Economics and Philosophy, **25** (2009) 1–25 Copyright © Cambridge University Press
doi:10.1017/S0266267108002253

PREFERENCE SATISFACTION AND WELFARE ECONOMICS

Daniel M. Hausman
University of Wisconsin-Madison

Michael S. McPherson
Spencer Foundation

The tenuous claims of cost-benefit analysis to guide policy so as to promote welfare turn on measuring welfare by preference satisfaction and taking willingness-to-pay to indicate preferences. Yet it is obvious that people's preferences are not always self-interested and that false beliefs may lead people to prefer what is worse for them even when people are self-interested. So welfare is not preference satisfaction, and hence it appears that cost-benefit analysis and welfare economics in general rely on a mistaken theory of well-being. This essay explores the difficulties, criticizes standard defences of welfare economics, and then offers a new partial defence that maintains that welfare economics is independent of any philosophical theory of well-being. Welfare economics requires nothing more than an *evidential* connection between preference and welfare: in circumstances in which people are concerned with their own interests and reasonably good judges of what will serve their interests, their preferences will be reliable indicators of what is good for them.

It is obvious that people sometimes make bad choices. (At least it is obvious to us that we do.) People often prefer what is bad for them. It requires no philosophical acumen to recognize human fallibility. An uncharitable observer might consequently conclude that economists are

Although organized and expressed quite differently, many of the ideas of this paper can be found in chapter 4 of our *Economic Analysis, Moral Philosophy, and Public Policy*. During the long period in which the ideas here germinated, many readers and audiences have helped us, and we are no longer able to acknowledge all these helpers individually, apart from Philippe Mongin, who gave us insightful suggestions on the penultimate draft.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

dunces, for mainstream normative economics ("welfare economics") apparently identifies welfare with the satisfaction of preferences. The tenuous claims of cost-benefit analysis to guide policy so as to promote welfare turn on measuring welfare by preference satisfaction and taking willingness-to-pay to indicate preferences. So it seems there is a problem. Does welfare economics rest on an elementary philosophical blunder?

Although economists are as prone to philosophical confusions as anybody else, this essay argues that welfare economics does not rely upon a false theory of welfare. At the same time, we argue at some length that it is a mistake to regard welfare as deriving from preference satisfaction. In other words, this essay attempts both to refute preference satisfaction theories of welfare and to defend welfare economists. What makes these two aims consistent is the possibility that welfare economics does not presuppose a preference satisfaction theory of welfare. Indeed we shall argue that welfare economics does not rely on any theory of welfare. When certain conditions are met, it is reasonable to take preference satisfaction as indicating welfare *regardless of what theory of welfare one accepts*. In laying out this somewhat paradoxical case, this essay will offer guidance concerning how to judge when the tools of standard normative economics can justifiably be applied to specific problems.

Here is a brief road map: Section 1 presents alternative theories of welfare and notes their inadequacies. Section 2 is concerned with preferences and explains why a preference satisfaction theory of welfare appeals to economists. Section 3 summarizes many well-known objections to preference-satisfaction theories of welfare, with particular emphasis on the problems posed by mistaken beliefs and by preferences that are not self-interested. Section 4 discusses four bad reasons why economists might nevertheless identify welfare and preference satisfaction. Section 5 presents a more plausible but ultimately unsuccessful defence of the identification of welfare and preferences, which rests on the claims that welfare is the satisfaction of properly spruced-up preferences and that actual preferences are proxies for spruced-up preferences. Section 6 argues that this defence fails, because welfare is not the satisfaction of preferences, no matter how spruced up. Section 7 then defends an evidential connection between preferences and welfare: if individuals seek to benefit themselves and are good judges of what is good for them, then economists can rely on information concerning preferences to judge what makes people better off. Section 8 explores implications of this view for the practice of cost-benefit analysis and concludes.

1. WELFARE

In speaking of someone's welfare or well-being, we are talking about how good their lives are *for them*, about how well their lives are going. Exactly

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

what is good for a particular agent, Ann, will depend on Ann's character, ability, tastes and circumstances, and what is good for Ann may be very different from what is good for Ben. Most of the differences between what makes Ann's life go well and what contributes to Ben's well-being concern *instrumental* goods – things that are good for people because they are *means* to something else. If one focuses on *intrinsic* goods – things that are good for individuals without regard to their consequences – then there is less variation from individual to individual. Bifocals may be good for Ann and bad for Ben, even though clear vision is good for both. Instrumental goods, like bifocals, are good only if the ends to which they conduce are good. If nothing were good for people without regard to its consequences with respect to something else, then nothing could be good as a means to some other end. There must be intrinsic goods in order for there to be instrumental goods. A central problem of moral philosophy has been to determine what things are intrinsically good for human beings. Thus Aristotle, for example, held that happiness was the sole intrinsic good.

In everyday conversation, as in moral philosophy, one finds conflicting theories of well-being. To many who are deeply religious, the ultimate good lies in a relationship with the divine. Many people have believed that only mental states are intrinsically good, but there is less agreement here than it seems, because there are so many different views concerning which mental states are intrinsically good. Jeremy Bentham maintains that well-being is pleasure (1789). John Stuart Mill disagrees (1861). He asserts that welfare is happiness. Henry Sidgwick argued for the hybrid view that welfare is any mental state that is intrinsically desirable (1901).

Few contemporary philosophers defend mental state views, though economists and psychologists such as Kahneman (2006), Krueger (Kahneman and Krueger 2006), and Layard (2005) are tempted by them. If Ann mistakenly believes that she is loved by her family and friends and that her projects have been successful, while Ben correctly believes just the same things, then their mental states may be the same. But it seems that Ben is better off.[1] Most philosophers conclude that welfare is thus at least in part objective, and several have attempted to list the objective goods whose possession makes life go well (notably Griffin 1986, 1996). But it is doubtful that the lists for different individuals coincide, and it seems that some further theory is needed to explain why items appear on Ann's and Ben's lists of intrinsic goods and why their lists differ. To invoke an individual's preferences seems a natural move. Although welfare and preference satisfaction do not coincide, many contemporary philosophers

[1] This argument is made by many including Griffin (1986: 9–10) and Parfit (1984: 494). Nozick's "experience machine" thought experiment makes the same point in a more extreme way (1974: 42–5).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

are wedded to the view that well-being consists in the satisfaction of self-interested "informed", "rational," or "laundered" preferences (Arneson 1990; Gauthier 1986, ch. 2; Goodin 1986; Griffin 1986).[2]

The theory of well-being is an unsettled area of philosophy, whose many theories all face serious difficulties. Yet at the same time it is important not to forget how much agreement there is concerning which things are generally intrinsically good for people and which things are generally bad for them. Even though there is no consensus on a philosophical theory that explains why loyal friends, robust health, and a ready sense of humour are good for people, there is little doubt that these are elements of a good life for almost everybody. Nor is there much doubt that violent rages, alcoholism, or severe pain are not elements of a good life. For the most part economists would like to avoid rather than to tackle the philosophical difficulties concerning welfare, and this essay will show how they can do so.

