On Sense and Preference

James Fanciullo

Forthcoming in the *Journal of Moral Philosophy*

Abstract: Determining the precise nature of the connection between preference, choice, and welfare has arguably been the central project in the field of welfare economics, which aims to offer a proper guide for economists in making policy decisions that affect people's welfare. The two leading approaches here historically—the revealed preference and latent preference approaches—seem equally incapable of so guiding economists. I argue that the deadlock here owes to welfare economists' failure to recognize a crucial distinction between two senses of "preference." I analyze and defend these two senses of "preference," and argue that each shares a close connection with just one of choice and welfare. This analysis reveals how economists should conceive of both the connections between "preference," choice, and welfare, and the proper roles of these concepts in welfare economics. I conclude by showing this analysis to best explain the plausibility of two leading alternative approaches from Hausman and Sugden.

The central project in the field of welfare economics—whose ultimate aim is to offer a proper guide for economists in making policy decisions that affect people's welfare—has arguably been that of determining the precise nature of the connection between preference, choice, and welfare. Debate here has largely focused on two leading approaches. More often than not, economists take preferences to determine choices, which are (therefore) taken to determine changes in welfare (or well-being).[1] On this received view, a person is made better off to the extent that their preferences are satisfied. Since their preferences also determine their choices, economists can measure

[1] For overviews here, see Hausman 2012 (ch. 3) and Sen 1982 (ch. 2).

improvements and reductions in welfare in terms of what is chosen. Something is chosen only if it is preferred, and it makes one better off only if it is preferred and satisfied. Thus, when it is chosen, the preference is satisfied, and the subject is necessarily made better off. All choice then informs us of what is preferred, which in turn informs us of what impacts welfare.

One motivation for this received view is the robust predictions it allows economists to make about how different market and policy changes will affect people's welfare. The tight connection this approach advances between preference, choice, and welfare makes both preference and welfare empirically measurable phenomena, as actual choices, which are of course easily measured, are all that are needed to measure preferences and welfare. But be that as it may, this may also be the *only* motivation for adopting the received view. The proposed connection between preference, choice, and welfare leads to entirely implausible views of welfare, preference, and human motivation more generally. As has been argued at length elsewhere, it cannot be the case that we always choose what we most prefer—at least, not in just any sense of "prefer"—nor can it be that our welfare consists simply in the satisfaction of just any of our preferences.[2] After all, there are clear senses in which we can do what we do not most prefer to do—such as when we act purely out of a sense of duty—and in which we can prefer things that would only make us worse off—such as when we prefer to take on some harm in place of someone else.[3]

The received view, then—or *the revealed preference approach*—takes our preferences to be "revealed" by our choices, and the satisfaction of these preferences, as indicated by our choices, to constitute our well-being. As this approach faces significant problems, an alternative has naturally been introduced. On this second leading approach, preferences are still central, but they do not have as close a connection to choice as the revealed preference theorist claims. The preferences that

[2] See e.g. Hausman 2012 (ch. 7). These claims will be made more precise below.
[3] For some other well-known criticisms of this approach, see Sen 1982.

matter to welfare may not be revealed by choices, on this approach, because a person's "true" or "latent" preferences may not always be expressed by the choices they make. As findings in behavioral economics have recently emphasized, human agents may be rational beings, but they are also imperfect such beings. Sometimes they miscalculate, suffer weakness of will, act on false beliefs, and fail to be perfectly rational in other ways. And these, it seems, are just the instances in which what they *actually* seem to prefer is not revealed by their choices, given their actual choices are just the result of these imperfections. Thus, on *the latent preference approach*, agents have a set of latent preferences, or preferences they would choose in accordance with if they were perfectly rational, and it is the satisfaction of these preferences that economists should be concerned with when considering changes in people's welfare.[4]

Still, this approach, too, faces significant worries. For one, it is unclear how economists could determine what a person's latent preferences are, and so how they could aim at satisfying them, given that as things stand people are far from perfectly rational. Of course, this epistemological worry doesn't show that latent preferences are not what economists should ultimately be focused on in assessing changes in people's welfare; it shows only that their ability to focus on these preferences is constrained. But this also hints at a deeper worry for the latent preference approach. In particular, it is far from clear that satisfying the preferences of one's perfectly rational self would make one's actual self any better off. These two versions of oneself might be quite different in terms of their interests. Thus if one were to satisfy the preferences of one's perfectly rational self, one might be left cold, or even deeply upset, by the result. More importantly, and fundamentally, as in the case of the revealed preference approach, it simply cannot be the case that welfare consists in the satisfaction of just any of one's preferences, whether latent or

[4] See e.g. Sunstein and Thaler 2003 and Thaler and Sunstein 2008; and, for discussion, see Infante et al. 2016. For a latent preference approach to well-being, see Railton 1986.

not. It seems obvious that one may prefer to take on some harm in place of another person without suffering any irrationality, meaning that one's perfectly rational self might prefer this as well. Clearly, though, the satisfaction of one's preference here need not make one any better off. So, if there is indeed some subset of preferences of human agents that perfectly track changes in their welfare—and so a subset of preferences that economists concerned with people's welfare should focus on—it cannot be the one the latent preference theorist identifies.

But if neither the revealed preference approach nor the latent preference approach gets the connection between preference, choice, and welfare right, then what is the correct connection between these concepts, and to what extent does this connection allow for measurement of the sort economists and policymakers concerned with welfare would like? More than just difficult, this question, as presently formulated, may even fail to admit of a single true answer. This is because, as some of the above claims have alluded to, there appears to be more than just one sense of "preference." If that is right, then, presumably, there will be more than just one connection between "preference," choice, and welfare. Indeed, there will presumably be one such connection for each sense of "preference" we specify. Examination of these distinct senses of "preference" and the connections they share with each of choice and welfare may shed light on a number of debates in welfare economics—such as that between revealed preference theorists and latent preference theorists—as well as on the limits of what we should expect or hope for debates in this field to ultimately achieve. More importantly, it may lead to analyses of the connections between "preference," choice, and welfare that should ultimately guide economists in making policy decisions that affect people's welfare.

