Economics and Philosophy, **31** (2015) 141–159 © Cambridge University Press
doi:10.1017/S0266267114000431
journals.cambridge.org/eap

HAUSMAN AND MCPHERSON ON WELFARE ECONOMICS AND PREFERENCE SATISFACTION THEORIES OF WELFARE: A CRITICAL NOTE

ALEXANDER F. SARCH*

Abstract: Hausman and McPherson defend welfare economics by claiming that even if welfare does not consist in preference satisfaction, preferences still provide good, if fallible, evidence of welfare. I argue that this strategy does not yet fully solve the problems for welfare economics stemming from the preference satisfaction theory of welfare. More work is needed to show that our self-interested preferences are sufficiently reliable, or in some other sense our best, evidence of well-being. Thus, my aim is to identify the challenges that remain and clarify what additional work is needed before Hausman and McPherson's defence of welfare economics succeeds.

Keywords: Welfare, Well-being, Welfare economics, Preference, Self-interest

A common criticism of welfare economics takes aim at the position's commitment to a preference satisfaction theory of welfare (or well-being).[1] Proponents of welfare economics[2] typically take it that one state of affairs is better than another just in case the former produces more welfare than the latter, where welfare, in turn, is taken to consist in the satisfaction of one's preferences. On such views, the more one's preferences are satisfied, the higher one's level of welfare. One benefit of such a view is that it enables economists to draw inferences about the welfare effects of various

* Department of Philosophy, University of Southern California, 3709 Trousdale Parkway, Los Angeles, CA 90089-0451, USA. Email: sarcha@usc.edu. URL: https://sites.google.com/site/alexsarch/

[1] As is common in this literature, I will be using the terms 'welfare', 'well-being', 'a person's good', and 'self-interest' interchangeably.

[2] See e.g. Kaplow and Shavell (2002: 15–28), cf. Hausman and McPherson (2009: 2–6).

policies from information about how well those policies would satisfy the preferences of the affected individuals.

There are, however, a number of well-known objections to the theory that welfare consists in the satisfaction of preferences.[3] In some cases, our preferences are based on faulty information. In other cases, even when not misinformed, we prefer things that intuitively would not enhance our welfare. For instance, when our preferences are manipulated or otherwise not formed autonomously, their satisfaction does not always seem to enhance well-being (Sumner 1996: ch. 6). Moreover, even when our preferences are autonomous, we might still prefer things that intuitively do not make a positive contribution to welfare – as in certain cases of self-sacrifice or masochism (Brink 1989: 227; Kraut 1994; Heathwood 2005, 2011; Bradley 2007).

A dizzying array of responses has been offered to these problems for preference satisfaction theories, and I will not attempt to canvass them all here. Rather, my aim is to investigate the merits of one novel solution recently proposed by Daniel Hausman and Michael McPherson on behalf of welfare economics (Hausman and McPherson 2009). Their strategy is to back off the claim that welfare *consists in* the satisfaction of preferences, and instead to merely assert that preferences provide good (if fallible) *evidence* of welfare.

Although I am sympathetic to Hausman and McPherson's strategy of taking preferences to be a useful source of information about well-being, I will argue that they have not yet fully solved the problems for welfare economics stemming from the preference satisfaction theory of welfare. One of the central notions in their proposal – i.e. self-interested preferences – allows for several different interpretations. However, I will argue that even on the best interpretation of this notion, more work is needed to successfully make the case that the relevant self-interested preferences are sufficiently reliable, or in some other sense our 'best', evidence of well-being. My aim in this note is not to argue that Hausman and McPherson's project fails, but only that it is incomplete in its current form. Specifically, I will identify a number of challenges that remain and clarify what additional work is needed before Hausman and McPherson's defence of welfare economics can be regarded as successful.

1. HAUSMAN AND MCPHERSON'S STRATEGY

The core of Hausman and McPherson's strategy is to deny that welfare consists in the satisfaction of one's preferences, and instead to posit an

[3] Some excellent discussions of these problems can be found, e.g. in Parfit (1984: 493–502), Griffin (1986: 21–38), Kagan (1998: 25–48), Heathwood (2005) and Hausman and McPherson (2009: 6–9).

evidential link between an agent's preferences and her welfare. Their claim, more specifically, is that under certain circumstances (soon to be specified), preferences are evidence of welfare in the following sense:

> *Evidential Link*: If one prefers state of affairs A to state of affairs B, then this is good though fallible evidence that one has higher welfare in A than in B.

What, then, *are* the conditions under which this evidential link is supposed to hold? As Hausman and McPherson put it,

> [o]ur 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. (Hausman and McPherson 2009: 16; emphasis added)

Thus there are two conditions that must be satisfied in order for the evidential link to hold:

(1) One's preferences are self-interested; and
(2) One is well-informed[4] (e.g. about the consequences of the relevant alternatives).[5,6]

Making this restricted claim about the connection that is supposed to hold between preferences and welfare offers two main benefits. First, Hausman and McPherson's strategy seems to allow proponents of welfare economics to sidestep the philosophical objections to the preference satisfaction theory of well-being. Hausman and McPherson's view is not committed to such a theory of well-being – indeed, they explicitly reject these theories.[7] Instead, they assume only that preferences can be taken

[4] Hausman and McPherson do not specify precisely how much information is needed for a preference to count as 'well informed'. Given their use of the phrase 'well informed', I assume it is something less than *full* information.