2. WELFARE AND PREFERENCES

Normative economists typically link welfare to preference satisfaction. Although preferences are central to mainstream positive economics as well as to normative economics, economists have not had much to say about what preferences are. Sometimes people's preferences are identified with their judgments concerning what is in their interests (Sen 1973: 67; 1977: 329). Many economists are attached to revealed-preference theory, which identifies preferences with choices or hypothetical choices (Samuelson 1947: ch. 5; Binmore 1994: 104–7). Although neither of these construals of preferences is defensible,[3] we need not enter into the details of those debates. This essay will rely only on those properties of preferences that are implied by the substantive core of ordinal utility theory.

Although there are different ways to formulate ordinal utility theory, all the formulations imply the following four axioms concerning preferences:

1. An agent's preferences are complete – that is, for any two alternatives x and y the agent either prefers x to y or y to x or is indifferent between x and y.
2. An agent's preferences are transitive – that is, for any three alternatives x, y and z, if the agent prefers x to y and y to z, then the agent prefers x to z and similarly for indifference.

[2] These authors disagree, however, about the ways in which they would purify or purge actual preferences. Griffin's text is not unambiguous, and he has denied that he meant to espouse any version of a preference satisfaction theory.

[3] For some criticisms, see Sen (1971, 1973) and Hausman (2000, 2005).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

3. An agent's preferences are context independent – that is, whether an agent prefers x to y does not depend on what other alternatives are available.
4. An agent's choices depend on his or her preferences – that is, if an agent knows which alternatives are feasible (which alternatives can be chosen), then (a) an agent chooses x only if x is feasible, and no alternative that the agent prefers to x is feasible, and (b) if there is some feasible alternative x that the agent prefers to all others, then the agent chooses x.

The ordinal representation theorem proves that when people's preferences satisfy these axioms and further technical conditions, then they can be represented by a continuous utility function that is unique up to a positive monotone (that is, order-preserving) transformation (Debreu 1959). "Utility" merely indicates preference ranking. It is not something people seek.

Axiom 4, which is often described as utility maximization, links preference and choice. It implies that preferences are not just judgments. They are instead intrinsically motivating. Since ordinal utility theory does not make much sense apart from the special idealized context of perfect knowledge, the axiom is often stated without any explicit reference to the agent's knowledge about the availability of preferred alternatives, though no one denies that agents may choose inferior alternatives when they are unaware of the availability of better ones.

The first three axioms have two other important implications concerning preferences. First, they imply that preferences, unlike mere wants or wishes, are comparative. Second, insofar as utility theory is used to explain behaviour, an agent's preferences among alternatives must reflect the agent's adjudication of *all* the considerations bearing on the evaluation of the alternatives or choice among the alternatives. How else could an agent's preferences be complete, transitive and context independent? A complete and transitive preference ranking is clearly a difficult intellectual achievement. It is the outcome of an unmodelled process of evaluation.

Despite the demands of completeness and transitivity, the substantive axioms of ordinal utility theory are usually regarded as very weak. They leave open, for example, whether preferences can be identified with choices in the way that revealed-preference theorists propose. In a formal treatment, more needs to be said about what constitute the "alternatives" among which agents have preferences; but we shall leave their nature unspecified here.

Since it is reasonable to assume that choice has some connection with welfare, it is natural that mainstream economists would take what they employ to explain choices to constitute well-being. So one

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

finds that economists regard levels of utility (preference satisfaction), which they employ to *explain* choices, as measures of well-being for evaluative purposes. If individuals are exclusively self-interested, then unless they are irrational, they will prefer x to y if and only if they believe that x is better for them than y is. If they are perfectly well-informed, then their beliefs will be true, and x is better for them than y if and only if they prefer x to y. Accustomed as they are to the idealizing assumptions that individuals are self-interested and perfectly well-informed, economists may be tempted to take well-being to be the satisfaction of preferences. In applied work, economists may rely on more objective measures of "real income" rather than on utility measures, but they typically regard real income as an imperfect proxy for preference satisfaction.

3. WHY WELFARE IS NOT PREFERENCE SATISFACTION

There are serious and well-known objections to a preference-satisfaction view of well-being. The first is that people care about more than their own well-being. People are sometimes altruistic and all too often malevolent. People sometimes sacrifice their own well-being to benefit others or to do harm to those they hate. Satisfying preferences sometimes has no bearing at all on one's welfare. Derek Parfit gives the following compelling example:

> Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire Fulfillment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory. (1984: 494)

What Parfit calls "the unrestricted desire fulfilment theory" says that, other things being equal, satisfying any of Ben's desires makes Ben better off. If the stranger recovers, Parfit's preference is satisfied. But this does not in any way make him better off. So the unrestricted desire fulfilment theory has been refuted.

A second objection arises from mistakes in people's beliefs. Even when entirely self-interested, people often prefer things that are bad for them because they mistakenly believe them to be beneficial. If the promises many religions make of an afterworld are false, then sacrifices of well-being in this world for benefits in the next are in vain. The dependence of preferences on beliefs leads to practical problems and has given rise to controversies concerning how to apply cost-benefit analysis (Hausman and McPherson 1994).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

Although these two problems of mistaken beliefs and non-self-interested goals are the most important objections to identifying welfare and preference satisfaction, we shall comment briefly on four subsidiary difficulties.

First, if welfare is the satisfaction of preferences, what should one say when preferences change (Gibbard 1986)? Should one care about satisfying preferences that a person no longer has (Brandt 1979: ch. 13; Parfit 1984: ch. 8)? Suppose that as a little girl, Ann had longed to be a flight attendant. Is there any reason now to satisfy her childhood desire? At this point in her life, she thinks not. Why not? Why should her current desires take priority over her past desires? There are two plausible answers, but both reject the identification of welfare and preference satisfaction. First, she knows more about herself and about the attractions of various careers. So her current preferences are more likely to track what will really make her life better. But this answer supposes that there is some notion of what makes her life go well for her other than what will satisfy her preferences. Second, giving Ann what she no longer wants is not likely to make her happy. This answer, too, implicitly surrenders the view that well-being depends on preferences. Preference changes pose practical as well as theoretical problems. Should the consequences for welfare of a policy that changes people's preferences be measured by people's preferences before the policy is implemented or by their preferences afterwards? Should one aim to satisfy existing preferences or to shape preferences to make them easier to satisfy?

Second, what should one say when there are conflicts in the person's preferences? The axioms of ordinal utility theory rule out conflicts among preferences, but that's a fact about the axioms, not a discovery about reality. There are often conflicts within a person's preferences, including tensions between "first-order" preferences, such as a preference for smoking and "second-order" or "metapreferences" such as a preference that one not prefer to smoke (Frankfurt 1971, Sen 1977). Such conflicts can give rise to internal struggle, feelings of regret and deprivation, and apparently inconsistent behaviour, such as purchasing cigarettes but then locking them away or flushing them down a toilet – and then running out to purchase more (Schelling 2006). If Ben has two or more preference rankings which conflict with one another in their ranking of x and y (which is, of course, excluded by the axioms), then on a preference-satisfaction view of well-being, there seems to be no way to say whether x is better for Ben than y.

