expected utility hypothesis] ... The type of behaviour I consider here is limited to choices in risky situations where both the outside observer and the decision maker agree on the probabilities and the ranking of the prizes. That is, probability is objective and the prizes are monetary. (1992, p. 21)

Objective probability has no influence on choice unless it agrees with the subjective probability judgement of the agent, and monetary prizes provide a shared ranking only if one begins with a shared preference for more money rather than less. Border is showing when and how preference can be inferred from choice *and* belief. He is not offering a theory of how choice by itself reveals preference.

If one takes beliefs as given, there is relatively little difficulty in taking choice to 'reveal' preference. But this fact is of no comfort to the revealed-preference theorist, because this is not the relevant sort of revelation. This revelation is uncontroversial and fully consistent with

[4] Consider these remarks of I. M. D. Little, 'The verb, "to prefer" can either mean "to choose" or "to like better," and these two senses are frequently confused in economic literature. The fact that an individual chooses A rather than B is far from conclusive evidence that he likes A better' (1949, pp. 91–2). Little goes on to argue that although 'preference' in the sense employed in folk psychology cannot be inferred from choice, economists can confine themselves to choice or to a notion of preference that can be revealed by choice alone. This view is addressed in the next two sections. I have, however, encountered economists who claimed not to understand what I meant when I spoke of preferences as mental states that, along with beliefs, are responsible for choices.

[5] Rosenberg 1992, p. 123. Sen's discussion of the ways in which consistency of choice depends on how agents conceive of their choices (1993, pp. 501–2) makes essentially this point, although not in these terms.

[6] This is particularly clear in Stanley Wong's book-length treatment of Samuelson's revealed-preference theory. He emphasizes that choices reveal preferences only given assumptions 'about the nature of the price-income or choice situation that the consumer faced' (1978, p. 88). Without certainty these will be assumptions about beliefs.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

folk psychology. It does not support the claim that choice defines preference.

3. PREFERENCE AND 'PREFERENCE*'

If *C(S)* is, as it is intended to be, the set of alternatives the agent chooses from *S*, and *R* is, as it is intended to be, the weak preference relation, then the revelation theorem shows that choices reveal preferences. The revelation theorem really is a theorem, and choices do not reveal preferences (as preferences are understood in folk psychology). Hence one must reject the interpretation of *C(S)* or of *R*. There is nothing obviously objectionable about either of these interpretations taken by itself. But they cannot be maintained simultaneously. The arguments in the previous section provide decisive reason to reject the interpretation of *R* as weak (subjective) preference.

There is, however, no reason why economists cannot define their own technical terms, and the fact that some term in economics does not mean exactly the same thing that it means at the corner pub is no criticism of economics. For example, the fact that people distinguish at the pub between preference and duty is no reason that economists cannot employ a notion of a preference ranking that derives from all the different factors – including duty – that may prompt individuals to act. When economists employ such a notion, they should of course be careful to avoid lay misunderstanding of their theories. They should, for example, warn that the fact that they take preferences to predict actions does not imply that they deny that considerations of duty influence people. More importantly, when economists assign technical meanings to terms, they must also avoid equivocating. Uses of their theory must not presuppose the corner-pub meaning of the term, which economists claim to have repudiated.

Suppose one grants the intended interpretation of *C(S)* and takes the revelation theorem as defining a special technical notion of 'preference'. To avoid equivocation, let us call this technical notion 'preference*'. Although it is defined in terms of choice, preference* may nevertheless correspond to a notion that is implicit in folk psychology. Notice that belief can be as it were embedded in preference. For example, one cannot predict why an investor purchases one stock rather than another merely from knowing that she wants larger returns. But her desire for larger returns and her beliefs about the returns from stocks may give rise to a preference for a particular action, a particular purchase. The connection between this choice and the 'final preference' – the preference for just *this* choice, *this* action – is not mediated by belief. Beliefs play their full role in generating final preferences. Preferences for actions may be like this in Savage's decision theory.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

Perhaps preference* – the technical notion of preference implicitly defined by the revelation theorem coupled with the interpretation of choice sets – captures the notion of 'final preference'. Perhaps – but can economists dispense with the notion of subjective (non-final) preference? To vindicate revealed-preference theory, economists must show that the notion of preference* will suffice. Can they make do without preference?