[5] This information requirement should not be taken to involve knowledge of the true theory of welfare, or else Hausman and McPherson's proposal would become unilluminating. After all, it begs the question to argue that well-informed preferences are good evidence of actual welfare when being well-informed is stipulated to include knowledge of the true theory of welfare. If being well-informed means knowing what is best for one, then it is trivially true that one's well-informed preferences are a reliable guide to one's welfare.

[6] Hausman and McPherson sometimes formulate this condition by saying that agents must be 'good judges' of their own welfare. (See e.g. Hausman and McPherson 2009: 18.) However, this formulation raises similar concerns to those discussed below. For instance, it is not clear what level of reliability in one's welfare judgements is required in order to be a sufficiently 'good judge'. (See generally Section 3 below.)

[7] More specifically, they reject both the actual preference satisfaction theory (according to which welfare is enhanced by the satisfaction of actual desires) and the idealized preference satisfaction theory (according to which welfare is enhanced by the satisfaction of the preferences one would have under certain idealized conditions). See Hausman and McPherson (2009: 14).

as a useful *guide* to well-being, provided the two conditions mentioned above are satisfied. As a result, even if there are some preferences whose satisfaction intuitively would not enhance one's welfare, economists could continue to regard preferences as a reliable indicator of welfare. Accordingly, economists could continue to draw normative inferences about the welfare effects of various policies from analyses of people's preferences as expressed in their market behaviour. When one policy produces more preference satisfaction than another, one could infer that the first policy is more productive of welfare than the other. Hausman and McPherson's view thus would allow economists to keep using the standard tools of economic analysis to draw evaluative conclusions about policy alternatives – at least when the conditions of self-interestedness and adequate information are satisfied.

Second, this strategy is supposed to vindicate welfare economics' reliance on preferences 'regardless of what theory of welfare one accepts' (Hausman and McPherson 2009: 2). If Hausman and McPherson are right that preferences are a reliable guide to welfare (provided the above conditions are met) regardless of the theory of welfare that proves correct, then it would show that 'welfare economics does not rely on any [particular] theory of welfare' (Hausman and McPherson 2009: 2). Accordingly, economists would not need to take a stand on the difficult philosophical question of what well-being actually consists in. It would allow them 'to defend their practice of making inferences concerning well-being from people's preferences without committing themselves to any theory of well-being at all' (Hausman and McPherson 2009: 16). Instead, the thorny evaluative questions would be largely left up to the individual, as each person's utility function ends up being a reflection of her own conception of her well-being.

Hausman and McPherson's proposal thus seems attractive. However, I will argue that it is not a complete solution as it currently stands. The trouble concerns an ambiguity in a core component of their proposal: the self-interestedness condition. Hausman and McPherson claim that preferences are a reliable guide to welfare only if they are 'self-interested' in some sense. This notion can be interpreted in a number of ways, but for each one, difficulties remain. I will begin by briefly considering two interpretations, which, while natural, can be disposed of quickly. Then I will consider at greater length a third interpretation that might better capture Hausman and McPherson's actual view. This interpretation, too, however, faces challenges.

2. TWO PRELIMINARY INTERPRETATIONS

To home in on the most promising interpretation of Hausman and McPherson's position, we should begin by noting the difficulties with some natural, but unsatisfactory interpretations of 'self-interestedness'. One interpretation that immediately suggests itself is to take the self-interestedness condition to require that one's preferences be *self-regarding*, as opposed to other-regarding. On this interpretation, a necessary condition of one's preferences being evidence of welfare is that they in some sense *concern oneself or one's life*, rather than the lives of others.

This proposal faces two main difficulties. Most importantly, the distinction between self-regarding and other-regarding preferences is notoriously difficult to draw. Suppose one decides to purchase toothpaste brand X rather than toothpaste brand Y for one's children. Is the preference for brand X over Y – or more generally, for promoting the welfare of one's children – expressed by this choice self-regarding or other-regarding? The object of the preference clearly involves others in important ways, and yet there is also an obvious sense in which it concerns one's own life. After all, one's own welfare clearly can be bound up with the flourishing of one's children. A great many of the preferences that are given expression in our market choices will similarly not fall neatly on either side of the divide between self-regarding and other-regarding preferences.

Even if some principled way to distinguish self-regarding from other-regarding preferences were found,[8] Hausman and McPherson would still reject this interpretation of 'self-interestedness', since they note a second sort of problem for it. They suggest that some preferences that are clearly self-regarding (i.e. which obviously concern one's own life) might nonetheless do nothing to enhance welfare.[9] Consider, for instance, the desire to be a chimera – i.e. one whose 'cells trace back to two or more separate fertilized eggs that fused into a single embryo' (Hausman and McPherson 2009: 13). This desire clearly concerns one's own life, and yet Hausman and McPherson think that its satisfaction *by itself* – i.e. without any enjoyment derived from *knowing* the desire is satisfied – intuitively would not enhance one's welfare (Hausman and McPherson 2009: 14). Granted, if this were an isolated case, it might not refute the claim that self-regarding preferences are fallible evidence of welfare. But Hausman and McPherson take the chimera case to demonstrate a deeper problem: namely, that obtaining the objects of one's preferences, even self-regarding ones, would not enhance welfare unless doing so is accompanied by some form of enjoyment or positive affect, broadly construed. Obtaining what one prefers would do nothing for one's welfare if, due to a failure to recognize or set store by that fact, it ultimately 'leaves one cold'. If

[8] For instance, Mark Overvold has defended the view that a self-regarding preference is one for states of affairs that require the existence of the agent. See Overvold (1984). Others have criticized this solution. See Hausman and McPherson (2009: 12 (footnote 8)).