Third, if one combines the view that welfare is preference satisfaction with views that take social policy to be concerned with the quantity or distribution of welfare, one is led to implausible moral conclusions in the case of expensive tastes, anti-social preferences, and preferences that result from manipulation or distorting psychological processes (Elster 1983). On

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

the preference-satisfaction view, one makes Ann better off by satisfying her preferences regardless of how idiosyncratic or obnoxious they are and regardless of how they were formed. If welfare is what matters to social policy, then social policy should be insensitive to the character and origins of preferences.

Consider those who have expensive tastes. If welfare is preference satisfaction, then a person who has cultivated a taste for "prephylloxera claret and plover's eggs" (Arrow 1973: 254) without an income that makes them affordable is worse off than someone with a similar income who wants only affordable beans and frankfurters (see also Ronald Dworkin 1981). Should social policy be responsive to expensive preferences such as these (Arneson 1990)? If one refuses to get all choked up about failures to satisfy expensive preferences, then either one is not concerned exclusively with the amount and distribution of welfare, or one does not accept a preference-satisfaction view of welfare.

Racist, sadistic and other anti-social preferences raise related problems. One reason why it is bad to satisfy them is that doing so frustrates other (and characteristically stronger) preferences. But should their effects on others be balanced against the "benefits" of satisfying them? Should anti-social preferences count at all (Harsanyi 1982/1977: 56)?

A final implausible moral implication concerns preference formation. People's preferences are sometimes the result of previous coercion or manipulation, and people may come to prefer things as the result of problematic psychological mechanisms. Some people, for example, want things precisely because they cannot have them ("the grass is always greener on the other side of the fence"), while other people spurn what they think is beyond their reach, like the fox who judged the unobtainable grapes to be sour (Elster 1983; Sen 1987, 1990). For example, women who are systematically denied roles in public life or equal shares of consumption goods may learn not to want these things. Women who have been systematically oppressed may not have strong preferences for individual liberties, the same wages that men earn, or even for protection from domestic violence. But liberties, high wages and protection from domestic violence may make them better off than giving them what they prefer. It seems that those who are concerned with welfare need to consider not just preferences, but the *origins* of preferences or the *justifiability* of preferences. Satisfying preferences that result from coercion, manipulation or "perverse" preference formation mechanisms may not make people better off. These suspect preferences sometimes reflect false beliefs, and some of the problems here are versions of problems we have already discussed. But coercion and manipulation may also distort the psychological mechanisms that lead people to value some things more than others. When the preferences of oppressed people derive from their

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

oppression, one cannot measure their welfare by considering how well their preferences are satisfied.[4]

4. FOUR BAD REASONS TO IDENTIFY WELFARE AND PREFERENCE SATISFACTION

The two main objections to taking welfare to be preference-satisfaction are not hard to see. First, what people want may have no bearing on their welfare, regardless of whether their beliefs are correct or mistaken. Second, even if they aim to benefit themselves, they may be mistaken about how to do so. Yet many economists continue to espouse the preference-satisfaction theory of welfare. Why?

Later in section 6, we shall give a good reason. But let us first consider four bad ones, which are unfortunately not unfamiliar:

1. "To maintain that welfare consists in preference satisfaction involves less controversial philosophical commitments than substantive theories about what is good or bad for people. By leaving the question of what is good for an individual up to the individual, normative economists show their philosophical modesty." Response: There is nothing philosophically modest about the theory that satisfying preferences constitutes welfare. This claim is an ambitious and false philosophical thesis.
2. "Rejecting a preference-satisfaction theory of welfare opens the door to unacceptable paternalistic policies. Since economists generally oppose paternalism, they should identify welfare with preference satisfaction." Response: If welfare is preference satisfaction, then the question of whether paternalistic policies are justified cannot arise. If people were never mistaken about what is good for them, it would be impossible to coerce individuals for their own benefit. But it is possible to coerce individuals for their own benefit. For example, suppose someone is about to step into an open manhole, and the only way you can protect them is physically to push them aside. In doing so, you are coercing them for their own good. Serious arguments against

[4] Heathwood (2005) responds to objections based on false beliefs or distorted or manipulated preferences by pointing out that when desires are "defective", and their satisfaction is consequently bad for someone, it is typically the case that satisfying them frustrates some other and more important desire. If Ann quenches her thirst by drinking water that is, unbeknownst to her, polluted and that thus makes her sick, she frustrates her desire to be healthy. In terms of her complete, transitive, and context-independent preference-ranking, she prefers (*ex post*) not drinking the water to drinking it. Whatever the merits of this partial defence of an actual desire-satisfaction view of well-being, it provides no comfort to welfare economists who take the satisfaction of her *ex ante* preferences as increasing rather than decreasing her expected well-being.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

paternalism, such as John Stuart Mill's, give reasons why it is generally better not to coerce people for their own benefit, rather than denying that it is possible to do so.[5]

3. "Welfare is a desirable mental state, such as happiness. Because what better satisfies preferences typically makes people happier, one way to implement a social policy that aims to make people happy (and which at the same time avoids paternalism) is to satisfy preferences. For example, in influential recent work on happiness, Kahneman maintains that, 'the correspondence of experience utility and decision utility is commonly taken for granted in treatments of choice' (2006: 501)." Response: Mental state theories are inconsistent with what economists say about welfare, and they are also indefensible. It is nevertheless easy to equivocate on the word "utility", sliding from interpreting it as an index of preference satisfaction to a substantive desirable mental state such as happiness. Economists often talk about individuals "seeking" utility, which makes no sense if utility is just a measure of the extent to which preferences are satisfied. And it is easy to equivocate again on the word "satisfaction". To satisfy a preference is for a preference to come true. When one knows that one's preferences have been satisfied, one may feel satisfied. But there is only this contingent connection between the satisfaction of a preference and the satisfaction of a person. The satisfaction of preferences is like the satisfaction of degree requirements. It has no necessary connection to any *feelings* of satisfaction. Parfit's preference that the stranger be cured is satisfied if the stranger is cured, even though Parfit never finds out and cannot *feel* any satisfaction at the event.
4. "If one accepts the standard view of rationality and the standard idealizations of positive economics (concerning self-interest and knowledge), then what people prefer necessarily coincides with what is good for them. Within the theoretical world depicted in standard economic models, welfare is preference satisfaction. Economists recognize that this world is not the real world, but they might regard the differences between theory and reality as mere noise." Response: This line of thought will not convince anyone who does not see the world the way mainstream economists do, but it helps explain why these economists are content to identify well-being with the satisfaction of preferences. To say that *if* people are rational, self-interested and well-informed, then what satisfies their interests will be what is good for them is not very impressive without some argument in defence of the antecedent as at least a reasonable approximation.

[5] Mill's argument against paternalism appears mainly in chapter 4 of *On Liberty* (1859). The example derives from Mill's example in chapter 5 of preventing someone from crossing an unsafe bridge. See also Dworkin (1971).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

5. WELFARE AND LAUNDERED PREFERENCES: THE APPROXIMATION RATIONALE

It is indefensible to maintain that welfare is a matter of satisfying actual preferences, since these need not be self-interested and may rest on mistaken beliefs. Yet many distinguished contemporary moral philosophers believe that there is something right about the attempt to link welfare and preference.[6] Probably the most common view of welfare is that it consists in the satisfaction of suitably "spruced-up" preferences. For example, Peter Railton suggest that "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free from cognitive error or lapses of instrumental rationality" (1986: 16). If well-being is the satisfaction of self-interested informed preferences, then identifying welfare with the satisfaction of actual preferences will be reasonable insofar as actual preferences approximate spruced-up preferences.