Those economists who want to explain or predict preferences* in terms of agents' utilities and subjective probability judgments cannot dispense with the notion of preference. For it is only preference (not preference*) that combines with belief to determine choice. Furthermore, if one eschews all talk of preference (or psychological cognates such as desire), then claims about beliefs or subjective probabilities are untestable and consequently scientifically illegitimate. The only evidence for claims about beliefs is action (including speech), but actions have no implications concerning beliefs without premises concerning preference. When one observes panic selling in the stock market, one infers that people believe that share prices are going to drop. This inference supposes that people prefer (in a subjective, folk-psychological sense) not to be poorer. Without some such premise concerning preference, no inference concerning belief could be made. If revealed-preference theory succeeded in banishing a subjective notion of preference from economics, it would also have to banish beliefs, expectations, and subjective probabilities.

The argument against revealed-preference theory could stop here. But the general difficulties can be made more definite by considering the difficulties that arise when game theorists attempt to make do without a notion of subjective preference.

4. WHY GAME THEORISTS NEED SUBJECTIVE PREFERENCES

Some game theorists have defended revealed-preference theory. For example, Ken Binmore writes:[7]

[7] Although Binmore claims to be a revealed-preference theorist, it is not clear whether he is. He writes, 'The preference [for *(hawk, dove)* over (*dove, dove*)] is deduced from the prior information that, if Adam knew he had to choose between only (*dove, dove*) and *(hawk, dove*), then he actually would choose (*hawk, dove*)' (1994, p. 106). In the context, Adam's knowing that he is choosing between (*dove, dove*) and *(hawk, dove*) requires his knowing what game he is playing and how the other player, Eve, has played. Here preference is defined in terms of a counterfactual concerning action given a state of knowledge. In fact, Adam's choice presumably depends on Adam's *beliefs*, not on whether his beliefs are true or whether they constitute knowledge. So Binmore seems to assume that to deduce Adam's preferences, one needs to know both what Adam does and what Adam believes. (Binmore points out, however, that one can use the sure-thing principle to argue that *some* beliefs may not matter (1994, p. 197)). But one does not need to know anything about belief to infer preference*. Beliefs link choices to subjective preferences, not to preferences*, which are defined by choices. If one takes choice to depend on belief, then,

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

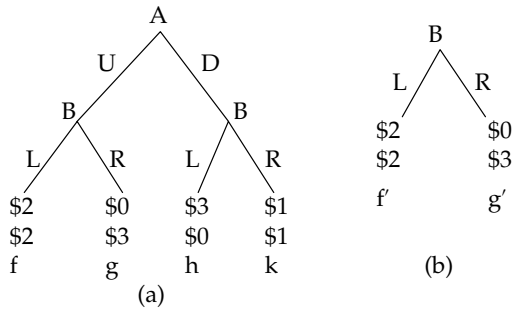


FIGURE 1

> Modern utility theory makes a tautology of the fact that action *b* will be chosen rather than *a* when the former yields a higher payoff by *defining* the payoff of *b* to be larger than the payoff of *a* if *b* is chosen when *a* is available. (1994, p. 169)

Binmore holds that once one specifies what game people are playing, the payoffs express how they would choose (1994, p. 164). Binmore sensibly insists that questions about how to play games should be sharply separated from questions about what games people are playing – that is, about how to model particular interactions among people. But when this salutary insistence is coupled with revealed preference theory, it turns out, I shall argue, that one has not specified explicitly what game people are playing until one has specified what strategy the players select. Revealed-preference theory thereby trivializes the tasks of advising players on what strategy to employ or predicting what strategies will be played.