[9] See generally Hausman and McPherson (2009: 13–14).

https://doi.org/10.1017/S0266267114000431 Published online by Cambridge University Press

Hausman and McPherson are right in this, then restricting the reliance on preferences to just those which are self-regarding still would not ensure that those preferences are a reliable guide to welfare.

Since this first interpretation of the self-interestedness condition faces problems and would be rejected by Hausman and McPherson themselves, I will set it aside here. A second natural interpretation takes the requirement that one's preferences be self-interested to mean that one's preferences must reflect what *really is good for one*. That is, the condition would require that whenever P prefers A to B, P *really* would have higher well-being in A than in B. Only if this condition is satisfied, the current thinking goes, would preferences provide a reliable guide to welfare.

This second interpretation, too, faces several serious problems. Most importantly, it does not bear out Hausman and McPherson's claim (2009: 1, 16) that economists can make do without a theory of well-being. After all, we can tell whether one's preferences reflect what really is good for one (i.e. whether the self-interestedness condition is satisfied) only if we have a theory that tells us what well-being consists in. Accordingly, Hausman and McPherson's strategy, thus interpreted, would require economists to take a stand on controversial questions about what in itself enhances well-being.

Hausman and McPherson indicate some awareness of this objection, which suggests that they may have envisioned the current interpretation of their view. The response they propose is this:

> The way out of this 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. (Hausman and McPherson 2009: 18)

Such an answer is unlikely to succeed, however. For it simply amounts to the claim that economists already possess intuitive views about well-being that are good enough to decide when the self-interestedness condition is satisfied – i.e. to decide when preferences reflect actual well-being. This, however, would not be consistent with Hausman and McPherson's attempt to render welfare economics 'independent of any philosophical theory of well-being' (Hausman and McPherson 2009: 1). Their strategy would not free economists from having to endorse particular views about well-being. Rather, the present response amounts to the claim that economists already *have* a sufficiently good intuitive understanding of the nature of well-being.

What's more, even if Hausman and McPherson are correct that economists' intuitive understanding of well-being is good enough to decide at least in *easy* cases whether preferences are self-interested in the current sense (consider, e.g. the preference for a world with an odd number of stars in the sky), it still seems insufficient to settle the more difficult cases that frequently arise. Would welfare in fact be enhanced by the satisfaction of, say, morally troubling preferences (e.g. the desire to engage in repulsive or morally reprehensible activities) (Brink 1989: 227; Feldman 2004: 40) or altruistic preferences (e.g. the desire to forego one's own flourishing for the sake of others) (Heathwood 2011)? Would welfare be enhanced by the satisfaction of preferences that one used to hold but has now abandoned (Bykvist 2003, 2007)? These are subtle normative questions that philosophers have spilt much ink debating. Even if economists possess an intuitive understanding of the concept of welfare, it is not likely to be sufficient to guide them in the hard cases that often come up. Thus it is too quick to assume that economists' intuitive views about welfare will allow them to reliably determine when the self-interestedness condition, on this interpretation, is or isn't satisfied. If economists cannot reliably determine when that condition is satisfied, then they will not be able to reliably decide when preferences serve as a guide to welfare.

3. A BETTER INTERPRETATION OF 'SELF-INTERESTEDNESS'

Accordingly, we need an interpretation of the self-interestedness condition that avoids the problems of the previous proposals. What is called for is an interpretation that both avoids the need for a substantive theory of welfare and does not rely on difficult-to-apply concepts like that of a self-regarding preference.

A more promising suggestion, then, is to take it that an agent's preferences are self-interested when they reflect her *own judgements* about what is in her interest. Indeed, Hausman and McPherson often talk this way,[10] so there is reason to think that this would be their preferred interpretation. More precisely, this interpretation amounts to the claim that preferences are good evidence of welfare only if the following version of the self-interestedness condition (stated in section 1) is met:

(1′) P's preferences among states of affairs reflect her *judgements* about what would be good for her. That is, she prefers A to B if and only if (and because) she judges that her welfare is higher in A than in B.

If this is what the requirement that preferences be 'self-interested' means, then it might seem plausible that economists would not need to adopt a

[10] '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 judgement concerning what is better for her*' (Hausman and McPherson 2009: 17) (emphasis added).

theory of well-being. For all it takes to decide whether condition (1′) is satisfied is the ability to decide whether P's preferences between states of affairs track her judgements about which state would be better for her. And that is something we do not need a theory of well-being to decide. Perhaps we could decide the matter merely by quizzing a representative sample of affected individuals. Thus on this interpretation of Hausman and McPherson's proposal, the question of what will promote the agent's well-being is left for the agent to decide for herself, and all the economist would have to do to decide whether (1′) is satisfied is figure out whether the preferences of the relevant agents bear the right relation to their judgements about their own good – i.e. whether the former track the latter.