It seems more plausible to maintain that well-being is the satisfaction of suitably "laundered" self-interested preferences than to maintain that it is the satisfaction of actual preferences. This view grants that people may prefer something that is bad for them because of ignorance or cognitive defects, or because of concern to benefit or harm others. Such preferences would not be in the relevant sense "rational", "informed", or "self-interested", and so satisfying them would not make the person better off. Furthermore, satisfying preferences that result from coercion and manipulation do not count as benefiting a person, if, but for ignorance or irrationality, the person would not have had such preferences. When actual preferences change, there need be no puzzle about how to make the person better off, if the person's rational preferences have not changed. Conflicts among actual preferences are no problem, if there would be no conflicts among laundered preferences. One might also maintain (although this is more controversial), that racist, sadistic, or other anti-social preferences are not rational, and so those who are concerned to benefit people can discount such preferences.

The view that welfare consists in the satisfaction of properly pruned and purified preferences thus apparently resolves most of the difficulties facing the actual preference-satisfaction view. Unfortunately, it also makes life more difficult for normative economists. It makes an individual's state of well-being even less measurable than does the actual-preference

[6] "Versions of the desire theory now define the orthodox view of the nature of welfare, at least in the Anglo-American philosophical world. In the theory of rational choice the equation of well-being with utility (preference-satisfaction) has achieved the status of an unquestioned axiom, while in ethics its more prominent recent defenders have included Brian Barry, John Rawls, R.M. Hare, James Griffin, and Joseph Raz" (Sumner 1996: 122).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

satisfaction view. Not only does the welfare economist face the problem of determining what people prefer, he or she now needs to determine to what extent actual preferences coincide with pruned and purified preferences. Later we shall argue that normative economists are indeed stuck with these problems, but not because welfare is the satisfaction of informed and self-interested preferences.

The problem with the approximation view – that is the view that actual preferences are good proxies for welfare because actual preferences are good proxies for spruced up preferences – does not lie in practical difficulties. The problem instead is that the fact that Ben prefers x to y does not make it the case that x is better for him than y, no matter what conditions one imposes on his preferences.[7] Consider the requirement that preferences be self-interested, which is introduced so as to avoid the implausible implication that satisfying Parfit's preference that the stranger be cured makes Parfit better off. According to Parfit, the "Success Theory" of welfare (in contrast to what he calls the "Unrestricted Desire Fulfillment Theory") takes well-being to be the satisfaction of preferences "about our own lives". But which of Ben's preferences are about his own life? Parfit maintains that if, unbeknownst to me, my child becomes a petty thief, that is bad for me, because it frustrates my desire to be a good parent, and that desire is about my own life (1984: 495). On the other hand, Parfit argues that my desire to have healthy children is not about my own life. If while I am in exile, Ben's daughter dies and he does not know it, her death does not make Ben worse off (1984: 494). Clearly it is not going to be easy to say which preferences are "about" oneself and which are not. Mark Overvold (1984) proposes that self-regarding preferences are preferences concerning states of affairs that entail the agent's existence.[8]

But the distinction that is needed is not between Ben's preferences that are about Ben and Ben's preferences that are not about Ben. If Ben prefers for some reason or other to harm himself, then his preference is certainly about himself, but satisfying it by definition will not make him better off. The distinction that is needed lies between those of Ben's preferences that are directed toward promoting his own well-being and those that are not. But that distinction cannot be drawn if well-being is a matter of satisfying preferences (Sumner 1996: 135).

The underlying problem is that preference satisfaction has by itself nothing to do with well being. Suppose Ann prefers alternative x to y

[7] If one maintains that it is rational to prefer whatever is good for oneself (in terms of some other account of what it is for something to be good for people), then it will be true that Ben will rationally prefer x to y if and only if x is better for him than y is. But the fact that Ben prefers x to y has no influence on whether x is better for him.

[8] For criticism of Overvold, see Sobel (1998: 266–9). Griffin is sceptical about the possibility of distinguishing self-regarding from non-self-regarding preferences (1986: 23–4).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

and that this preference is as rational and well-informed as anyone could demand. Why should one conclude that Ann would be better off in the state of affairs where x obtains than she would be in the state of affairs where y obtains? One answer is that Ann finds out whether x or y obtains and is both pleased with x (her anticipation of her enjoyment may have been the reason why she preferred x to y) and pleased that her preference is satisfied. But in that case what makes her better off is not the fact that her preference is satisfied, but the pleasures she experiences because x obtains. Recall that for a preference to be satisfied is for the world to be as the agent prefers. Whether or not the satisfaction of a preference contributes to the agent's *feeling satisfied* is a separate question (Kraut 2007: 98–9).

Suppose that when Ann finds out that her preferred state of affairs obtains, she is not pleased or that x's obtaining has no effect on Ann's mental state. Are there still any grounds to conclude that the fact that x obtains implies that (*ceteris paribus*) Ann is better off? How does Ann's preference ranking by itself make the fact that x obtains good for Ann? Richard Kraut puts the challenge in the following way:

> More precisely: is it the case that whenever S wants (rationally, with proper information, and reflectively) P to occur, and P does occur, that is good for S simply *because* it is something he wants in that way?
>
> . . . The hybrid theory [that takes the satisfaction of desire to be only one of the sources of welfare] says that in some cases our desires should reach out to objects because they are good for us to have; but in other cases our desires just happen to reach out to objects that are not in themselves worth wanting, and when they motivate us to acquire those objects [and we succeed in doing so], that is good for us. (2007: 118)

Consider again Parfit's example of his preference that the stranger be cured. This seems to be a case where satisfying a preference does not increase well-being. Should one conclude, as Parfit and others have, that his is the wrong sort of preference, or should one conclude instead that the fact that some state of affairs satisfies Parfit's preferences does not by itself contribute to Parfit's well-being? The fact that his preference that the stranger be cured intuitively has nothing to do with him is, we suggest, irrelevant. For example, suppose that Ben learns that some people are chimeras – that is, their cells trace back to two or more separate fertilized eggs that fused into a single embryo – and he forms a preference that he be a chimera.[9] Clearly Ben's preference is about himself. Suppose further that the desire to be a chimera is not an idle wish, and Ben

[9] This is a real, not a hypothetical phenomenon. See for example the *New Scientist*, vol. 180, issue 2421 (15 November 2003), p. 34.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

goes to considerable trouble and expense to arrange for genetic tests on a sampling of his organs, which show that they all trace back to the same egg. Ben is disappointed to learn that he is probably not a chimera. Unbeknownst to Ben, however, he is in fact a chimera; his untested spleen traces back to a separate fertilized egg. Is there any reason to believe that the fact that Ben's self-regarding preference is satisfied enhances his well-being, that it makes his life in any way better? The fact that a state of affairs satisfies Ben's preferences does not make the state of affairs good for Ben.[10] We conclude that preference satisfaction theories of well-being, whether the preferences in question are actual or informed, are mistaken.

