ABANDONING THE PARETO PRINCIPLE

Stephen Ellis
University of Oklahoma

ABSTRACT: Standard economic theory assigns a dual role to preferences: in positive economics they help determine behavior; in normative economics they are the criteria for welfare. It is no secret that there are problems with using preference as a foundation for an account of well-being. Unfortunately, many economists ignore this insight and even its advocates fail to appreciate its full force. Satisfaction of actual preferences (the ones that determine behavior) doesn't entail well being; further, preference satisfaction isn't even a useful indicator of welfare. In this paper, I argue that we should completely abandon any normative principle that treats the satisfaction of preferences as a proxy for well being.

A situation *A* is a Pareto improvement (PI) over a situation *B* if and only if at least one person prefers *A* to *B* and no one prefers *B* to *A*.[1] PIs play a central role in normative economics. Virtually all economists agree with (some version of) the Pareto principle (PP): if *A* is a PI over *B* then *A* is better than *B* (e.g., Hausman and McPherson 2006, 136; Tungodden 2003, 2, 8; Feldstein 1999, 34; Sen 1985, 9; Rescher 1979, 170; Sen 1979, 18; Sudgen and Weale 1979, 111).[2] The PP is, for example, a key premise in standard (consequentialist) arguments for market systems (e.g., Hausman and McPherson 2006, 65-7; Gloria-Palermo and Palermo 2005; Nelson

[1] On the standard terminology, *A* is a *weak* PI over *B* since only one person must prefer *A* to *B*. *A* would be a *strong* PI over *B* if everyone prefered *A* to *B*. I frame my discussion in terms of weak PIs, but the arguments apply to strong PIs as well, *mutatis mutandis*.

[2] There are a number of fundamentally equivalent ways to put the PP: Pareto superior outcomes are better than the outcomes they Pareto dominate (Sudgen 1984, 505, 506); Pareto optimal (PO) outcomes (i.e., outcomes where no more PIs are possible) are socially rational (Rescher 1979, 169-70); Pareto inefficient (i.e., non-PO) outcomes ought not to be chosen (Sudgen 1984, 507, 508). I focus on PIs rather than PO states in this statement of the PP. I take it that "better than," "socially rational," and "ought" all express moral (or at least morally relevant) judgments. If not, it is hard to see how the PP could play the role it does in economic debates.

1994; Sen 1985). In standard welfare analysis the PP makes it possible to judge that welfare has been increased without even knowing what it is that people want: the issue is not who ended up with what but did people end up with what they wanted (e.g., Koszegi and Rabin 2008, 1821; Becker 2009). The appeal of the PP stems from the purported connections among preferences, well-being, and morality: a PI seems to make at least one person better off and no one worse off; it seems morally good to contribute to someone's well being (at least if other things are equal).[3]

The PP is not without its critics. A number of thinkers are leery of it on the grounds that moral analysis goes beyond issues of well being. They allow, for example, that *A* might be a PI over *B* and still be worse because well-being isn't the only morally relevant issue in evaluating *A* and *B* (e.g., Hausman and McPherson 2006, 65-7; Sen 1985, 10; Sudgen 1984; Sen 1979, 549-54; Sen 1976, 220-6, 235). Such thinkers usually conclude that PIs are still have some appeal: PIs are moral improvements, *other things equal*, even if particular PIs aren't, *on balance*, good, (e.g., Hausman and McPherson 2006, 138; Chang 2000, 20-2; Sen 1985, 10). I reject even this weaker claim. The problem I will focus on is that the PP doesn't get well being right. PIs are determined by what people actually prefer.[4] As certain philosophers and like-minded economists have noted, however, well being and the satisfaction of actual preferences can come apart. Sometimes, for example, people want things that are bad for them (e.g., Hausman and

[3] On the standard terminology, this leads to a *ceteris paribus* version of the *strong* PP since it appeals to *weak* PIs. The strong PP involves weak PIs; the weak PP involves strong PIs, i.e., if everyone prefers *A* to *B* then *A* is better than *B* (e.g., Gravel 2001, 164; Chang 2000, 175-80; Sen 1976, 217).