Promising though this route might seem, I will argue that it, too, faces problems. Before discussing the three main areas of concern, however, some clarification is in order. Specifically, it should be noted that the current interpretation leaves little work for *preferences* to do. Recall that according to welfare economics, an individual can be taken to have higher welfare in A than B if and only if she *prefers* A to B. The resulting information about individual welfare is then aggregated to construct an overall ranking of the alternatives in terms of how much welfare they contain. Hausman and McPherson's proposal is that alternatives may be ranked in this way only if the underlying preferences are both well-informed and self-interested. However, if the current interpretation of 'self-interested' is adopted, then (setting aside the informational requirement for the moment) individual welfare will be determined as follows:

> P has higher welfare in alternative A than alternative B if and only if [P *prefers* A to B] AND [it's also the case that P prefers A to B if and only if P *judges* that A is better for her than B].

This claim has the form P iff (Q & (Q iff R)), which simply reduces to P iff R. Thus for each individual, the relevant alternatives would really be ranked according to this simpler principle:

> P has higher welfare in alternative A than in alternative B if and only if P *judges* that A is better for her than B.

Thus on the current interpretation of the self-interestedness condition, (1′), there is no separate work for preferences themselves to do in ranking alternatives according to how much welfare an agent possesses in them. Considerations of theoretical simplicity, accordingly, would suggest that agents' preferences should drop out of the theory.[11]

[11] Of course, as an anonymous reviewer pointed out, preferences would still have work to do in other areas – for example, when it comes to explaining actions (as suggested by the Humean theory that actions are motivated by belief/desire pairs). Thus I'm not suggesting

Thus if we interpret the self-interestedness condition along the lines of (1′), it is misleading to construe welfare economics as being particularly concerned with *preference satisfaction*. The standard construal of welfare economics tells us to look at an agent's preferences to determine how well off she is in each of the relevant alternatives. However, if the present interpretation of Hausman and McPherson's proposal is adopted, then we should instead be considering agents' *judgements* about what is in their interest. Thus, instead of constituting a defence of economists' standard practices, Hausman and McPherson's proposal would imply that economists' concern with preference satisfaction is misguided. This, of course, would not warrant the rejection of welfare economics, but would entail some revisions to the standard understanding of it.

Nonetheless, the present interpretation of Hausman and McPherson's proposal also faces more serious challenges. A discussion of these reveals what additional work is needed before Hausman and McPherson's defence of welfare economics can be regarded as successful.

3.1 An Incomplete Solution

The first problem arises from the fact that the present interpretation of Hausman and McPherson's proposal does not fully solve the sort of problems for preference satisfaction theories that Hausman and McPherson were seeking a solution to from the outset. On the current interpretation, one's judgements about what one's welfare would be under various alternatives are assumed to provide a reliable guide to how well off one actually is in each. However, I argue that the agent's welfare judgements are likely to fail to track her actual welfare in very many of the same cases as her *preferences* fail to track welfare. Still, as we will see, we are not yet in a position to know whether this problem is so widespread as to be decisive, since it is not clear how reliably preferences must track welfare in order for Hausman and McPherson's proposal to succeed.

To see the initial concern that welfare judgements might fail to track actual welfare in many of the same cases as preferences do, consider some of the core cases that have led theorists to reject theories on which welfare consists in the satisfaction of preferences. Some objections concern 'idiosyncratic' or 'pointless' preferences – say, a preference for devoting one's life to hopeless pursuits like counting blades of grass (Rawls 1971: 432; Parfit 1984: 500; Hausman and McPherson 2009:

that the concept of preferences can be eliminated altogether. The present point is merely that on the current interpretation of 'self interested preferences', it is not necessary to appeal to preferences in order to determine which policy alternatives are more productive of welfare. Rather, to rank alternatives in terms of welfare, it would be enough to rely directly on individuals' judgements about their welfare, without considering their preferences as well.

8). Other objections concern 'obnoxious', 'sadistic' or otherwise 'anti-social' preferences – say, the preference for inflicting gratuitous harm on others (Hausman and McPherson 2009: 8). Still other objections concern preferences that are not fully autonomous – say, preferences that one has only because of brainwashing, coercion or more pervasive but subtle forms of manipulation (Sumner 1996: ch. 6; Hausman and McPherson 2009: 8). The intuition that the satisfaction of such preferences would not enhance welfare (or at least would do so significantly less than the preference satisfaction theory entails) is a large part of what motivates Hausman and McPherson's proposal in the first place. The trouble, however, is that someone who has a preference for such things seems likely to also be disposed to *judge* that they promote her welfare. If one has preferences that are pointless, sadistic or non-autonomous, one is likely to judge that getting those preferences satisfied is welfare-enhancing as well. Accordingly, many of the cases in which one's preferences intuitively do not track welfare are likely to also be cases where one's welfare judgements do not track welfare.

Of course, this does not hold for all the intuitively 'defective' preferences that seem to pose problems for preference satisfaction theories. If one had a masochistic preference (i.e. a preference to harm oneself), for instance, it seems unlikely that one would judge that it is in one's interest to have this preference satisfied. Similarly, if one had an altruistic preference for sacrificing one's own good in order to promote the good of a loved one, then one also might not judge the satisfaction of that preference to be in one's interest (though this case is less clear-cut than that of the masochistic preference). Accordingly, to be fair, one might think that the move to judgements of self-interest does help with *some* of the cases that pose problems for the preference satisfaction theory of welfare.