Another way to make this point is to ask what moral "pull" satisfying Ann's preferences should have on Ben. Thomas Nagel has argued that if something is valuable to people *only* because they want it, then their getting it has *no* direct moral importance for others (1986: 9). Ben has no reason to satisfy Ann's preferences unless he can make sense of why what she wants is worth wanting, or why her life will be better in some substantive way if she gets what she wants. For example many Boston residents desperately wanted the Red Sox to win the World Series in 2004. Their happiness when the Red Sox won gave others – even Yankees fans – some reason to want their preferences to be satisfied. But, as Nagel maintains, the mere preferences of Red Sox fans, as opposed to their happiness or unhappiness, should be of no moral importance to others. Since the welfare of others, unlike the satisfaction of their preferences, is of moral importance, it follows that the fact that Ann prefers x to y does not make it the case that (other things being equal) x is better for Ann than is y.

Views that take welfare to be the satisfaction of self-interested rational preferences seem more acceptable in part because they provide a foot in the door for arguments concerning substantive goods, goods that ought to be preferred (Railton 1986; Griffin 1986: 30). To take well-being to be the satisfaction of informed self-interested rather than actual preferences shifts the emphasis from what people prefer to what it is rational to prefer in order to benefit themselves.[11] If it could be shown, for example, that

[10] It may be the case that it is better to be a chimera, because one's body is then compatible with more prospective organ donors. In that case the satisfaction of Ben's preference does improve his expectations, but not, of course, because it satisfies his preference.

[11] "The shift to informed desires also represents an important change in the role of desires as determinants of well-being. If a full appreciation of the ways in which my life would be changed if I could speak French well would lead me to have a strong desire to master that language, then it is likely both that I have reason to do this, and that doing it would contribute to my well-being. But what role does the desire that I would have play in making these things true?" (Scanlon 1998: 115).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

it is rational for everyone who is concerned with his or her own welfare to prefer happiness to unhappiness or virtue to vice, then a view of well-being as the satisfaction of rational self-interested preferences could derive from a substantive view of well-being as involving happiness or virtue, coupled with the view that it is rational to prefer what is better for oneself. But there is nothing in the fact that Ann prefers some alternative x to y that makes x better for Ann than y is.

How could philosophers have gone so wrong? As Kraut argues (2007: 66–94), to say that *G* is good for *S* is assert that a relationship holds between *G* and *S*. That relationship depends on the properties of both *G* and *S*. So, for example, sexual promiscuity is good for young bonobos, but arguably not for young humans (2007: 90). Something that is good for most people may be bad for Ann because of distinctive features of her character, environment, physical traits or history. For example, even though Griffin maintains that autonomy is valuable to everyone, "a particular person might be made so anxious by being autonomous that it would be best for him not to be" (1986: 58). Just as Ben's anxieties can help determine what is good for him, so can Ben's desires. The extent to which Ben gets what he wants bears on how sensible and effective he is, and those who are effective and sensible are, other things being equal, better off than those who are not.

As Joseph Raz (1984: ch. 12) and Thomas Scanlon (1998: 118–21) point out, there are many ways to live well, and all of them require choosing to pursue some things rather than others. If people make bad choices about what to strive for, success in their projects may make their lives worse. But there are many good choices, many worthwhile objectives to pursue. Succeeding in the pursuit of worthwhile objectives and projects makes one better off (other things being equal), and failing makes one worse off. So one component of Ann's well-being is her success in the worthwhile endeavours that she has chosen. For example, saving whooping cranes and saving Siberian tigers may have very similar consequences for most people. Both are, we shall suppose, worthwhile endeavours. The extent to which saving either contributes to Ann's welfare depends on whether either of these endeavours is Ann's. If Ann has devoted time and effort over a prolonged period to saving whooping cranes, then if whooping cranes are saved, and Ann's endeavours have helped to save them, then Ann's life has in one way been successful, quite apart from any pleasure she may take from what she learns about the effectiveness of her efforts. Whether Siberian tigers are saved, on the other hand, has nothing to do with whether Ann is successful in any worthwhile pursuit, because saving Siberian tigers has not been one of Ann's projects. So Ann's choice of projects, which in turn reflect Ann's preferences, certainly matters. It is because of Ann's preferences that saving whooping cranes became one of

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

Ann's projects. Since this project was also worthwhile, its success enhances the value of Ann's life. But what makes Ann better off is the success of a worthwhile project, not the satisfaction of a preference.[12]

6. PREFERENCES AND WELFARE: AN EVIDENTIAL VIEW

There is another connection between preference satisfaction and well-being, which can justify some applications of welfare economics.[13] Even though well-being is not preference satisfaction, there is an evidential connection between preferences and well-being. If people are more or less self-interested with respect to certain alternatives, then economists can use people's preferences to make inferences concerning what people believe will benefit them. And if it is also reasonable to suppose with respect to the policies being considered and their consequences that individuals are good judges of what will benefit them, then economists can use people's preferences as evidence concerning what in fact makes them better off. In this way welfare economists can defend their practice of making inferences concerning well-being from people's preferences without committing themselves to any theory of well-being at all. Preferences are sometimes good evidence for welfare.

Even though what satisfies Ann's preferences does not necessarily make her better off, Ann may be sufficiently self-interested and well informed that her preferences are the best *guide* others have to what is beneficial to her. What better way is there to determine what will benefit people? Legislators know less of Ann's circumstances than she knows, and they have a less tender concern for her well-being than Ann probably does. The judgments of legislators about how to make Ann better off are often likely to be worse than her own judgments. It is also *safer* to rely on people's preferences. The mistakes individuals make about their own good will be to some extent unsystematic, and legislators will not arrogate to themselves the power to substitute their judgment for people's own judgments.

Our claim is that when preferences are self-interested and people are well informed, then their preferences will be a good (though fallible) guide to what will make them better off. We are not arguing – and it is not true – that satisfying preferences will always fail to benefit people when these

[12] Thus Raz writes both "[I]t is generally the case that the value of various situations for a particular person depends to a large extent on his actual goals, as they are or will be throughout his life" (1984: 290) and "At no point do I wish to suggest that the fact that a person wants something is in and by itself a reason for action either for him or for others" (1984: 291n).

[13] Scanlon (1998: 116–18) suggests a related interpretation of the link economists forge between preference and welfare, which helped to form our thinking.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

conditions are not met. It is perfectly possible to prefer what is better for oneself even if one is a bad judge of what is in one's interest and one is not seeking one's own advantage. Furthermore, as we saw in the whooping crane example, successfully carrying out a project that is animated by non-self-interested preferences can make one's life better. But when people are bad judges or are not seeking their own advantage, there's little reason to take their preferences to be evidence concerning what will benefit them.