Game theorists cannot make do with the notion of preference*, because they seek to advise players on what strategies to play or to predict what strategies will be played.[8] To do this, they need to show how the choice of strategies depends on preferences for outcomes. They are concerned with how 'final preference' is formed, and so they cannot accept a view that takes final preference as already given. For example, consider the strategic situation shown in Figure 1(a). *A* moves first and plays up or down, *U* or *D*. Then *B* plays left or right, *L* or *R* in response. *f*, *g*, *h*, and *k* are *B*'s utilities for the outcomes. The top number in each of the dollar pairs represents the payoff to *A*, and the bottom number the dollar payoff to *B*. Figure 1(a) specifies the game form, but until

as argued above, one needs a subjective notion of preference; and one cannot be a revealed-preference theorist.

[8] Game theorists may also want to *explain* choices, but since revealed-preference theorists define preferences* – and hence payoffs – in terms of choices, they must deny that explanation is a legitimate goal.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

something is known about *f, g, h*, and *k* and the corresponding utilities for *A*, no game has been specified, and game theory has nothing to advise or predict.

If individuals always seek the largest monetary payoff for themselves, then there is apparently no problem filling in the utilities. As in a prisoner's dilemma (of which this game is a sequential variant), *B* has a dominant strategy, and *(D, R)* is the unique Nash equilibrium.

Although it is reasonable in many contexts to suppose that choices depend on nothing but a player's own winnings, this is obviously not always the case. When individuals are brought into a laboratory for experimental economics and placed in situations such as the one depicted above, a sizeable minority do not play what appear to be dominant strategies. If *A* believes that *B* is the sort of player who will play *L* in response to *U* and *R* in response to *D*, then *A* should play *U* rather than *D*. If on the other hand *A* knows that $g > f$ and $k > h$, then *A* knows that *B* will play *R* regardless of what *A* does, and *A* should, of course, play *D*. But the problem is that *A* only knows the structure of the game and the financial payoffs. To know how to play (and for the game theorist to know how to advise *A* to play), *A* and the theorist need to know *B*'s preferences or preferences*.

There are four possible types of *B* players, each of whom plays one of *B*'s possible strategies. So the situation can be modelled as a game of incomplete information in which nature moves first to determine *B*'s type, then *A* chooses to play up or down, and then *B* finally moves and the dollar payoffs are determined. To know how to move, *A* needs to judge what type of player *B* is. Since the revealed-preference game theorist has eschewed all talk of subjective preference, a judgment of *B*'s type is, of course, a judgement of *B*'s preferences*.

Suppose one attempts to determine the preferences* by offering *B* choices among the two pay-off pairs separately. For example, as shown in Figure 1(b), one could present *B* with a choice between \$2 for both *B* and *A* and \$3 for *B* and nothing for *A*.[9] The problem is that *f* need not be the same as f' and *g* need not be the same as g'. There is no reason why it could not be the case that $f > g$ and $f' < g'$. Choosing *L* after *A* has already chosen *U* is not the same thing as simply choosing between two pairs of monetary payoffs. Experimental subjects in the role of *B* who play *L* because they want to repay a kindness or to reward a trust draw such a distinction. So choices in circumstances like the one depicted in Figure 1(b) do not reveal whether $f > g$.

Another way of making this point is to notice that it is misleading to

[9] Note that figure 1(b) depicts a separate degenerate game, not a subgame of the game whose form is given in Figure 1. In the corresponding subgame the (utility) payoffs for *B* may depend on how *A* played.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

state that ($2, $2) is the outcome that results when *A* plays *U* and then *B* plays *L*. The outcome is actually: (*A* plays *U*, then *B* plays *L*, then A and *B* each receive $2 from the experimenter), and *f* is the utility for *B* of this outcome. There is nothing in game theory or revealed-preference theory that tells us that *f* is the same as *B*'s utility when *A* and *B* both receive $2 in some other circumstance.