The aim of this paper is to specify two such senses of "preference," and to examine the implications of this distinction for the connection between, and measurement of, preference, choice, and welfare, as well as for welfare economics more generally. Section 1 introduces and defends these two senses of "preference." Section 2 examines the implications of this distinction for the connections between preference, choice, and welfare, and ultimately offers a novel analysis of the proper role of these concepts in welfare economics. Section 3 discusses the implications of this analysis for the measurement of preference, choice, and welfare. And section 4 discusses two of the leading alternative approaches to these topics offered by Daniel Hausman and Robert Sugden, and concludes that the approach offered here is not only more plausible than, but also best explains the plausibility of, these alternatives.

1. Two senses of "preference"

While the existence of the two distinct senses of "preference" I have in mind has gone largely unrecognized by economists, it has been touched on, if not fully explored, by several philosophers.[5] Perhaps the most notable such treatment has come from Daniel Hausman, who begins his extended analysis and critique of welfare economics by noting a distinction between what he calls "overall" preference and "total" preference. Despite the less-than-fully-illuminating labels, we can understand this distinction in terms of what the two types of preference take into consideration. Whereas overall preferences consider, as Hausman puts it, "*most* of what matters" to a subject, total preferences consider "everything that matters" to them.[6] To clarify, Hausman adds that whereas overall preferences are generally taken to *compete* with some other factors, such as obligations or duties, total preferences, as they take "everything" that matters into account, also account for these other factors. To illustrate, he offers the following examples:

[5] See e.g. Hausman 2012 (pp. 3-4), Sumner 1996 (p. 120), and Wedgwood 2017 (pp. 78-79).
[6] Hausman 2012 (p. 3).

(1) Jill drank water rather than wine with dinner, despite preferring to drink wine, because she promised her husband she would stay sober.

(2) Jill drank water with dinner because she preferred to do so. But for the promise she made to her husband to stay sober, she would have preferred to drink wine rather than water with dinner.[7]

Here there are two entirely natural senses of "prefer," and Hausman takes his distinction to explain them: the first sense corresponds to "overall" preference, while the second sense corresponds to "total" preference. So while the terminology may, again, tend toward confusion, the idea here is that whereas "overall" preference corresponds to some partial aspect of what matters to the subject at the time of choice (and thus competes with Jill's promise), "total" preference corresponds to all of what matters to the subject at that time (and thus accounts for her promise).

This, though, hardly explains the difference between Jill's psychological states in (1) and (2). It is unclear both which aspects of "what matters" to Jill are left out in the case of "overall" preference, and, more fundamentally, what is meant in the first place by "what matters" to her. Of course, this is not to deny that Hausman is on to something in offering his distinction—quite the opposite. There are clearly two distinct senses of "preference" here. But his analysis is not precise enough to allow for the kind of robust implications about the connection between preference, choice, and welfare, or about the state of welfare economics, that we are looking for here.

[7] Hausman 2012 (p. 3).

Consider then what is meant by "what matters" to Jill.[8] Presumably this means something like: what is (and isn't) in Jill's interest. And this, presumably, means something like: what promotes or contributes to (and hinders) Jill's goals or needs. More precisely, it seems to mean: what would satisfy or frustrate Jill's desires and (their negative counterpart) aversions. To be sure, there are a number of leaps in reasoning here, but consider: as it relates to a choice scenario, it seems that what matters to someone in that scenario will be affected just in case the satisfaction or frustration of something they desire or are averse to is affected in that scenario. It matters to Jill, for instance, whether she has wine with dinner, both because she desires to have it, and because she is averse to breaking her promise. If she had just one of these attitudes, it would still matter to her whether she had the wine, either because she simply wanted it, or because she was simply averse to breaking her promise. Yet if she had neither of the attitudes, and no other desires or aversions relating to the choice, then it would not matter to her whether she had wine with dinner—she would be indifferent. While this single case certainly does not settle the matter, it at least gives us good reason to think that what matters to a subject, and so what preferences serve to take into account, is a function of that subject's desires and aversions.[9,10]

You may find this explanation of preference in terms of desire and aversion implausible. But notice that the proposed link here leads naturally to a particularly illuminating explanation of our two senses of "preference," and an explanation of which of our interests are bracketed in the case of just one of these senses. This explanation begins by pointing out the intuitively obvious yet oft-neglected difference between desires had when we are genuinely attracted to or enthused by something, and desires whose possession is merely entailed by our being disposed to act so as to

---

[8] My focus here on analyzing what it is for something to matter to a subject is meant only as an intuitive way of reaching, from Hausman's starting point, a deeper analysis of what preferences serve to take into account. Whether we commit to the former analysis is largely unimportant for present purposes.
[9] Davis (1981) offers a similar analysis.
[10] I'll sometimes speak just of desire or aversion below, but I take the pair, as positive and negative compliments of one another, to be fundamentally relevant to a subject's preferences.

satisfy them in certain circumstances. That is, sometimes when we desire things, we're genuinely attracted to the things—we're not merely disposed to get them, but are enthused or excited by them. Following Chris Heathwood, we can call desire of this sort *genuine* desire.[11] On the other hand, there are also cases where our desiring something entails only that we are disposed to get the thing in certain circumstances. We have desires of this sort when, for instance, we visit the dentist despite loathing the very thought of doing so. Since we go to the dentist, we must have in one sense been motivated to go, and must have in that same sense desired to go. Clearly, though, it's not the case that we genuinely desired to go; and, indeed, what we genuinely desired to do was *not* to go. Call desire in this stripped-down, merely motivational sense *behavioral* desire.

It seems clear that "desire" admits of these two senses. It also seems clear that "preference" admits of two similar senses. The most plausible explanation for this, it appears, is that "preference" has this feature because "desire" does. That is, there are two senses in which we can prefer because there are two senses in which we can desire, and preference is partly constituted by desire. To see this, consider the following analogs of Hausman's examples:

(1*) Jill drank water with dinner, despite desiring to drink wine, because she promised her husband she would stay sober.