Nonetheless, the move from preferences to welfare judgements is of little or no help for many of the core cases of 'defective' preferences that have lead theorists to abandon the preference satisfaction theory of welfare. Since one's welfare judgements are likely to be mistaken or misguided in many of the same cases in which preferences intuitively fail to track actual welfare (e.g. preferences that are idiosyncratic, anti-social or non-autonomous), Hausman and McPherson's proposal, on the current interpretation, seems not to completely avoid the problems for the preference satisfaction theory that they were seeking a solution to at the outset. Although this may not entirely undermine the claim that welfare judgements and the preferences reflecting them are fallible evidence of welfare (since some unspecified amount of error is compatible with this evidentiary claim), it does raise questions about how much of an improvement Hausman and McPherson's proposal is over the preference satisfaction theory they sought to replace.

In response to this worry, Hausman and McPherson might appeal to the second condition in their proposal. Their view, recall, is that preferences are a reliable guide to welfare as long as (i) the preferences reflect one's judgements about what promotes one's welfare and (ii) one is well-informed. The informational requirement might be thought to significantly reduce the frequency of cases in which the agent's judgements about her own welfare, and thus the relevant sub-set of self-interested preferences, miss the mark. Perhaps there won't be many *well-informed* people who are incorrect in their judgements about their own welfare.

Nonetheless, there are reasons to doubt that the information requirement will fully solve the problem either – i.e. ensure that welfare judgements will reliably track actual welfare. First of all, we can be reasonably sure that, even when granted full information, people will still have *conflicting* conceptions of what welfare consists in. All of these conflicting conceptions of what constitutes a life high in individual welfare cannot be correct at the same time. (That just is what it is for them to conflict.) Thus we can be quite sure *ex ante* that there will be a range of cases in which even our well-informed judgements about our own welfare will fail to reflect our actual welfare.[12]

A further argument supports the same conclusion. Start by noting that only one theory of welfare can be true. Whatever theory this turns out to be, it is not likely that everyone will believe it; only some will. Accordingly, there is yet more reason to think that even well-informed judgements about what promotes our welfare will frequently be misguided. For example, suppose that some Aristotelian theory of welfare turns out to be true.[13] Even so, it would be unlikely that everyone will come to believe this theory of welfare. There would still be many Hedonists, many who endorse some objective list theory of welfare, many who accept the conception of the good life that is promulgated by some religion or other, and so on. But if we are supposing that an Aristotelian theory of welfare is true, then a wide range of welfare judgements of people who accept *other* theories of welfare will not be accurate – *even supposing that these individuals are otherwise well-informed*. Consider a Hedonist who is as well informed as you might like – i.e. she is fully informed about the consequences of the alternatives on offer and, indeed, is afforded all relevant factual knowledge.[14] Even so, her

[12] This objection of course would not have traction against those who reject the existence of objective evaluative facts. But this kind of radical anti-realism does not appear standard among proponents of welfare economics. Nor is it taken on board by Hausman and McPherson.

[13] For some helpful discussion of this kind of view, see e.g. Hurka (1993) and Haybron (2007).

[14] The person in question here obviously cannot be given all relevant *normative* knowledge – e.g. told what the true theory of welfare is – without rendering Hausman and McPherson's

judgements about what will enhance her welfare will not be particularly reliable evidence of her *actual* welfare, since, by hypothesis, an Aristotelian theory of welfare is supposed to be true. Now, this is not a special fact about Aristotelianism. Regardless of which theory of welfare actually proves correct, we can be sure that there will be many cases of just this sort – i.e. where people who subscribe to some *other* theory of welfare make systematically misguided judgements about what enhances their own welfare, despite being well-informed. Thus no matter what theory of welfare turns out to be true, there is reason to think that for many people, even their well-informed judgements about welfare will fail to track their actual welfare. (Although one might respond that different theories of welfare might largely agree in their first-order implications about what things promote welfare, and differ only as to the explanation given of why these happen to be the welfare goods, this view of the welfare debate would have the implausible consequence that there is little or nothing substantive at issue between different theories of welfare. Philosophers of well-being do not seem to be engaged merely in the academic exercise of identifying the most convincing story to tell about a putative consensus concerning the goods that promote welfare; they also seem concerned to make progress on controversial first-order questions about welfare.[15])

At this point, one might wonder how great a threat to Hausman and McPherson's view these considerations really pose. After all, one might grant that there will be some cases in which our well-informed judgements about our own welfare miss the mark, while still insisting that there is at least a *reliable* connection between such judgements and our actual welfare. To succeed in their project, Hausman and McPherson do not need our well-informed welfare judgements to *always* reflect actual welfare, so long as they *usually* do so. Accordingly, perhaps the rate of

view unhelpful. After all, if our preferences or judgements are supposed to offer reliable evidence of welfare only when we *know* the true theory of welfare, then economists could hardly be expected to be able to discern *when*, exactly, our preferences or judgements are good evidence of well-being unless the economists themselves knew the correct theory of well-being.