The fact that one cannot infer whether $f > g$ or $h > k$ by examining choices like those shown in Figure 1(b) does not imply that *A* or the revealed-preference theorist is unable to learn how *B* will play. For example, if *B* has played the same game before, *A* can draw conclusions concerning how *B* will play this time. Furthermore, *A* can make use of information about how others have played similar games to make inferences about what *B* will do.

Although it is possible for a revealed-preference theorist to gather evidence concerning what strategy *B* will play, the revealed-preference theorist cannot consistently compare *g* and *k* or *f* and *h*, because there is no way that *B* could ever face a choice between playing *L* when *A* has played *U* and playing *L* when *A* has played *D*. The impossibility is conceptual rather than causal. The existence of such a choice is logically contradictory. Any situation in which *B* faces a choice between the payoff pairs ($2, $2) and ($3, $0) or between ($0, $3) and ($1, $1) involves a different game, and a choice among these payoff pairs in a different game does not reveal a preference* between the outcome pairs resulting in this game from *(U,L)* and *(D,L)* or from *(U,R)* and *(D,R)*. Although the revealed-preference theorist thus cannot fully specify the game, the difficulty might seem unimportant, because the undefined preferences* of *B* will never affect the play. The revealed-preference theorist (and *A*) can predict *B*'s strategy, and no further information concerning *B*'s preferences is needed.

The true seriousness of the problems only becomes evident when one turns to *A*'s choice. In order to predict how *A* will play or to advise *A* on how to play, the game theorist needs to know *A*'s payoffs, which the revealed-preference theorist equates with *A*'s preferences*. Since *A* never faces a choice between the outcomes *of this game*, the revealed-preference theorist cannot (on pain of logical contradiction) assign preferences* over the *outcomes* to *A*. Since revealed-preference theorists define preference* in terms of choice, they must deny that there is a preference* where no choice is logically possible. There is no context in which *A* ever faces a choice among the outcomes of *this* game. So the revealed-preference theorist cannot offer any account of how *A*'s choice among strategies depends on *A*'s preferences* over outcomes. The only choice *A* faces in this game is whether to play *L* or *R*, and the only preferences* *A* has are for these moves.

The revealed-preference theorist could say that although the fact that

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

one cannot write in *A*'s preferences* for outcomes makes the extensive form of the game appear rather naked, the essential information needed to predict *A*'s choice is there. Since *A* chooses to play up or down, *A* has a preference* for up or down; and that preference* for up or down, combined with *B*'s preference* for left or right in response determines the outcome. But the revealed-preference theorist has now trivialized game theory. The only prediction or advice that the revealed-preference theorist can offer *A* and *B* is to take whatever action they prefer* – that is, to do what they do. To suppose that both players know the preferences* is to suppose that both players know what moves both are making – but then the task of deciding what strategy to play has been accomplished in specifying the game.

This gutting of game theory is inconsistent with the practice of game theorists, who would maintain that *A*'s choice should depend on *A*'s (subjective) preferences over the outcomes and *A*'s (subjective) expectations about how *B* will respond. Preferences, not preferences*, are those states of individuals that combine with beliefs (or subjective probabilities) to produce choice. Game theory is concerned with how choice or 'final preference' depends on beliefs and subjective preferences – that is, with precisely the phenomena that the revealed-preference theorist seeks to expunge from economics. Reference to subjective preferences is unavoidable, if game theory has any role to play in predicting or guiding the choice of strategies.

In response to this argument, a defender of revealed-preference theory might note that individuating objects of choice too finely renders all principles of rationality empty. For example, with an ultra-fine individuation of alternatives, choices will never be intransitive, since choosing x from the pair (x, y) is different from choosing x from the pair (x, z). The independence condition loses its teeth, too, since, for example, the choice in Figure 1(b) is different from the choice in Figure 1(a) in the circumstances where *A* has already played *U*. A revealed-preference theorist could object that the above critique depends upon an unreasonably fine-grained individuation of the objects of choice. Once they are individuated in some coarser way – as they must be – the conceptual criticism collapses.