These considerations fall far short of a general justification for relying on actual preferences to determine what will benefit people, even in a rough and ready policy context. All that one has here are some grounds for diffidence about overruling preferences and a sketchy identification of the considerations that determine whether it is more or less reasonable to take preferences as indicating welfare. The assumptions that people's preferences are determined by their well-founded judgments of what will benefit them and that economists should defer to people's judgments of what will benefit them are not always justified. Clearly they are more reasonable in some contexts than in others. For example, applications of cost-benefit analysis to questions concerning the protection of endangered species such as whooping cranes and Siberian tigers will require some other sort of justification (if they have any justification at all). The values placed on protecting these species by contingent valuation studies do not reflect well-founded judgments of individuals concerning the extent to which the continued existence of whooping cranes and Siberian tigers benefits them.[14] The survival of Siberian tigers and whooping cranes bear to some extent on the welfare of individuals because people take pleasure in seeing these magnificent creatures or in merely contemplating their survival from the edge of extinction. But for most people, these pleasures are modest, and there are other non-self-interested reasons why one would prefer that these creatures not become extinct. In the case of endangered species, preferences are a poor measure of benefit.[15]

In arguing that Ann's preferences concerning endangered species are not mainly self-interested, we mean that her preferences concerning endangered species do not derive mainly from her judgment concerning what is better for her. Rather than attempting to distinguish the preferences

[14] As David Sobel puts it (1998: 251), "Would you rather that 10,000 acres of rain forest be preserved or that your career get an important boost? Commonsensically, we think that a person could prefer the former without this signaling that rain forest preservation promotes that person's well-being more than career advancement." For an extended critique of the use of so-called "existence values" in cost-benefit analysis, see Sagoff (2004).

[15] Policies that secure the survival of whooping cranes may help bring to fruition Ann's project to save them, and in that way they may contribute to Ann's well-being. On the other hand, they may render her efforts irrelevant to the survival of the cranes and thereby make her worse off while at the same time securing her objective. The measure of benefit to Ann is not her preferences.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

of agents that concern themselves from the preferences that do not, economists should ask whether the agent's preferences are directed toward the agent's own well-being. If Ben does not believe that there is any advantage to being a chimera, Ben's preference that he be one is not self-interested, and if he is right, then satisfying the preference does not benefit him. On the other hand, his preference that his daughter be acquitted of armed robbery might well be largely self-interested, and the acquittal might benefit him.

The evidential view and the related defence of taking preferences to indicate welfare do not support the thesis that welfare *is* the satisfaction of preferences. Instead they deny that normative economists need any philosophical theory of welfare. Regardless of what philosophical theory of human well-being one accepts, the best *indicator* of well-being in certain circumstances is the extent to which preferences are satisfied.

Economists cannot however sensibly take people's preferences to be evidence concerning what is good for them, unless they have some notion of what is good for people. They would have little reason to believe that preference satisfaction is the best measure of welfare unless they had some idea of what welfare is. They must rely on that idea both to distinguish between self-interested and non-self-interested preferences and to decide when people are good judges of what serves their own interests. Both economists and non-economists might be sceptical about whether economists possess any such notion of welfare and whether they may reasonably be asked to make such judgments. It might appear that we've given economists a way of jumping out of an uncomfortably warm preference-satisfaction frying pan directly into the fierce philosophical fire that heats it.

The way out of the difficulty lies in platitudes concerning well-being rather than through embracing some other philosophical theory. Assuming that human beings are generally capable of judging what sorts of things are good for themselves and others, presumably economists are capable of doing so, too.[16] For example, economists can easily see that the continued existence of Siberian tigers and whooping cranes makes some contribution to individual welfare, because they know that enjoyment contributes to welfare and that people enjoy looking at Siberian tigers and whooping cranes and imagining them in the wild. On the other hand, it is hard to see what other contribution to individual welfare the continued existence of these species could make, because it is hard to see how their existence bears

[16] It is worth noticing that people's preferences will not indicate what benefits themselves, unless individuals have some notion of what is good for themselves. It is, we suppose, possible that people could judge what benefits themselves without having any idea what is good for others. But it is hard to take this possibility seriously, and we shall have nothing more to say about it.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

on other intrinsic goods. This argument assumes, as economists and people in general typically do, that pleasures contribute to well-being, but it does not assume that nothing else does. It offers no exhaustive list of intrinsic goods and depends on no philosophical theory that specifies what things are intrinsically good for people. Philosophers would certainly like to do much better, but economists and human beings in general do not have to wait for a well-articulated and justified philosophical theory of welfare before they can venture opinions concerning what makes people's lives better or worse. We conjecture that welfare economists in fact have many beliefs about what makes people better off, though their philosophical modesty makes them reticent about trumpeting those beliefs. Not all of those beliefs need be defensible – and it is certainly not our purpose to defend them. Indeed we suspect that economists are prone to exaggerate the role of commodities and of pleasures in a good life. Whether this speculation is correct or not, economists know or think they know enough about the things that make people's lives good or bad that they can make good sense of the evidential view of the relationship between preference satisfaction and welfare.

7. CONCLUSIONS AND PRACTICAL IMPLICATIONS

Welfare economics influences policy evaluation explicitly mainly via cost-benefit analysis, though appraisal of macroeconomic policy also rests implicitly on theoretical welfare economics. Cost-benefit analysis assesses policies in terms of their net benefits: the total amount the beneficiaries of a policy would be willing to pay to implement the policy minus the total amount that those who would prefer that the policy not be adopted would require in compensation for agreeing to the policy. If willingness to pay reflects preference satisfaction, then the policy with the greatest net benefit is the most efficient at satisfying preferences. It has the greatest *capacity* to satisfy individual preferences (Kaldor 1939; see also Hicks 1939). Economists resist going further and maintaining that the policy with the greatest net benefits in fact leads to the greatest preference satisfaction, because that would require interpersonal comparisons of preference satisfaction, about which most economists have serious qualms. Furthermore, the extent to which people's preferences will actually be satisfied as the result of building a bridge or implementing a safety standard also depends on the distribution of income. So net benefit captures just one component of policy evaluation. It is not meant to be the sole criterion. Welfare economists usually assume that net benefit is an important consideration in policy choice and the consideration to which economics has the most to contribute, but they do not deny that there are other relevant considerations.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

Cost-benefit analysis raises many questions besides those concerning the relations between preferences and welfare, which have been discussed in this essay. Willingness-to-pay depends on wealth as well as preferences, and the techniques employed to measure net benefits either from market behaviour or surveys raise a host of further difficulties. But let us set those problems aside.

If preferences indicate individual welfare, then satisfying preferences makes people better off, and an increased capacity to satisfy preferences, as an increased capacity to make people better off, is a consideration in favour of a policy. Since, as we have argued, welfare is not preference satisfaction, satisfying preferences need not make people better off. Satisfying preferences makes people better off if and only if people prefer what is in fact in their interest. They may do this by accident or because succeeding in a worthwhile but non-self-interested endeavour may make someone's life better, but what people prefer will be more likely to match what benefits them in those circumstances where they know what is in their interests and their preferences are self-interested.

Net benefit can be a relevant consideration in policy evaluation when (a) economists are able to measure it, (b) net benefit measures capacity to satisfy preferences, (c) preferences concerning the policy and its consequences are mainly self-interested and (d) there is no way to improve upon people's expectations concerning the policy and its consequences for themselves. Retrospectively, what matters is whether people's beliefs are true. Prospectively, what matters are how well supported people's expectations are. Conditions (a) and (b) are well-known and much discussed in the literature on cost-benefit analysis (Sugden and Williams 1978; Mishan 1971, 1981; Layard and Glaister 1994; Adler and Posner 2006). This paper has argued for conditions (c) and (d).