(2*) Jill drank water with dinner because that is what she most desired to do. But for the promise she made to her husband to stay sober, she would have most desired to drink wine.

---

[11] Heathwood 2019. The same distinction has been made in different terms by a number of philosophers—see e.g. Lewis 1988 (p. 323), Parfit 2011 (p. 43), and Sumner 1996 (p. 120).

Here we can see a clear parallel between our two senses of "desire" and Hausman's two senses of "preference." In (1*), Jill's desire is genuine—just as, in (1), her preference is "overall;" and in (2*), Jill's desire is behavioral—just as, in (2), her preference is "total." As we have seen, in the case of preference, it seems some considerations, like Jill's promise, compete with her overall preference, but not her total preference. This is naturally explained by the current analysis: acting on her promise is clearly not something to which Jill is genuinely attracted, and so this consideration competes with her overall preference, or with what her genuine attractions and aversions tell in favor of; yet, when deciding on her act, Jill clearly took her promise into account, which means that her promise factored into her total preference, or into what she most desired, in the behavioral sense, to do. Given the plausibility of this explanation, as well as the relative inconvenience of Hausman's terminology, we can again follow Heathwood's lead and call preferences which take all of a subject's relevant desires and aversions into account *behavioral* preferences, and preferences which take only a subject's relevant genuine desires and aversions into account *genuine* preferences.

In addition to plausibly explaining the two senses of "preference" that Hausman points out, the analysis on which preference is partly constituted by desire accounts for what would otherwise be a striking overlap in the fundamental features of desire and preference. Consider the following typical features of preferences: (i) they entail motivation, or dispositions to act so as to make their objects more likely to obtain; (ii) they tend to cause pleasure when they are seen as or imagined to be satisfied, and displeasure when they are seen as or imagined to be frustrated; (iii) they tend to direct our attention toward things we associate with their objects; and (iv) they tend to have amplified such effects when we have sensory or imaginative representations of things we associate with their objects.[12] Thus, when Jill prefers to have wine rather than water with dinner, she is disposed to act so as to make it more likely that she will have wine; she will tend to experience pleasure when she

[12] For discussion of these features as they relate to desire, see Sinhababu 2017 (ch. 2).

sees herself as having wine with dinner or imagines that she will, and displeasure when she sees herself as not having wine or imagines that she won't; she will be apt to attend to things like the wine glass on the table or corkscrew in the drawer; and these effects will be amplified when she sees, smells, or imagines the taste of, the wine. All of these things will seem equally true of her when she desires to have the wine. The most plausible explanation here seems to be that these things are true of her when she has this preference in virtue of their being true of her when she has this desire.

It may be wondered whether this analysis puts things the wrong way around. Granted, desire and preference share these features. But perhaps desire is partly constituted by preference, rather than the other way around.[13] While this suggestion is worth considering, however, there seem to be a number of strong reasons to reject it.[14] For one, preferences are essentially comparative, whereas desires are not. You cannot prefer A when A is all there is. Of course, you can prefer A to nothing, but then there is something else, namely nothing, to which A is preferred. In contrast, you can clearly desire A when A is all there is. It does not matter whether there is anything else—you desire A just the same. This suggests that preference is a comparison or ranking of something more basic, like desire. For another, it seems that you cannot have a preference as to A and B when you are affectively neutral as to everything relevant to A and B. If you have no desire for or aversion to anything relating to either A or B, it seems that you cannot have a preference with respect to them. Of course, you can have a preference when you are affectively neutral as to everything relating to one of A and B: you may desire or be averse to something relating to the other. But this is not

---

[13] This seems to be the view of e.g. Spurrett (2021). See also Lichtenstein and Slovic 2006; and, for discussion, see Schroeder 2020.

[14] An additional such reason is due to Pollock (2006, pp. 22-27), who argues that the number of basic pairwise preferences that would be required just to determine facts about the extent to which any given thing was preferred over any other given thing would be far greater than the number of particles in the universe. Thus if preferences were the more basic attitude, facts about the strengths of desires could only be determined by an astronomically large number of more basic facts, somehow stored in the brain, about pairwise preferences. Alternatively, if desires were the more basic attitude, Pollock suggests, the very same facts about the strengths of desires and preferences could be determined by just 600 basic facts about desires. And so, taking desires to be the more basic attitude seems the more psychologically realistic view.

possible when you are neutral as to all that relates to both. If, for instance, there is nothing relating to the fact that 2+2=4 that you have a desire for or aversion to—if you are affectively neutral as to everything relating to that fact—and you stand in the same position toward the fact that 3+3=6, you cannot prefer one fact to the other. So, it seems, you cannot have a preference without having a relevant desire or aversion. In contrast, you can desire A or B without having any preference as to A or B. If, for instance, you now desire a cold glass of water, and also desire not to miss your flight later, you may not prefer one of these things to the other, because you have not been compelled to weigh them against each other—or, because you have no reason to think they are in tension. Yet if you are then told that you will miss your flight if you go to get the water, your newly conflicting desires will suddenly give way to a preference for one over the other. Consequently, it seems you can have desires and aversions without having relevant preferences, but cannot have preferences without having relevant desires or aversions. This, too, suggests that desire is more basic than, and indeed partly constitutes, preference.

It seems, therefore, that there are two senses of "preference," and that this is ultimately explained by the facts that there are two senses of "desire," and that desire is partially constitutive of preference. While I take this explanation to offer the best motivation for and illumination of the proposed distinction between genuine and behavioral preferences, though, it is also worth noting that the distinction can be accepted even without endorsing my preferred explanation of it. Thus even those who reject the view that preference is partially constituted by desire can accept the intuitive distinction between genuine and behavioral preferences, and so follow us in exploring the implications of this distinction for debates in welfare economics. As we have seen, it seems clear, more generally, that recognizing this distinction in types of preference is crucial to assessing claims about preference: even something as simple as whether Jill prefers to drink wine or water with dinner cannot be determined without specifying the sense of "prefers" we invoke. When economists and others typically talk of preference, however, this distinction goes entirely ignored. As may be expected, this has led to some deep tensions in the ideas and goals of welfare economics, whose most fundamental challenge has arguably been to determine the precise nature of the connection between preference, choice, and welfare.