[4] Standard welfare economics gives a dual role to preferences: the preferences that determine behavior are the same preferences that determine whether a PI exists (e.g., Hausman and McPherson 2006, 120). This is no accident. If PIs were determined by actual benefits or 'cleaned up' preferences then there would no longer be any connection between market outcomes (which are determined by actual motives) and positively evaluated outcomes (which would be determined by what is good or ought to motivate) (Gloria-Palermo & Palermo 2005, 66-7). I take it that what counts as a preference, at least for economists and those who want to communicate with them, is fixed by positive economic theory. For an account of different notions of preference in economics, see Hausman 2005. My argument will hold on any of the notions Hausman discusses.

McPherson 2006, 120-8). When preference and welfare diverge, PIs aren't even *ceteris paribus* moral improvements.

Concerns about the PP haven't had much influence on economic thought. The search for PIs, for example, still guides normative analysis in technical spheres of economics (e.g., market design – see Abdulkadiroglu, et. al, *forthcoming*). This isn't simply a slow spread of ideas across economic sub-fields: recent theoretical work on welfare also endorses the PP (see, e.g., Koszegi and Rabin 2008; Beshears, et al. 2008). Even the critics of the PP bow to the widespread conviction that the PP is onto *something* when they allow that unanimous consent is a reason to morally endorse a transaction (see, e.g., Hausman and McPherson 2006, 138; Sen 1976, 235). In this paper I argue that this sort of concessive approach gives the PP too much respect. The real importance of the preference-versus-welfare critique is underappreciated. The problem is not merely that preference satisfaction imperfectly tracks welfare. The point, rather, is that there is a general distinction between what people *actually* want and what enhances their well being (and so what they *ought to* want). Recognizing this distinction is fatal to the PP – to determine whether satisfying a person's preferences really enhanced her welfare, we would need independent access to the notion of well being; with such access, however, there would be no need for appeal to preferences. The PP is simply a red herring. No pattern of actual preference satisfaction is sufficient to establish any welfare claim. Given the many ways in which people's actual preferences come apart from that they ought to prefer, it isn't even useful to track PIs for normative purposes.

1. The Pareto Principle and the Preference-versus-welfare Criticism

The argument for the PP involves two key principles.

Unambiguous Welfare Gain: The existence of a PI implies that there is welfare gain for someone without a decrease in anyone's welfare.

Minimal Benevolence: It is (morally) good to increase welfare, other things equal.

If both principles hold then PIs increase welfare, increasing welfare is good, *ceteris paribus*, and so PIs are good, *ceteris paribus* (e.g., Koszegi and Rabin 2008, 1821-3; Hausman and McPherson 2006, 64-5; Gloria-Palermo and Palermo 2005, 66-7; Cudd 1996, 1). The argument involves a dual appeal to dominance reasoning (i.e., if something is better along one [set of] dimension[s] and no worse along any other then it must be better with respect to those dimensions all together). If *A* is a PI over *B*, *A* is supposed to be better than *B* with respect to one facet of the good – welfare – because it is better for at least one person and no one is worse off. Further, if other non-welfare values are the same then *A* is better than *B simpliciter.*

The unlimited-welfare-gain claim is controversial.[5] Economists routinely just identify welfare with utility (and so ultimately with preference satisfaction) (e.g., Koszegi and Rabin 2008, 1821, 1823, 1824; Nelson 1994, 663; Sudgen 1984, 507).[6] It is unclear whether we should take this practice as simply a modeling convenience or as a reflection of a serious philosophical claim. At any rate, most non-economists are unwilling to follow suit. Clearly, "well being" is not just another term for preference satisfaction. The best evidence for this is that well being and preference satisfaction come apart, sometimes quite dramatically. The clearest illustrations involve people who sometimes act from completely implausible views of the good. Rachel, for example, lost everything she ever loved to her methamphetamine addiction. To hold that she was better off in some way for fulfilling her desire for meth doesn't make sense – preferring a

[5] Minimal benevolence is a pretty weak claim: it asserts only that well being (whatever it is) is one dimension of moral value. This is generally recognized as a reasonable view, so I won't discuss it here.