15 More specifically, while it is not implausible that the competing theories of welfare might agree in their first-order implications about a large class of 'standard' cases, this does not change the fact that they likely also disagree on a large number of important cases. To take just one example, hedonism and the preference satisfaction theory would offer significantly different advice to one who is trying to decide whether to either (i) put in the difficult work needed to achieve a set of long-standing desires even though it would bring little enjoyment, or else (ii) to opt for a pleasant existence in which those desires remain unfulfilled. This is just one example, but I do not think it is uncommon. It is of course difficult to estimate the relative sizes of (i) the class of cases in which various theories of welfare agree in their first-order implications and (ii) the class of cases in which they disagree. But I think it is clear that the competing theories on offer really do have different substantive implications in a range of cases.

error that is suggested by the above considerations is small enough to still be acceptable. Welfare economics, after all, is meant to be a practical tool for evaluating actual policy choices, and so the approach could still be employed even if there are some cases in which well-informed and self-interested preferences fail to track actual well-being.

However, we are not yet in a position to tell whether this reply succeeds. Whether it does or not depends on *how reliably* our well-informed welfare judgements must track actual welfare in order to vindicate welfare economics' reliance on preferences under the conditions Hausman and McPherson identify. Do our welfare judgements – and more generally our self-interested preferences – have to track the facts about welfare with a *near perfect* degree of reliability? Must they merely be *more likely than not* to accurately reflect actual welfare? What ratio of accurate to erroneous welfare judgements is required in a given domain in order for economists to be justified in relying on the self-interested preferences in that domain as a guide to welfare? Hausman and McPherson do not tell us.

These questions require principled answers.[16] Without a clear way to specify the degree of reliability with which welfare judgements must track actual welfare, we are not yet in a position to determine whether the rate of error suggested by the considerations above is too high to be acceptable. What is more, without specifying the required level of reliability, we would not (even in principle) be able to determine whether welfare judgements in a given domain of the market are sufficiently reliable for welfare economics to be justifiably applied to that domain. Nor would we be able to determine whether such domains are common or rare. Accordingly, we would not be able to ascertain whether the *widespread* reliance on preferences by welfare economists is justified. The upshot is that more ground-level work is needed in order to get clear on the extent to which Hausman and McPherson's proposal would vindicate welfare economics.

[16] The issue is further complicated by the fact that whether welfare judgements (or preferences) reflect actual welfare is not a binary question. Rather, they can depart from the welfare facts to greater or lesser degrees. Suppose A would enhance your welfare three times as much as B would. If you judge that A would enhance your welfare only twice as much as B (or have a preference for A that is twice as strong as your preference for B), then this judgement (preference) would reflect the facts about welfare at least *somewhat*, since it gets the ordinal ranking of A and B right; but the match would be less than perfect. Thus any answer to the question of how reliable our welfare judgements (or self-interested preferences) have to be must account for the fact that it is a matter of *degree* how well such judgements (or preferences) reflect actual welfare.

3.2 Are Well-Informed Self-Interested Preferences Our 'Best' Evidence of Welfare?

Now, Hausman and McPherson may be able to avoid the difficult task of specifying the precise level of reliability required for economists to be justified in relying on preferences (if well-informed and self-interested). Even without being able to affirmatively establish that well-informed welfare judgements – or the preferences that track them – rise to the requisite level of reliability (whatever it is), economists might still plausibly claim to be justified in relying on preferences reflecting such judgements *if these are our best available evidence of welfare*. In general, it seems justified to rely on whatever proves to be our best available evidence of welfare. If something besides the relevant sub-set of preferences were *better* evidence, though, the reliance on such preferences would not be obviously justified.

Thus, even without being able to identify in the abstract what level of reliability is required, Hausman and McPherson's defence of welfare economics might still be regarded as successful if they were able to show that preferences (when well informed and self-interested in the current sense) are our *best* available evidence of welfare. In fact, Hausman and McPherson themselves claim that 'the *best* indicator of well-being in certain circumstances is the extent to which preferences are satisfied' (Hausman and McPherson 2009: 18).

Nonetheless, even under the limited circumstances of adequate information and self-interestedness that Hausman and McPherson insist on, it is doubtful that our preferences are going to the *best* available evidence of actual welfare – at least if 'best' here is understood to mean 'most reliable.' After all, it is in principle possible to gather much more information about each individual affected by a given policy proposal, besides the person's preferences, that would provide additional good evidence of her well-being. For instance, we might try to discern what would be good for the people in question through detailed and empathetic conversations with a representative sample of affected individuals,[17] or perhaps through psychological evaluations aimed at determining the factors that would provide stable and lasting happiness. Of course, this would be a very time-consuming and expensive way to go about evaluating alternative policy proposals. Still, the important point for present purposes is just that there is a range of other sources of information about individual welfare which, together with preference information, would likely provide even better – that is to say, *more reliable* – evidence of welfare than preferences alone ever could. Thus it

[17] This is perhaps akin to the sort of 'notice and comment' procedures that government agencies often use to aid decision-making.

seems incorrect to take it that preferences alone, even if self-interested in the relevant sense and well-informed, would provide the most reliable evidence of welfare.