2. The connections between "preference," choice, and welfare

Recall that on the revealed preference view in welfare economics, well-being consists in the satisfaction of preference, and something is chosen only if it is most preferred. While this view faces many familiar problems, it is not entirely obvious exactly where it goes wrong. Indeed, each of its claims even seems to have a certain ring of truth: we tend to be better off when our preferences are satisfied as opposed to when they are not, and our choosing something seems to entail that we were most motivated, and so most preferred, to choose that thing. Once we recognize the distinction between genuine and behavioral preference, however, we can see precisely where this approach goes wrong.

Since behavioral preference is just preference that takes every motivating factor in a choice situation into account, it seems clear that the connection between behavioral preference and choice must be quite strong. Indeed, it seems that strength of behavioral preference is most plausibly understood as a function of hypothetical choice: the more circumstances in which one would (be most motivated to) choose the thing, the stronger one's behavioral preference. More importantly, since it follows by definition that one always chooses in accordance with one's strongest behavioral preference, the connection between this sort of preference and choice is even necessary. Obviously enough, though, the connection between choice and *genuine* preference can be nowhere near this strong. After all, we constantly choose to do things that we do not most genuinely prefer to do: we go to the dentist despite genuinely preferring to stay home; we go to the early meeting despite genuinely preferring to stay in bed. Every day we make choices that do not reflect our strongest genuine preferences. So we have, on the one hand, a sense of "preference" that shares a necessary connection with choice; and, on the other hand, a sense of "preference" whose connection to choice is not only not necessary, but too weak to even be evidential.

Of course, this divergence alone does nothing to threaten the revealed preference approach, since it might be that the type of preference that shares a necessary connection with choice also shares the sort of connection with welfare that the approach proposes. That is, it might be that the satisfaction of behavioral preference is what constitutes well-being. In that case, the revealed preference approach will get things exactly right: we choose something only if it is what we most (behaviorally) prefer, and our well-being is enhanced if and only if our (behavioral) preferences are satisfied; thus, when we choose something, we are made better off by getting that thing. It again seems obvious, however, that behavioral preference shares no such connection with welfare. Indeed, despite its necessary connection with choice, behavioral preference shares a merely incidental connection with welfare—its satisfaction is neither necessary nor sufficient for, but may sometimes occur alongside, improvements in welfare. If, for instance, you most behaviorally prefer to be tortured for ten minutes rather than to be tortured for twenty minutes, your being tortured for ten minutes will not in itself make you better off; and if you are unaware that having a delicious meal is available as an alternative to being tortured for any amount of time, your getting the delicious meal might in itself make you better off despite your failing to behaviorally prefer this. Still, sometimes we manage to choose what is best for us: you might behaviorally prefer to have the delicious meal upon learning that it is an option. Sometimes, then, our behavioral preferences are satisfied, and our welfare is improved. But not always. And, arguably, not even that often. Like choice, though, welfare does seem to share a close connection with one of our types of preference. In this case, however, it is genuine preference. Just as behavioral preference could be plausibly linked to choice, so genuine preference can be plausibly linked to welfare.[15] Indeed, there is a good deal of plausibility in the claim that a subject's welfare is enhanced if and only if their genuine preferences are satisfied.[16] After all, if they genuinely prefer something (like the delicious meal, but not like the ten minutes of torture), then they are more genuinely attracted to that thing than some other thing; and so, between getting the one thing or the other, they seem to do better by getting what they are more genuinely attracted to. Of course, it might be that they now genuinely prefer something that will later be bad for them, like addictive drugs, to something else that will later be better for them, like sobriety. But, at least at the time of getting the drugs rather than sobriety, considered in isolation, it seems plausible that their welfare is enhanced by the enjoyment they experience. So, there is no satisfaction of genuine preference without an (immediate) improvement of welfare here.[17] More generally, and precisely, it seems plausible that whenever one's genuine preferences appear to one to be satisfied,

[15] In fact, Heathwood's (2019) aim in introducing his distinction between behavioral and genuine desire is to show that only genuine desires are directly relevant to well-being. Given my arguments that (genuine) preference is partly constituted by (genuine) desire, then, his conclusions seem to support the current point.

[16] Importantly, however, I myself will not endorse the "only if" portion of this biconditional when making my view more precise just below. This is because, as I've argued earlier, it seems possible for a subject to have a genuine desire for A without also having formed a genuine preference for A (over, say, B). And since, plausibly, the (apparent) satisfaction of a genuine desire for A entails the same sort of intuitively-welfare-promoting experience as the (apparent) satisfaction of a genuine preference for A, it seems this sort of experience can be had, and welfare can intuitively be enhanced, without a corresponding genuine preference having even been formed, let alone (apparently) satisfied.

More importantly, though, since this line of reasoning clearly depends on my earlier argument for the view that preference is partly constituted by desire, those who resist this argument or view more generally do not face similar pressure to refrain from endorsing the biconditional in full. That is why I mention it as an alternative here.

[17] As I appeal explicitly to enjoyment in suggesting that the subject is made better off in this case, it may be wondered, more generally, whether the account offered here ultimately constitutes a form of hedonism, which takes pleasure (or enjoyment) to be what is ultimately good for us. After all, when we genuinely prefer something and appear to be getting it, it seems plausible to categorize us as simply being pleased. Here I have three points in response. First, as Heathwood (2019) similarly notes, the debate over the nature of pleasure is contentious, and nothing I say here requires taking a robust stand on it. The account I'm offering concerns the connection between genuine preference and welfare; should the true theory of the nature of pleasure turn out to be relevant here, so be it. Second, even if this theory of pleasure implies that the apparent satisfaction of genuine preference constitutes pleasure, it still might be that the true theory of *welfare* turns out to be either hedonism, or a pluralistic view on which pleasure is one of multiple things that directly contribute to well-being, and in either case the claims here regarding the connection between genuine preference and welfare would be vindicated. Finally, to the extent that apparent genuine preference satisfaction does overlap with pleasure, this may be a welcome result from a practical perspective, as a good deal of psychological research on the measurement of pleasure and its relation to welfare has recently been inspired by (most notably) the work of Daniel Kahneman. See e.g. Kahneman et al. 1999.

one is made at least somewhat better off.[18] At the very least, this seems to be the closest we will get to plausibly linking the satisfaction of "preference" with improvement in welfare.