[6] There is a Libertarian argument for market exchange, but it isn't based on the value of facilitating preference satisfaction. Voluntary trades are supposed to be good because they respect freedom, whether or not they enhance welfare (e.g., Williams 1996)..

pleasant stupor to a (quite satisfactory) family life was an error for Rachel. A moment's reflection suggests that people often want things that are not good for them: bigots desire to avoid those they see as inferior; the overconfident want to avoid correction; the self-loathing want to be inappropriately punished. When someone gets what she wants in such cases it doesn't count as any sort of welfare gain because the desires satisfied are just mistaken (e.g., Chang 2000, 179ff; Sen 1976, 220-3, 225-6, 229-31, 232; Scanlon 1975, 657).

While dramatic cases where people have seemingly crazy views about welfare serve to make the conceptual point that preference satisfaction and welfare aren't the same thing, they tend to obscure the extent to which desire satisfaction and well being diverge in people's lives. While most of us avoid becoming addicted to meth, few (if any) avoid making a host of smaller mistakes.[7] Even when reasoning from their own views about welfare, people make mistakes in forming their preferences. Satisfying such mistaken preferences won't be conducive to an agent's well being *by her own lights*. In the most prosaic cases, people have false beliefs that lead them to want one thing when it would make sense (given their more basic desires) for them to want another. I might, for example, desire money and so come to want shares in CompuGlobalHyperMegaNet (CGHMN) because I believe (erroneously, it turns out) that an investment in CGHMN will make money.[8] My proximate desire for shares of CGHMN does not track my more basic desire for money and so satisfying it doesn't make me better off by my own view.[9] Erroneous beliefs of this sort give rise to PIs where at least one person will be worse off by her own lights. Someone must be selling shares of CGHMN if I am able to buy them. A transaction between us might well be a PI: I want to buy, the seller wants to sell, and no one else

[7] Even with regard to major value errors, most people will avoid any particular one but it is much less clear that most people avoid them all.
[8] Most of our desires are derived from more basic desires and beliefs in this way.
[9] Actually getting money might not be what it is cracked up to be either. It is possible for someone to end up where she ought to be by failing to get what she wants.

cares. No matter what happens to CGHMN stock, one of us will fail to achieve our ends: if the price goes down, I will regret my purchase; if the price goes up, the seller will regret the sale.[10] Both of us want to make the transaction but one of us will fail to get what we really want (e.g., Hausman and McPherson 2006, 124, 137-8; Sen 1992, 143-4). Even if we allow that a person's welfare depends on what she values, it is practically inevitable that her preferences will fail to track her well being in some ways.

False beliefs aren't the only source for mistaken preferences. Psychology tells us that people have other trouble bridging the gap between their overarching goals and the situation-specific preferences that guide their behavior: they have a hard time resisting nearer but lesser goods, overweight the influence of small probabilities, are too risk-averse for possible gains, and are too willing to gamble in order to avoid even trivial losses (for an overview, see Tetlock and Mellers 2002 and Camerer 1999). People are sometimes attracted to things in an irrational way: they find forbidden fruit more appealing, suffer from sour-grapes reasoning, or fall prey to group-think (Hausman and McPherson 2006, 128-9). People often form beliefs, desires, and preferences without attending to all of the elements of the situations they consider important (Schick 2001; Hayden and Ellis 2007, 661-675; Ellis 2008).[11] Reasoning goes awry in many ways, so on any account of the good it is practically certain that people will desire things that are not beneficial on their own view of the good (Sen 1976, 220-3, 225-6, 232).[12] These sorts of possibilities show that PIs wouldn't guarantee welfare gains even if people sought what was

---

[10] In general (e.g., special circumstances and portfolio effects aside), people prefer to buy shares when they think their value will go up and sell when they think their value will go down.