Nonetheless, there are other senses of ‘best evidence’ that might be employed when claiming that preferences are our best available evidence of welfare. It may not be implausible to claim that preferences that reflect our well-informed welfare judgements are a source of evidence that provides *the most desirable balance between reliability and practicability*. Perhaps this, then, is the sense in which Hausman and McPherson claim that well-informed and self-interested preferences are ‘the *best* indicator of well-being in certain circumstances’ (Hausman and McPherson 2009: 18). Alternatively, one might think that there are certain normative constraints that limit the kinds of information about individuals that can be deemed admissible as input to a policy analyst’s welfare assessments. Perhaps more private kinds of information (e.g. interview information, fMRI scans, purchasing habits, patterns of web-browsing) would not be deemed admissible absent the consent of the individuals in question.[18]

However, a great deal more argument is required on Hausman and McPherson’s part to establish the conclusion that preferences are the best indicator of well-being in some sense that goes beyond mere reliability. At a minimum, they would have to canvass the main ways to acquire evidence about well-being on offer, weigh the reliability and practicality of each, and then show that the admissible information about self-interested and well-informed preferences strikes the most desirable balance between these two values. This in turn requires a principled explanation of how reliability and practicality are to be traded off against one another – no easy feat given the ways in which these values might (at least *prima facie*) seem incommensurable. Furthermore, any applicable admissibility constraints would have to be considered, and it is far from clear that all sources of information about preferences will automatically pass muster, while non-preference information will be barred.[19]

While I suspect that some such strategy may well be the avenue by which welfare economics’ reliance on preferences will ultimately have to be justified, the important point for present purposes is only that Hausman and McPherson have not yet made the case for this conclusion.

[18] Thanks to an anonymous reviewer for raising this point.

[19] Although one might think that such an admissibility constraint would allow in preference information, but rule out more qualitative forms of welfare information, I suspect that the matter will prove more complicated. In particular, preference information can be gathered in many ways, some of which will be more intrusive than others. Aggressive data collection with respect to purchasing habits or online activity might be inadmissible, while interview information, questionnaires or psychological treatment history might turn out to be admissible if anonymity is preserved or consent is obtained. Thus, a detailed investigation of any applicable admissibility constraints cannot be avoided.

More work is needed to establish that preferences, if well-informed and self-interested (on the present interpretation) are in some sense or other our 'best' available evidence of welfare.

3.3 Theory-dependence

A third problem concerns Hausman and McPherson's claim that economists are justified in relying on preferences that are well-informed and self-interested (in the current sense) regardless of what theory of welfare turns out to be true. The trouble is that our well-informed judgements about our own welfare likely would not reliably track the facts about what at least some theories say actually enhances welfare.

Anecdotal evidence suggests that many people (perhaps most) are sympathetic to some form of welfare subjectivism.[20] According, to such theories, welfare depends mainly on one's attitudes or experiential states. Suppose for the sake of argument that most people today form judgements about what is good for them in some subjectivist manner – say, on the basis of what they think would make them *feel* best. In that case, judgements about our own welfare would not conform well to what another important class of theories – i.e. the *objective* theories – says would enhance welfare. Accordingly, whether Hausman and McPherson are correct that our welfare judgements and the self-interested preferences that track them provide reliable evidence of welfare depends on whether subjectivism or objectivism turns out to be true. In the scenario just sketched, where most people are subjectivists of some sort, Hausman and McPhersons' claim would plausibly be true if subjectivism about welfare proves correct, but likely false if objectivism proves correct.

The upshot is that Hausman and McPherson's defence of welfare economics is partially dependent on what theory of welfare is true. Whether our well-informed welfare judgements – and thus the preferences that are self-interested in the current sense – provide reliable evidence of actual welfare depends on (i) the contingent facts about how we form our welfare judgements (including facts about what theories of welfare enjoy popular support) and (ii) what theory of welfare turns out to have been true all along. More specifically, it depends on whether (i) matches up with (ii). Accordingly, the likelihood that Hausman and McPherson's defence of welfare economics is successful is not entirely independent of what theory of welfare is true.

[20] Cf. Haybron (2007: 1), discussing the 'present age of subjectivism'.

https://doi.org/10.1017/S0266267114000431 Published online by Cambridge University Press

4. AN OBJECTION TO HAUSMAN AND MCPHERSON'S PROPOSAL, HOWEVER INTERPRETED

We have seen that the most natural interpretations of the notion of self-interested preferences continue to face serious questions. Of course, other interpretations are possible,[21] and it might well prove fruitful to pursue some of them further. However, even if some interpretation can be found that avoids the above problems, a final concern remains. No matter how we understand the notion of self-interested preferences, additional argumentation is needed to vindicate the widespread use of the economist's technique of drawing inferences about agents' welfare from the choices they make in the marketplace.

Hausman and McPherson's proposal would justify the technique of drawing inferences about welfare from market choices only when we are dealing with a domain in which market choices are guided mainly by self-interested preferences (rather, say, than some mix of self-interested and non-self-interested preferences). As they put the point, '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' (Hausman and McPherson 2009: 17). People often decide to purchase certain items rather than others not because they see it as in any way connected to their own welfare, but rather because they believe that it will benefit their loved ones, promote a favoured ideological cause, or what have you. When this is the case, economists would not be in a position to use analyses of market behaviour to read off information about welfare.[22]

The trouble is that it is not immediately obvious which areas of the market are such that consumer choices reflect self-interested preferences. A great deal of subtle psychological information about the considerations

[21] Another interpretation (which likely departs from what Hausman and McPherson had in mind) would be to say that a preference is 'self-interested' when motivated or appropriately caused by a concern with oneself – more precisely, when it is produced by a sense of *self-love* (cf. Darwall 2002).