But regardless of how plausible we find this approach, it's worth taking a step back to notice the implications of these arguments for the nature of the connections between "preference," choice, and welfare. The revealed preference approach takes (i) well-being to consist in the satisfaction of "preference," and (ii) choice to be determined by strength of "preference." Once we recognize the distinction between genuine and behavioral preference, however, we are able to see that any plausibility had by this view is based on a conflation of two senses of "preference." Its first claim has a ring of truth in virtue of the nature of genuine preference, and its second claim has a ring of truth in virtue of the nature of behavioral preference. As we have just seen, however, its first claim also seems clearly false in virtue of the nature of behavioral preference, and its second claim seems clearly false in virtue of the nature of genuine preference. Thus, neither sense of "preference" outlined here allows for a plausible connection between itself, choice, and welfare of the proposed sort. And so, some part of the revealed preference approach must be given up. The connections between these concepts cannot be as strong as many economists and philosophers had hoped.

---

[18] This is formulated in terms of apparent satisfaction, rather than actual satisfaction, so as to avoid counterexamples involving genuine preferences that are satisfied despite their subject failing to know it. Thus, for instance, I might genuinely prefer for your life to go well, and it might go well despite my being entirely unaware of it (Parfit 1984, p. 494). In this case, the satisfaction of my preference does not seem to make me any better off, since I never actually consider its satisfaction. Its satisfaction only seems good for me provided it appears to me to be satisfied—or, provided the thing I'm genuinely attracted to appears to me to be the case.

Of course, this focus on apparent or "subjective" satisfaction as opposed to actual or "objective" satisfaction brings with it a number of familiar worries. Most notably, cases where subjects are deceived into falsely believing that their genuine preferences—for, say, having a loving family and successful business—are being satisfied seem to suggest that apparent satisfaction cannot be all that matters to well-being (Kagan 1994). After all, the deceived subject's appearances might be identical to those they would have if they were not deceived, yet if they were not deceived they would intuitively seem better off. As I see it, however, these cases aren't particularly troubling for the current proposal. This is because, even when subjects are deceived, the apparent satisfaction of their relevant genuine preferences, in itself, still intuitively makes them *at least somewhat* better off. This at least puts the apparent satisfaction approach in a better position to respond to its worries than the objective satisfaction approach, since the cases undermining the latter approach seem to show that the objective satisfaction of preferences sometimes intuitively makes subjects *no* better off.

In any case, having noted this qualifier about appearance in the view, I will largely omit it in what follows.

Consider next the implications of the current analysis for the alternative, latent preference approach in welfare economics. This approach takes (i) well-being to consist in the satisfaction of perfectly rational "preference," and (ii) choice to be determined by strength of actual (imperfectly rational) "preference." Again, any apparent plausibility here is explained by the conflation of genuine and behavioral preference. As in the case of the revealed preference approach, its second claim has a ring of truth in virtue of the nature of behavioral preference, which is defined so as to make the claim a necessary truth. Its first claim, on the other hand, has a ring of truth not just in virtue of the nature of genuine preference, but also in virtue of the fact—which ultimately motivates the move from revealed preferences to latent preferences with respect to analyzing well-being—that our perfectly rational selves would presumably come closer to maximizing the satisfaction of our genuine preferences than we do: with full information and rationality, they would presumably make fewer errors in bringing about states of affairs that were most genuinely attractive. Still, since this claim ignores the distinction between genuine and behavioral preference, it implies that the satisfaction of the behavioral preferences of our perfectly rational selves would improve our welfare in just the same way the satisfaction of their genuine preferences would. And this means that the approach implies—implausibly, as we've seen—that the satisfaction of our perfectly rational selves' merely behavioral preference to take on some harm in place of another person would directly improve our well-being.[19] Again, then, both the apparent plausibility and ultimate failure of the latent

[19] Here it might be objected that my own view faces a similar worry. Specifically, if one had a genuine preference to take on some harm in place of another person, the view might seem to imply that the satisfaction of this preference would make one better off, despite clearly harming one. Here, however, two things are worth noticing. First, it is not the actual satisfaction, but the apparent satisfaction, of this genuine preference that the view claims would make one better off; and second, the view claims just that this apparent satisfaction would *in itself* make one *at least somewhat* better off. Thus if one genuinely preferred to take on the harm, and ultimately did so, the view implies that one would be made *at least somewhat* better off *specifically in virtue of* the thing one was attracted to appearing to one to be the case. It does not imply that actually saving the other, in itself, would make one better off; nor that doing so would make one better off, all things considered. Indeed, so long as the harm involved in saving the other entailed the apparent satisfaction of some of one's genuine aversions (like one's genuine aversion to pain) or the apparent frustration of one's genuine desires (like one's genuine desire to exercise), and these experiences outweighed the fleeting experience of apparently satisfying one's genuine desire to save the other, the proposed view would tell us that saving the other made one worse off, all things

preference approach is plausibly explained in terms of the conflation of behavioral and genuine preference.

We are left, then, with the following analysis of the connections between "preference," choice, and welfare. Behavioral preference, by definition, shares a necessary connection with choice: our choices necessarily reveal our behavioral preferences, in that we necessarily choose what we most behaviorally prefer. Nonetheless, behavioral preference shares a merely incidental connection with welfare: it is only when we choose in accordance with our genuine preferences, which then by definition overlap with our behavioral preferences, that the satisfaction of our behavioral preferences is accompanied by a change in welfare. Genuine preferences, on the other hand, share a close connection with welfare: their apparent satisfaction seems in every case to come with at least some improvement in welfare. Yet, as evidenced by choices we make every day, our choices do not always, and arguably do not even often, accord with what we most genuinely prefer. It seems, therefore, that neither sense of "preference" can play the role that economists concerned with changes in people's welfare would like. Neither seems to share a close enough connection with both choice and welfare to warrant the measure of welfare in terms of choice.