This question of how to individuate the objects of choice has recently been discussed by a number of theorists (see particularly Broome, 1991 and Bacharach and Hurley, 1991, pp. 12–17), but only Rosenberg (1992, pp. 121–2) has commented on its relevance to revealed-preference theory. Some principle to individuate the alternatives among which the agent chooses is necessary before the set of alternatives *S* and thus the choice set *C(S)* are well defined. If one adopts the principle 'same dollar payoffs implies same object of choice' then the conceptual difficulty I pointed out above would indeed disappear. But if one individuates the

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

objects of choice this way, then one will find that game theory often makes false predictions and that it often gives bad advice. People do distinguish choices that involve reciprocating a kindness or repaying a trust (such as *B*'s choice of *L* after *A* has played *U*) from mere choices among dollar payoffs. Unless one refuses to face this fact, one cannot evade the conceptual criticism. Which choices are the same depends on how agents *conceive of* the choices, on what agents *believe*, and on what agents *value*. Alternatives cannot be individuated non-arbitrarily without reference to subjective beliefs and wants.

As the game depicted in Figure 1 illustrates, game theorists need to be able to talk about the dependence of strategy on belief. The point is particularly obvious in pure coordination games, but it is a general one. Consider the following normal form of a game in which Row and Column move simultaneously and independently:

Row	**Column** *left*	*right*
up	(1, 1)	(3, 0)
down	(0, 2)	(2, 3)

As game theory is ordinarily understood, the game theorist uses this normal form to predict how Row and Column will play, to advise them how to play, and to explain how they play, by treating choices as deriving from the payoffs *and* the players' beliefs. The players may use the normal form in the same way in deciding what to do. To use the normal form to decide what to do, the players have to *believe* that the numbers correctly predict what players would do *if* they had the relevant *beliefs*. For example, Row has to believe that if Column *believed* that Row had played *down*, then Column would play *right*. The numbers do not say that Column will play *right* if Row in fact plays *down*. How Column plays depends on what Column believes Row will do, not on how Row actually plays.

The numbers could not possibly represent preference* rather than preference. If they did, then Row would know that Column would play *left* if Row plays *up* and that Column would play *right* if Row plays *down*, and so Row should play *down*! This is a bizarre and perhaps even incoherent way to read a normal form, and it is obviously *not* the way game theorists read it. Game theorists would instead note that *up* is a dominant strategy, and that Row should play *up*. *Since strategy depends both on payoffs and beliefs, payoffs must represent preference instead of preference**. If the payoffs were indices of preference*, then they would say how individuals would choose. They would already incorporate the

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

influence of belief, and belief could play no further role. If the revealed-preference theorist were right and payoffs already represented what strategy was chosen, there would be nothing left for game theory to do.

The same point applies of course to expected utility theory. Whether I choose to carry my umbrella may depend on whether I think it is going to rain. To employ expected utility theory to advise me on whether to carry the umbrella or to predict whether I will carry it, the theorist needs to know both my subjective probabilities and my subjective preferences. The task is to determine what I will choose. Preferences* are defined in terms of choice and are given only when the task of determining how to choose is completed. If there is no room in the theory for subjective preferences, then expected utility theory can play no role in predicting, advising, or explaining.

5. SUBSIDIARY DIFFICULTIES

When one has made a case on strong grounds, it is dangerous to go on to weaker ones. The reader will wonder why the author is bothering, and doubts about the weaker *arguments* can easily turn into doubts about the *conclusion*. Yet, it is nevertheless worth briefly recalling other problems with revealed-preference theory.

The *raison d'être* of revealed-preference theory was philosophical. It was supposed to enable economists to rid economic theory of references to subjective preferences or to make those references respectable. Revealed-preference theory does not accomplish these objectives, and they are in fact misconceived. A reasonable empiricism does not demand that preference be defined by choices. It demands only that claims about preferences be testable. Because revealed-preference theory misstates the relationship between preference and choice, it does not show how to test claims about preferences, and it is of no use to a reasonable empiricist.