One problem with this proposal, however, is that there seem to be very few preferences that are motivated *purely* by self-love. In reality, most preferences – even those that seem to chiefly concern ourselves – are going to be motivated by an array of concerns, aims, habits, wishes, emotions and impulses. But it only seems plausible to take preferences stemming from self-love to provide good evidence of welfare when the preferences in question are *solely* motivated by self-love, or at least nearly so. If one of my preferences – e.g. to win the contest – stems not only from my sense of self-love, but also, say, from my sense of pride (which notoriously can come apart from welfare), then it is by no means clear that the satisfaction of this preference would promote my welfare. Accordingly, this interpretation of 'self-interested' is problematic, too, since our preferences actually stem from a variety of motives that can come apart from and conflict with welfare.

[22] Hausman and McPherson acknowledge several contexts (e.g. wildlife conservation) in which peoples' preferences do not seem self-interested (Hausman and McPherson 2009: 17).

that drive choices in this or that domain will be required. Accordingly, it cannot simply be assumed without argument that there exist a large number of contexts in which consumers' choices are driven chiefly by self-interested preferences. However, in order to vindicate a *widespread* reliance on market behaviour as evidence of welfare – not merely in a few isolated contexts – some argument is needed for thinking that there will be a significant number of areas in which self-interested preferences are indeed driving consumer behaviour.

Accordingly, to provide a general defence of the technique of reading off welfare information from market behaviour, Hausman and McPherson must offer not only a principled way to identify domains in which market behaviour is driven primarily by self-interested preferences, but also some reason to think that such domains will not be few and far between. Thus, as before, more work is still required in order to successfully defend the widespread use of welfare economics.

5. CONCLUSION

I have argued that more work is needed before Hausman and McPherson can be taken to have fully overcome the difficulties for welfare economics that stem from the preference satisfaction theory of welfare. Their suggestion that economists may rely on preferences as evidence of welfare only when these preferences are self-interested and well-informed does not yet provide a full vindication of the normative credentials of welfare economics.

This is not to say that Hausman and McPherson would not be right to claim that there likely are some contexts in which preferences can indeed provide useful evidence of welfare. However, we do not yet have a clear way to determine when we have encountered such a context. Moreover, more work is needed to make the case that preferences generally are the most reliable, or in some other non-reliabilist sense, our *best* evidence of welfare. At the very least, placing welfare economics on a secure normative footing will require more than the claim that preferences would provide *some* useful evidence of welfare if self-interested and well-informed. Thus, my conclusion is that while Hausman and McPherson have begun the important task of saving welfare economics from the problems that are due to preference satisfaction theories of welfare, that task is not yet complete. I hope to have made some small contribution to the project by indicating what steps remain.

ACKNOWLEDGEMENTS

The author would like to thank Stephen Campbell, Scott Hershovitz, Stephen Nyak-Young and an anonymous reviewer for this journal for helpful comments on previous drafts of this paper.

REFERENCES

Bradley, B. 2007. A paradox for some theories of welfare. *Philosophical Studies* 133: 45–53.
Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
Bykvist, K. 2003, The moral relevance of past preferences. In *Time and Ethics: Essays at the Intersection*, ed. H. Dyke, 115–136. Dordrecht: Kluwer Academic Publishers.
Bykvist, K. 2007. Comments on Dennis McKerlie's 'Rational Choice, Changes in Values Over Time, and Well-Being'. *Utilitas: A Journal of Utilitarian Studies* 19: 73–77.
Darwall, S. 2002. *Welfare and Rational Care*. Princeton, NJ: Princeton University Press.
Feldman, F. 2004. *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Clarendon Press.
Griffin, J. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.
Hausman, D. and M. McPherson. 2009. Preference satisfaction and welfare economics. *Economics and Philosophy* 25: 1–25.
Haybron, D. 2007. Well-being and virtue. *Journal of Ethics and Social Philosophy* 2: 1–27.
Heathwood, C. 2005. The problem of defective desires. *Australasian Journal of Philosophy* 83: 487–504.
Heathwood, C. 2011. Preferentism and self-sacrifice. *Pacific Philosophical Quarterly* 92: 18–38.
Hurka, T. 1993. *Perfectionism*. New York, NY: Oxford University Press.
Kagan, S. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
Kaplow, L. and S. Shavell. 2002. *Fariness Versus Welfare*. Cambridge, MA: Harvard University Press.
Kraut, R. H. 1994. Desire and the Human Good. *Proceedings and Addresses of the American Philosophical Association* 68: 39–54.
Overvold, M. 1984. Morality, self-interest, and reasons for being moral. *Philosophy and Phenomenological Research* 44: 493–507.
Parfit, D. 1984. What makes a person's life go best? In *Reasons and Persons*, 493–502. New York, NY: Oxford University Press.
Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Sumner, L.W. 1996. *Welfare, Happiness, and Ethics*. New York, NY: Clarendon Press.

BIOGRAPHICAL INFORMATION

Alexander Sarch is a Postdoctoral Fellow in Law and Philosophy at University of Southern California. He received his Ph.D. in philosophy from University of Massachusetts, Amherst and his J.D. from University of Michigan. His current scholarship lies at the intersection of law and moral philosophy, focusing particularly on the role of culpability in the criminal law and the normative foundations of law and policy analysis in general.