3. The measurement of "preference," choice, and welfare

Assuming the current analysis is correct, is there any hope of carrying on with the welfare economist's goal of reliably measuring welfare, in terms of choice, via preference? Well, while there still may be hope of reliably measuring welfare in terms of preference—at least in one sense of "preference"—it seems that the sense of "preference" that would allow for such reliable measurement could not be reliably measured in terms of choice, and so that any reliable measure of

considered, even though the apparent satisfaction of one's genuine desire to save the other, in itself, made one at least somewhat better off. This seems to me exactly right.

welfare in terms of preference will not focus on choices. That is, since genuine preference shares a close connection with welfare, but not choice, and behavioral preference does not share a close connection with welfare, the current analysis implies that any reliable measure of welfare via preference will not focus on choices. Insofar as the economist is interested in measuring changes in well-being via preference, the type of preference she should be concerned with is, not behavioral preference, but genuine preference. And since genuine preferences do not share a close connection with choices, she will have to find some other way of measuring the satisfaction of genuine preferences to get to evidence of improvements in welfare. The challenge, then, is to determine how the satisfaction of genuine preference might be measured.

Immediately, however, a serious problem arises: it is not at all clear how to reliably measure the contents of people's mental states. While we could, if we knew these contents, in many cases easily measure whether they were satisfied, there does not appear to be any easy way of determining the precise contents themselves. Of course, we might try the most obvious method available here, which is to simply ask people what they genuinely prefer. But this method will only take us so far. After all, people are often mistaken about their own mental states, especially their affective or emotional ones.[20] What's more, people often succumb to the sorts of failures that the latent preference theorist was concerned to rule out: they are often unsure of all the relevant alternatives, and suffer from failures of rationality like weakness of will. In these cases, people's reports of their own genuine preferences may not reflect what they actually genuinely prefer, and so may not be a reliable measure of what would enhance their welfare.

So simple self-reports will not be enough. Is there any other way we might reliably measure the satisfaction of people's genuine preferences? Well, one optimistic suggestion is that the satisfaction of these preferences will be reliably tracked by the satisfaction of people's behavioral

[20] See e.g. Haybron 2008 (ch. 10).

preferences. Thus, when people make choices, and so act on their strongest behavioral preferences, they provide reliable evidence of what they most genuinely preferred when making those choices. But this suggestion is obviously implausible. As we've seen, we constantly do things that we do not genuinely prefer to do—and, indeed, things that we genuinely prefer *not* to do—in our everyday lives. If self-reports are not reliable enough to reliably measure genuine preference, then behavioral preference will not even come close.

If there is any way of reliably measuring the satisfaction of people's genuine preferences, it is at least not immediately clear what it might be.[21] So as to avoid an entirely negative conclusion here, though, it may be worth mentioning (however briefly) one alternative that welfare economists might adopt in attempting to track and improve people's welfare. This is to focus on *opportunity for* (apparent) genuine preference satisfaction. This would presumably leave them to focus on people's access to things like health, work, hobbies, social groups, and other things that have an overwhelming tendency to promote (apparent) genuine preference satisfaction. In addition to securing at least some evidential link to people's welfare, this approach would also leave the actual realization of the relevant opportunities up to the people themselves, and so with respect to policy decisions would seem consistent with considerations of liberal neutrality. While this is, admittedly, no more than the most cursory of sketches of the alternative in mind, it at least seems worth offering in hopes of future development.

[21] At least at present. In principle, though, given the ultimate focus here is on the psychological phenomenon of apparent genuine preference satisfaction, it certainly seems that the particular portion of the brain associated with this phenomenon could ultimately be discovered and measured directly. Also, as I've mentioned, the path to such reliable measurement may be considerably clearer if there turns out to be a significant enough overlap between the apparent satisfaction of genuine preference and pleasure—see fn. 17.

4. Hausman and Sugden on welfare economics

Before closing, it's worth considering how the current analysis of the connection between "preference," choice, and welfare compares to, and ultimately improves on, influential analyses offered by other welfare economists. Consider first then Daniel Hausman's view of how economists should understand the connection between these concepts. Since, as ongoing debate illustrates, we are of course unsure of what well-being consists in, Hausman notes that relying on any particular proposal for the purposes of welfare policy would seem unjustified. Nonetheless, he thinks, economists can instead rely on certain "platitudes" about well-being, which together constitute a "folk theory" of the concept.[22] These include claims such as "having good friends typically makes people's lives better," and "other things being equal, adequate nutrition makes a person better off." Now, since Hausman takes the sense of "preference" that is relevant to welfare economists to be behavioral preference (or, in his terms, "total" preference), his aim is to link this "folk" conception of welfare to behavioral preference and thus choice. So, his question is: if economists accept his folk theory of well-being, under what conditions can they justifiably take people's choices, and so the satisfaction of people's behavioral preferences, to indicate improvements in well-being? And his answer is, when the people meet the following three conditions: (i) they are adequately informed regarding their feasible option set; (ii) they exercise good judgment over those options; and (iii) they choose on the basis of self-interest—or, on the basis of what would promote their own well-being.[23] Plausibly, when people meet these three conditions—which economists can determine by appealing to Hausman's simple folk theory—economists are justified in taking people's choices, and so the satisfaction of their behavioral preferences, to provide evidence of what is good for them.

[22] Hausman 2012 (ch. 8).
[23] Hausman 2012 (p. 89).