In addition, as economist's knowledge of cognitive biases and the nitty-gritty of motivation improves, it may be possible to start with willingness-to-pay information, whether inferred from market behaviour or elicited via contingent valuation studies, and to correct for the factors that make these raw data a reasonable guide to preferences but a mistaken guide to individual welfare. In this way, the domain within which some variant of cost-benefit analysis can usefully inform economic policy might be considerably extended.

Decisions about whether to make such corrections to people's preferences and whether to employ cost-benefit analysis at all should rest on the determination of whether people are both self-interested and good judges of their own interests with respect to the effects of the policy alternatives under consideration. These determinations are fallible, and they are bound to be controversial, but what else is new? Unrestrained application of cost-benefit analysis or blanket condemnation are not less controversial. By determining whether these assumptions of

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

self-interest and good judgment are reasonable, economists can judge when welfare economics can provide sensible policy guidance. In this way, the evidential connection between preferences and well-being provides guidance concerning when those who employ cost-benefit analysis should or should not be willing to take uncorrected willingness-to-pay information gleaned from market behaviour or surveys as the basis for policy evaluation.

Economists may not welcome the tasks they must undertake if preference satisfaction is related to welfare in only the evidential way discussed above. Consider these three tasks separately. First, economists must assess whether people's preferences are self-interested, rather than taking them simply as givens. But practitioners of cost-benefit analysis do not always take preferences simply as givens. As Adler and Posner point out, to some extent, cost-benefit analysts already avoid measuring non-self-interested preferences:

> The USDA does not ask animal rights activists whether they care about the existence of slaughterhouses. The U.S. Postal Service does not ask individuals whether they care about the existence of pornography in the mail. Funding agencies do not ask people how much they would be willing to pay to prevent those agencies from funding morally controversial research projects involving stem cells. (2006: 126–7)

In each of these cases Adler and Posner mention, there is an element of self-interest. For example, animal-rights activists would be happier if there were no slaughterhouses. But their preferences are not mainly self-interested and taking them into account would be a way of making animals rather than people better off.[17] In pointing out (as Adler and Posner also do) that the distinction between self-interested and non-self-interested preferences is implicit in some current applications of cost-benefit analysis, we do not mean to suggest that the restriction to self-interested preferences is already accepted by welfare economists. The relationship between welfare and preferences defended in this paper implies a significant change in welfare economics. The point is rather that economists already make judgments concerning the nature of people's preferences and thus cannot take the need to do so as grounds to refuse to change their practice in the ways we are recommending.

In addition to restricting the preferences from which net benefit is calculated to those which are self-interested, the evidential connection between preferences and welfare implies that net benefit calculations should depend only on preferences that reflect true beliefs and the best

[17] There are, of course, many other benefits that would accrue to people from ceasing to slaughter animals for food and hides. The point here is just that the preferences of animal-rights activists for eliminating animal slaughter do not indicate the benefit to them.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

supported expectations about the bearing of the consequences of policies on the individual's welfare. The issues here have already been canvassed in the controversy over so-called *ex ante* and *ex post* approaches to cost-benefit analysis (Harris and Olewiler 1979; Broome 1991: ch. 7; Hammond 1983; Hausman and McPherson 1994). On the assumption that welfare is preference satisfaction, the *ex ante* approach takes welfare to be the expected level of preference satisfaction, given the subjective probabilities of the individuals in question. *Ex post* welfare is how well preferences are in fact satisfied by the outcome. The *ex ante* approach evaluates outcomes in terms of *ex ante* preference satisfaction, while the *ex post* approach (which must of course also assess alternatives prospectively, if it is to guide policy choice) evaluates outcomes in terms of "objective" expectations of preference satisfaction – that is, expectations that rely on objective probabilities.

Although the *ex ante* approach, which asks no questions about whether individuals have mistaken beliefs or biased expectations, has many defenders, both in principle and in practice there is little justification for relying on preferences based on misapprehensions. Adler and Posner provides some nice examples of the way this realization influences current practice:

> Before asking them about air quality over the Grand Canyon, the EPA showed survey respondents photographs of the site with different levels of pollution. The EPA's goal was presumably to provide information on environmental aesthetics, about which respondent's intrinsic preferences were uninformed. For a regulation involving labeling of meat and poultry products, the USDA relied on CVs for health benefits people would enjoy if they altered their behavior in response to the labels, rather than people's CVs for nutrition disclosure. The agency appeared to take the intrinsic preference (for health) as a given and to circumvent the problem of imperfectly informed instrumental preference (for nutritional disclosure).[18]
>
> Those evaluating regulations can draw rough and ready distinctions between the contexts in which people are likely to be good judges of their own interests and the contexts in which they are likely to make mistakes, and as the last examples show, economists can take steps to provide people with better information and to correct mistaken beliefs or to avoid relying on them.

This leads to the third and most controversial task: correcting for mistakes, biases, and non-self-interested motives that make preferences a misleading

[18] Adler and Posner (2006: 127–8). The information concerning the EPA is from Deck (1997: 267), and the information concerning the USDA is from Department of Agriculture (1991). "CVs" are "compensating variations", that is, how much would have to be added or subtracted to people's wealth if the project were instituted so as to make individuals indifferent between the project and the status quo. Compensating variations are the most common interpretation of the more general idea of willingness to pay.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

guide to welfare. Contemporary psychology has had considerable success in identifying contexts in which people are likely to make mistakes, and policy analysts can use these findings both to help decide whether to take people's preferences as guides to their welfare and to consider methods that will permit them to correct for cognitive distortions. One advantage of not identifying preference satisfaction and welfare is that economists and regulators can make clear sense of people's preferences being mistaken – since what they are interested in are people's judgments about what serves their interests, not specifically preferences at all.

Accepting the view that the connection between preferences and well-being is purely evidential permits economists to avoid offering any philosophical theory of well-being.[19] The challenges facing those who hope to employ cost-benefit analyses to guide regulation would be clearer, too, since they would know that they need to address two central questions: (1) Is this a context in which willingness-to-pay information provides good evidence concerning what individuals believe is beneficial to themselves? and (2) Is this a context in which individuals are good judges of what is beneficial to themselves?

REFERENCES

Adler, M. and E. Posner. 2006. *New foundations of cost-benefit analysis*. Cambridge, MA: Harvard University Press.

Arneson, R. 1990. Liberalism, distributive subjectivism, and equal opportunity for welfare. *Philosophy and Public Affairs* 19: 158–94.

Arrow, K. 1973. Some ordinalist-utilitarian notes on Rawls' theory of justice. *Journal of Philosophy* 70: 246–63.

Bentham, J. 1789. *An introduction to the principles of morals and legislation*, ed. W. Harrison, 1967. Oxford: Basil Blackwell.

Binmore, K. 1994. *Playing fair: Game theory and the social contract*. Cambridge, MA: MIT Press.

Brandt, R. 1979. *A theory of the right and the good*. Oxford: Oxford University Press.

Broome, J. 1991. *Weighing goods*. Oxford: Basil Blackwell.

Cooter, R. and P. Rappoport. 1984. Were the ordinalists wrong about welfare economics? *Journal of Economic Literature* 22: 507–30.

Debreu, G. 1959. *Theory of value*. New York: Wiley.