Furthermore, revealed-preference theory does not improve the empirical respectability of utility theory. Unless choice sets are singletons, observations of what is chosen from S do not establish the full contents of $C(S)$. If an individual P is purchasing bundles of commodities and P's beliefs about the alternatives are correct, then a choice of x reveals that x is in P's choice set. But even a repeated choice of x does not reveal that some other bundle y is not in the choice set. P might be indifferent between x and y. If P subsequently chooses y when x is available, there need be no violation of WARP (Wong 1978, § 5.3). One might say that P is indifferent between x and y only if P is equally likely to choose x or y (Harsanyi 1977, pp. 27–8). The rationale for this claim is obscure, but even if it is correct, the empirical determination of what belongs to an individual's choice set turns out to be an arduous process. And in the background lurks the problem that the agent's tastes may change.

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

Revealed-preference theory also impoverishes (1) the empirical theory of choice, (2) the normative theory of rationality, and (3) welfare economics.

1. We have already seen how it impoverishes the empirical theory of choice. If one wants to ask why people play certain strategies – as many game theorists do – then one cannot accept a view that denies that choice can be explained. *Diagnosing* behaviour is not merely indulging curiosity. It also helps economists to predict behaviour. If, for example, economists discover that what causes certain agents to violate WARP is a particular false belief, then economists can improve their predictions concerning how the agents will behave in the future.

2. One controversial question in the theory of rationality is whether to require that preferences be complete. Economists cannot ask this question if they accept revealed-preference theory. For then unless choices violate WARP, they automatically reveal a preference. Accepting revealed-preference theory creates spurious intransitivities in indifference, which arise from limits to people's ability to detect small differences.

3. If WARP is satisfied, revealed-preference theorists who identify welfare with the satisfaction of preference must say that people always choose what is best for them. It is implausible to identify welfare with the satisfaction of (subjective) preferences (Hausman and McPherson, 1996, Ch. 6) and even more implausible to identify welfare with choice. Furthermore, since choice does not reveal subjective preference, economists cannot say both. If, owing to a mistaken belief, an agent chooses y even though the agent prefers x, the agent is better off according to a choice theory of welfare and worse off according to a preference-satisfaction view. The difficulties are not just philosophical niceties. If economists accept revealed-preference theory, then they are committed to the implausibilities of an ex-ante interpretation of cost-benefit analysis (see Broome, 1991, Ch. 10; Hausman and McPherson, 1994). Furthermore, as Sen has emphasized (1977), revealed-preference theorists cannot consider whether individuals might sometimes choose (perhaps for moral reasons) something that they do not prefer.

Commitment to revealed-preference theory also seems to involve self-deception. The sense and point of the theory are parasitic on the subjective notion of preference that it supposedly disowns (Sen ,1973, pp. 242–4; Rosenberg, 1992, p. 120). Economists are interested in choices, which are intentional human actions, not unintentional movements; and to draw this distinction requires the notion of subjective preference that revealed-preference theorists disavow. Furthermore, as Sen has argued (1993), consistency conditions such as WARP can only be justified by an interpretation of choice informed by the objectives of the choosers.

Revealed-preference theory misstates the relations between preference and choice. Its purported empirical advantages are spurious, and

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

the theory impoverishes the theory of choice while presupposing the subjective notions it attempts to avoid.

6. CONCLUSIONS

Although economists are often more concerned with aggregate outcomes of individual choices than with individual choices themselves, they are stuck with a theory of individual choice. That theory builds upon folk psychology – that is, the explanation of individual choice in terms of, on the one hand, beliefs, and, on the other, of preferences, desires, commitments, values and so forth. Economists spruce up beliefs as subjective probabilities, and amalgamate an agent's desires, values, and commitments into a preference ordering. Since preference orderings can be represented by utility functions, folk psychology turns into a mathematical theory of utility maximization. In this theory, choices can be predicted, explained, criticized or justified in terms of the subjective beliefs and preferences of agents. Choice does not depend on belief alone, and it does not depend on preference alone.