The approach offered here seems well-positioned to explain the apparent plausibility of Hausman's view. The idea of choosing on the basis of "self-interest" can be plausibly understood as meaning that one chooses on the basis of one's relevant genuine desires and aversions, or on the basis of all of what one is genuinely attracted and averse to in the choice context. So, when people meet Hausman's third condition, it seems plausible that they are choosing from the apparent alternatives on the basis of what they most genuinely prefer; which, the current account says, is precisely what is needed for well-being to be improved. However, since people might be unaware of some option that they would genuinely prefer (or would genuinely prefer even more), we must also assume that the people are adequately informed in order to conclude that their choice really is evidence of what is good for them—that is, we must assume his first condition. And since people sometimes act irrationally, we must also assume that their choice here is not the result of any obvious such error—that is, we must also assume his second condition. Putting all these together: when Hausman's conditions are met, the relevant subjects rationally satisfy their strongest genuine preference in the relevant circumstances. That is, with respect to their genuine desires and aversions and the conditions they're in, they choose optimally. In that case, it is clear why it seems plausible that economists are justified in taking their choices to provide evidence of what is good for them here. They are doing the very best they can relative to both their genuine attractions and circumstances.

Now, obviously enough, this explanation of Hausman's view in terms of the current analysis does little, if anything, to threaten his view. It is simply an explanation of why his view seems to have some intuitive plausibility. Notice, though, that if we adopt the analysis offered here of behavioral and genuine preference, and the connections each shares with choice and welfare, then we have available a deeper or more basic explanation of why Hausman's view might be true. This suggests that the current analysis goes further than Hausman's own in getting at the precise nature of the connections between "preference," choice, and welfare. And this suggests that, regardless of whether we agree with Hausman's approach, its truth or apparent truth will ultimately be best explained in terms of the analysis offered here.[24]

Consider next Robert Sugden's view, which departs from the analyses discussed here by focusing on options, rather than choices or preferences.[25] Earlier, I mentioned a potential approach to normative welfare economics that focuses on opportunity for (apparent) genuine preference satisfaction. The motivation behind this approach was that, if genuine preference is indeed the sort of preference that is relevant to welfare, then because genuine preferences, as mental states, are exceedingly difficult to measure or track, we might do better to focus on people's opportunities for genuine preference satisfaction, since this is more easily measurable in terms of the things like jobs and health resources that are available to them, than we would by focusing on people's actual mental states. Still, these mental states are ultimately what we want to get at, on this approach—they are what economists should ultimately aim to track or measure. Sugden's view is similar in that it claims that economists should be focused on something antecedent to preference in designing policies that affect welfare, but it is dissimilar in that it claims that this antecedent thing is also what economists should ultimately be aiming to track or measure. That is, Sugden argues that *opportunities* should be what economists focus on in designing policies that affect people's welfare, and that this is because opportunities themselves, rather than any preference satisfaction they might allow for, are ultimately the thing that matters or is good for people. Thus, on this view, "as viewed by each citizen separately, more opportunity for that person is better than less," and "individuals collectively should have as much opportunity as possible to carry out whatever voluntary transactions they might want

24 A similar point holds for any version of the latent preference approach that adopts conditions like Hausman's (requiring all relevant information, good judgment, and pure self-interest) in specifying the set of preferences whose satisfaction directly contributes to well-being.

25 Sugden 2018 and Sugden 2019.

to make" (Sugden 2019, p. 421). The idea is then that opportunities are not only a useful way of measuring what should be of ultimate importance to welfare economists, but are themselves what is of ultimate import.

Again, it seems any apparent plausibility of Sugden's view is ultimately explainable in terms of the analysis offered here. Opportunities are good for people in virtue of the fact that they are necessary means to satisfying preferences: a person obviously cannot satisfy their preferences if there are no opportunities for them to do so. Still, to say that opportunities are the source of goodness in this equation seems to mistake the means to goodness for goodness itself.[26] After all, we can easily imagine a subject who has "as much opportunity as possible to carry out whatever voluntary transactions they might want to make" throughout their entire life, yet who spends their entire life sleeping or inactive. In this case, it seems implausible to say of this person's life after it has ended that it went just as well as a life with the same amount of opportunity as well as a great deal of genuine preference satisfaction. To take the point to the extreme: Sugden's view implies that a public policy designed to ensure that everyone alive will sleep, or at least choose not to act, for the rest of their lives might do just as well in terms of making people's lives go well as a public policy designed to maximize everyone's lifetime (apparent) genuine preference satisfaction, provided each policy ensured that the people had the same amount of opportunity to carry out potential voluntary transactions.[27] That seems, to me, implausible. And while this is of course a fanciful example, I think it gets at an important point about the philosophical foundation of Sugden's view. Specifically, it

[26] For a recent, related argument against the view that freedom of choice is intrinsically valuable, see Gustafsson 2020.
[27] Of course, a similar objection might be leveled against my opportunity-based proposal for measuring changes in people's genuine preference satisfaction. This proposal too may see little difference between two policies that were equal in terms of the opportunity for genuine preference satisfaction they allowed for—at least initially. However, unlike Sugden's proposal, it would certainly recognize a difference between the policies if just one were to actually lead to improvements in genuine preference satisfaction. This is because it recognizes that what economists should ultimately be concerned with—improvements in genuine preference satisfaction—would be realized with only one of the policies, making the other policy a failure with respect to what ultimately matters. Sugden's approach, in contrast, makes no distinction between opportunity's practical use and its ultimate importance, and so cannot recognize any similar failure in the analogous case.

suggests that while opportunities might be a good target for welfare economists' focus in measuring and tracking what ultimately matters, they are not themselves what ultimately matters. Instead, what ultimately matters is how well people's lives go. And whereas, as we've seen, the quality of people's lives can be plausibly analyzed in terms of genuine preference satisfaction, their quality does not seem plausibly analyzable in terms of opportunity. Thus, while opportunities might represent an effective means of measuring or tracking what should ultimately matter to welfare economists, they do not seem to be what should be taken to ultimately matter themselves.[28]