Deck, L. 1997. Visibility at the Grand Canyon and the Navajo generating station. In *Economic analyses at EPA: Assessing regulatory impact*, ed. R. Morgenstern, 267–302. Washington, DC: Resources for the Future.

Department of Agriculture. 1991. Nutrition labeling of meat and poultry products. 56 *Federal Registry* 60, 023 (November 27).

[19] For those who believe that policy should depend in part on interpersonal comparisons of welfare, the evidential view would spare them the hopeless task of attempting to explain how it is possible to make interpersonal comparisons of preference satisfaction. There are serious difficulties in making interpersonal comparisons of mental states or of different objective lists of goods accruing to different individuals, but these pale in comparison to the special conceptual difficulties attached to comparing the "extent" to which Ann and Ben's preferences are satisfied (Hausman 1995).

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

Dworkin, G. 1971. Paternalism. In *Morality and the Law*, ed. R. Wasserstrom, 107–36. Belmont, CA: Wadsworth Publishing Company.
Dworkin, R. 1981. What is equality? Part 1: equality of welfare. *Philosophy and Public Affairs* 10: 185–246.
Elster, J. 1983. *Sour grapes: Studies in the subversion of rationality*. Cambridge: Cambridge University Press.
Elster, J. and A. Hylland, eds. 1986. *Foundations of social choice theory*. New York: Cambridge University Press.
Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68: 5–10.
Gauthier, D. 1986. *Morals by agreement*. Oxford: Oxford University Press.
Gibbard, A. 1986. Interpersonal comparisons: preference, good, and the intrinsic reward of a life. In *Foundations of social choice theory*, ed. J. Elster and A. Hylland, 165–94. New York: Cambridge University Press.
Goodin, R. 1985. Laundering preferences. In *Foundations of social choice theory*, ed. J. Elster and A. Hylland, 75–101. New York: Cambridge University Press.
Griffin, J. 1986. *Well-being: Its meaning, measurement and moral importance*. Oxford: Clarendon Press.
Griffin, J. 1996. *Value judgment: Improving our ethical beliefs*. Oxford: Oxford University Press.
Hammond, P. 1983. Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In *Social choice and welfare*, ed. P. Pattanaik and M. Salles. Amsterdam: North-Holland.
Harris, R. and N. Olewiler. 1979. The welfare economics of *ex post* optimality. *Economica* 46: 137–47.
Harsanyi, J. 1982. Morality and the theory of rational behavior. *Social Research* 44 (1977): 623–56. Republished in *Utilitarianism and Beyond*, ed. A. Sen and B. Williams, 1982, 39–62. Cambridge: Cambridge University Press.
Hausman, D. 1995. The impossibility of interpersonal utility comparisons. *Mind* 104: 473–90.
Hausman, D. 2000. Revealed preference, belief, and game theory. *Economics and Philosophy* 16: 99–115.
Hausman, D. 2005. Sympathy, commitment and preference. *Economics and Philosophy* 21: 33–50.
Hausman, D. and M. McPherson. 1994. Preference, belief, and welfare. *American Economic Review Papers and Proceedings* 84: 396–400.
Hausman, D. and M. McPherson. 2006. *Economic Analysis, Moral Philosophy, and Public Policy*. Cambridge: Cambridge University Press.
Heathwood, C. 2005. The problem of defective desires. *Australasian Journal of Philosophy* 83: 487–504.
Hicks, J. 1939. The foundations of welfare economics. *Economic Journal* 49: 696–712.
Kahneman, D. 2000. Experienced utility and objective happiness: a moment-based approach. In *Choices, values and frames*, ed. D. Kahneman and A. Tversky, 673–92. New York: Cambridge University Press and the Russell Sage Foundation.
Kahneman, D. 2006. New challenges to the rationality assumption. In *The construction of preference*, ed. S. Lichtenstein and P. Slovic, 487–503. Cambridge: Cambridge University Press.
Kahneman, D. and A. Krueger. 2006. Developments in the measurement of subjective well-being. *Journal of Economic Perspectives* 20: 3–24.
Kaldor, N. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* 49: 549–52.
Kraut, R. 2007. *What is good and why*. Cambridge, MA: Harvard University Press.
Layard, R. 2005. *Happiness: Lessons from a new science*. London: Penguin.
Layard, R. and S. Glaister, ed. 1994. *Cost-benefit analysis*, 2nd edn. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press

Mill, J. S. 1859. *On Liberty*. Republished, ed. C. V. Shields., 1985. New York: Macmillan.
Mill, J. S. 1863. *Utilitarianism*. Republished in *The Philosophy of John Stuart Mill*, ed. M. Cohen, 1961, 321–98. New York: Modern Library.
Mishan, E. 1971. *Cost benefit analysis: An introduction*. New York: Praeger.
Mishan, E. 1981. *An introduction to normative economics*. Oxford: Oxford University Press.
Nagel, T. 1986. *The view from nowhere*. Oxford: Oxford University Press.
Overvold, M. 1984. Morality, self-interest, and reasons for being moral. *Philosophy and Phenomenological Research* 44: 493–507.
Parfit, D. 1984. *Reasons and persons*. Oxford: Oxford University Press.
Railton, P. 1986. Facts and values. *Philosophical Topics* 24: 5–31.
Raz, J. 1984. *The morality of freedom*. Oxford: Oxford University Press.
Sagoff, M. 2004. *Price, principle, and the environment*. Cambridge: Cambridge University Press.
Samuelson, P. 1947. *Foundations of economic analysis.* Cambridge, MA: Harvard University Press.
Scanlon, T. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.
Schelling, T. 2006. *Strategies of commitment and other essays.* Cambridge, MA: Harvard University Press.
Sen, A. 1971. Choice functions and revealed preference. *Review of Economic Studies* 38: 307–17.
Sen, A. 1973. Behaviour and the concept of preference. *Economica* 40: 241–59. Republished and cited in *Choice, welfare and measurement*, 1982, 54–73. Oxford: Blackwells.
Sen, A. 1977. Rational fools. *Philosophy and Public Affairs* 6: 317–44.
Sen, A. 1982. *Choice, welfare and measurement*. Oxford: Blackwells.
Sen, A. 1987. The standard of living: Lecture I, concepts and critiques. In *The standard of living*, ed. G. Hawthorn., 1–19. Cambridge: Cambridge University Press.
Sen, A. 1990. Gender and cooperative conflicts. In *Persistent inequalities*, ed. I. Tinker, 123–49. London: Oxford University Press.
Sen, A. and B. Williams, ed. 1982. *Utilitarianism and beyond*. Cambridge: Cambridge University Press.
Sen, A., J. Muellbauer, R. Kanbur, K. Hart and B. Williams. 1987. *The standard of living*, ed. G. Hawthorn. Cambridge: Cambridge University Press.
Sidgwick, H. 1901. *The methods of ethics*, 6th edn. London: Macmillan.
Sobel, D. 1998. Well-being as the object of moral consideration. *Economics and Philosophy* 14: 249–81.
Sugden, R. and A. Williams. 1978. *The principles of practical cost-benefit analysis*. New York: Oxford University Press.
Sumner, L. W. 1996. *Welfare, happiness, and ethics*. Oxford: Clarendon Press.

https://doi.org/10.1017/S0266267108002253 Published online by Cambridge University Press