Extreme empiricist economists might object: 'We cannot be respectable scientists if we refer to unobservable mental states. The ''operational significance'' of preference is choice. We must do away with mysterious mental preferences, and define preferences by choices. The revelation theorem shows how to do this'. But the preferences* that choices define do not do the work of preferences. They do not enable one to infer people's beliefs. They do not justify welfare conclusions. They do not enable one to predict, explain, or recommend choices. They close off questions in welfare economics, positive theory, and the theory of rationality.

There was never any reason to develop revealed-preference theory in the first place. There is nothing wrong with referring to subjective preferences and beliefs. There is nothing to be ashamed of in the notion of a subjective preference ranking, and it is no good pretending that economics can make do without it.

REFERENCES

Bacharach, Michael and Susan Hurley. 1991. 'Issues and advances in the foundations of decision theory'. In *Foundations of Decision Theory: Issues and Advances*, pp. 1–38. Michael Bacharach and Susan Hurley (eds.). Blackwell

Binmore, Ken. 1994. *Playing Fair*. MIT Press

Border, Kim. 1992. 'Revealed preference, stochastic dominance, and the expected utility hypothesis'. *Journal of Economic Theory*, 56:20–42

Broome, John. 1991. *Weighing Goods*. Blackwell

Gravelle, Hugh and Ray Rees. 1981. *Microeconomics*. Longman

Green, Edward and Kent Osbard. 1991. 'A revealed preference theory for expected utility'. *Review of Economic Studies*, 58:677–96

Harsanyi, John. 1977. *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press

Hausman, Daniel. 1992. *The Inexact and Separate Science of Economics*. Cambridge University Press

Hausman, Daniel and Michael McPherson. 1994. 'Preference, belief and welfare'. *American Economic Review, Papers and Proceedings*, 84:396–400

Hausman, Daniel and Michael McPherson. 1996. *Economic Analysis and Moral Philosophy*. Cambridge University Press

Henderson, James and Richard Quandt. 1980. *Microeconomic Theory: A Mathematical Approach*. 3rd edn. McGraw-Hill

Houtthaker, H. S. 1950. 'Revealed preference and the utility function'. *Economica*, 17:159–74

Little, I. M. D. 1949. 'A reformulation of the theory of consumers' behaviour'. *Oxford Economic Papers*, 1:90–99

Little, William, H. Fowler, and J. Coulson. 1955. *The Oxford Universal Dictionary on Historical Principles*. 3rd edn. Clarendon Press

MacKay, A. 1980. *Arrow's Theorem: The Paradox of Social Choice. A Case Study in the Philosophy of Economics*. Yale University Press

Rosenberg, Alexander. 1992. *Economics: Mathematical Politics or Science of Diminishing Returns?* University of Chicago Press

Samuelson, P. 1938. 'A note on the pure theory of consumers' behaviour'. *Economica*, 5:61–71

Samuelson, P. 1947. *The Foundations of Economic Analysis*. Harvard University Press

Sen, Amartya. 1971. 'Choice functions and revealed preference'. *Review of Economic Studies*, 38:307–17

Sen, Amartya. 1973. 'Behaviour and the concept of preference'. *Economica*, 40:241–59

Sen, Amartya. 1977. 'Rational fools'. In *Philosophy and Economic Theory*, pp. 87–109. Frank Hahn and Martin Hollis (eds.). Oxford University Press, 1979

Sen, Amartya. 1993. 'Internal consistency of choice'. *Econometrica*, 61:495–521

Wong, Stanley. 1978. *The Foundations of Paul Samuelson's Revealed Preference Theory: A Study by the Method of Rational Reconstruction*. Routledge

https://doi.org/10.1017/S0266267100000158 Published online by Cambridge University Press