In any case, there does seem to be a ring of truth in the thought that having more options is invariably better for people than having fewer. This, I think, is because genuine preferences could in principle have any object. That is, any possible alternative could in principle be genuinely preferred by a subject over any other. That being so, for any possible alternative, having the alternative available to one could allow one to satisfy one's potential genuine preference for it; and not having that alternative available could only prevent one from satisfying one's potential genuine preference for it. If one never actually genuinely prefers the alternative, no harm is done: one would never have attempted to satisfy it in the first place. But having the alternative available leaves open that one's potential genuine preference for it could be satisfied. To put the point slightly differently: it's never in itself bad to have more alternatives available to one.[29] Either one does not genuinely prefer the alternative, in which case the alternative is irrelevant to one, or one does genuinely prefer the alternative, in which case one can attempt to bring it about. Since this is true of every possible alternative, it follows that having more alternatives available to one could in itself only be good, and

[28] Arneson (1999) appeals to a similar point in renouncing his view of egalitarian distributive justice on which the "currency" of egalitarian justice, or what justice requires that people be made equal in, is opportunity for welfare.
[29] Here it might be objected: wouldn't it be in itself bad to have more alternatives available to one, at least given the view of welfare proposed here, if one simply genuinely preferred to have fewer alternatives? It would not. This is because, even here, what would be in itself bad for one on the view could not be the fact that one actually had more alternatives (or the actual frustration of one's genuine preference), but could only be the appearance one had of having more alternatives (or the apparent frustration of one's genuine preference).

could never be bad. Of course, sometimes having too many alternatives available to one can lead to excessive indecision. And that might be bad. But here it is not the having of the alternatives itself, but rather the lack of decisiveness, that is bad. Accordingly, for any alternatives, having more could in itself only be good, and having fewer could in itself only be bad. Importantly, though, and as this discussion illustrates, this is not because having more alternatives is in itself good for one and having fewer alternatives is in itself bad for one. Rather, these things are good and bad for one in virtue of their leaving open or closing the possibility of one's genuine preferences for them being satisfied. In that case, it is again the satisfaction of genuine preference that explains the plausibility of Sugden's key hypothesis.

I conclude, therefore, that the current analysis should be preferred to Sugden's. It offers a plausible explanation of the apparent plausibility of Sugden's view, reaches more intuitively plausible verdicts about particular policy choices, and offers a more plausible account of what should ultimately matter to economists with respect to well-being public policy.

5. Conclusion

Let me sum up. Economists who are concerned with changes in welfare have generally focused on preference, and have taken one of two approaches to analyzing preference's connections to choice and welfare. As we've seen, these revealed preference and latent preference approaches fail, in large part due to their neglect of the distinction between genuine and behavioral preferences. Once we recognize this distinction, we see not only where exactly these approaches went wrong, but also what it is that welfare economists should be focused on in making policy decisions that affect people's welfare. Specifically, they should be focused on the apparent satisfaction of people's genuine preferences. But since, as a mental state, such apparent satisfaction does not seem amenable to reliable measurement, welfare economists must likely instead focus on something else, perhaps like opportunity for apparent genuine preference satisfaction, in attempt to measure or track how different policies might affect apparent genuine preference satisfaction. In addition to its intuitive plausibility, this approach plausibly explains both the ring of truth in, and ultimate inadequacy of, two other leading analyses of these issues in welfare economics. Each of these analyses seems plausible to the extent it does in virtue of facts about the distinction between behavioral and genuine preference, as well as facts about the ultimate importance of apparent genuine preference satisfaction for welfare. I conclude, then, that the current analysis significantly clears up and advances the central debate in the field of welfare economics, or the debate regarding the precise nature of the connections between “preference,” choice, and welfare.[30]

[30] Many thanks to two anonymous referees for helpful comments. And special thanks to Tyler DesRoches for many helpful comments, suggestions, and discussions.

References

Arneson, R. (1999). "Equality of Opportunity for Welfare Defended and Recanted." *Journal of Political Philosophy* 7: 488-497.

Davis, W. (1981). "A Theory of Happiness." *American Philosophical Quarterly* 18: 111-120.

Gustafsson, J. E. (2020). "A Paradox for the Intrinsic Value of Freedom of Choice." *Noûs* 54: 891-913.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.

Haybron, D. M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.

Heathwood, C. (2019). "Which Desires Are Relevant to Well-Being?" *Noûs* 53: 664-688.

Infante, G., G. Lecouteux, and R. Sugden. (2016). "Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioral Welfare Economics." *Journal of Economic Methodology* 23: 1-25.

Kagan, S. (1994). "Me and My Life." *Proceedings of the Aristotelian Society* 94: 309-324.

Kahneman, D., E. Diener, and N. Schwarz (eds.). (1999). *Well-Being: Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Lewis, D. (1988). "Desire as Belief." *Mind* 97: 323-332.

Lichtenstein, S. and P. Slovic. (2006). *The Construction of Preference*. New York: Cambridge University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

---. (2011). *On What Matters. Vol. 1*. Oxford: Oxford University Press.

Pollock, J. (2006). *Thinking About Acting: Logical Foundations for Rational Decision Making*. New York: Oxford University Press.

Railton, P. (1986). "Facts and Values." *Philosophical Topics* 24: 5-31.

Schroeder, T. (2020). "Desire." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). https://plato.stanford.edu/archives/sum2020/entries/desire/

Sen, A. (1982). *Choice, Welfare and Measurement*. Oxford: Blackwell.

Sinhababu, N. (2017). *Humean Nature*. Oxford: Oxford University Press.

Spurrett, D. (2021). "The Descent of Preferences." *British Journal for the Philosophy of Science* 72: 485-510.

Sugden, R. (2018). *The Community of Advantage: A Behavioral Economist's Defence of the Market*. Oxford: Oxford University Press.

---. (2019). "The Community of Advantage." *Economic Affairs* 39: 417-423.

Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.

Sunstein, C. and R. Thaler. (2003). "Libertarian Paternalism Is Not an Oxymoron." *The University of Chicago Law Review* 70: 1159-1202.

Thaler, R. and C. Sunstein. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Wedgwood, R. (2017). "Must Rational Intentions Maximize Utility?" *Philosophical Explorations* 20: 73-92.