[11] I might, for example, form the intention to go for a cup of coffee with a colleague without attending to either a previously scheduled engagement or my recently diagnosed ulcer. Even important values won't influence a person's action-guiding preferences where those values aren't activated.

[12] This is why there is a standard distinction between manifest (revealed) preferences and true (normative) preferences (Beshears, et al. 2008, 1787; Chang 2000, 193; Rescher 1979, 176-7). A person is motivated to act by her manifest preferences but those preferences may not track what she ultimately wants. It also explains the appeal of laundered or amended preferences in normative analyses (Hausman and McPherson 2006, 128-9; Chang 2000, 183ff).

actually good. Given the many ways in which actual preferences can be based on mistakes, it is a poor idea to read welfare conclusions off of the mere existence of PIs.

The point of the preference-versus-welfare criticism is not merely that utility is imperfectly correlated with well being. The real lesson, rather, is that every account of the good must distinguish what *someone thinks is* good from what *is* good. A person's preferences capture what she thinks is good (or at least what she thinks is worth doing).[13] Welfare, on the other hand, is concerned with what is actually good (or at least part of the good). No one who reflects on the difference between *good* and *thought to be good* can understand satisfying even her own preferences (i.e., doing what she thinks is good) as simply equivalent to doing what is good because people are prone to mistakes – there is always a conceptual gap. Investigating well-being is simply distinct from investigating what people want. Appealing to the PP to make welfare assessments is sort of like taking a poll to find the answer to a math problem. In both cases you learn what people think but the method itself can't determine whether they have the correct answer. The conceptual distance between *good* and *thought to be good* implies that preference satisfaction isn't even a satisfactory indicator of welfare. We can't reach welfare conclusions from on preference-satisfaction evidence where a person is wrong about what is good or choiceworthy. Intuitions may diverge about exactly how likely such cases are, but we must have an independent examination of what is good for people (and so what they should want) to determine which intuitions are more accurate. If we had information about what was good for someone, of course, we wouldn't care about the status of her proximate preferences in the first place. As with the math problem analogue, there isn't much point in taking a poll once you've

[13] This shows that preference satisfaction and welfare are connected after a fashion, albeit not in a way that helps the PP. This is also why (even though they aren't interdefinable) utility is connected to choice in descriptive economics – it captures what people see as choiceworthy on the whole.

done the calculations carefully.[14] Mere preference, then, has no real role to play in normative assessment.

2. Response: Controversial Views of the Good

Despite the foregoing argument, most economists still rely on the PP and most philosophers seem to be willing to let them. Defenders of the PP often attack the preference-versus-welfare argument itself on the grounds that it depends on controversial premises. Any distinction between welfare and preference satisfaction, they argue, must presuppose a controversial view of the good (e.g., Sen 1985, 10; Rescher 1979, 177). A critic, it seems, must go outside of an agent's view of the good to argue that she desires something that is not beneficial for her: after all, she sees what she wants as worth pursuing. Arguments across conceptions of the good, however, are notoriously intractable. There is no agreement about what is good, so criticisms of preferences are inevitably tendentious (e.g., Rescher 1979, 175-6). This view is usually amplified by the claim that it is not an economist's job to work on moral truths anyway (e.g., Hausman and McPherson 2006, 119-20; Atkinson 2001, 194-5). Given disagreements about morality, it seems best to leave welfare (and other facets of the good) in the eye of the beholder (e.g., Hausman and McPherson 2006, 120; Gravel 2001, 164; Sudgen 1984, 507).[15]

This controversial-views-of-the-good response to the preference-versus-welfare criticism of the PP involves some questionable burden shifting. The fact that people disagree does not show that some view about the good is not the right one. At a minimum, arguments seem to rule out

[14] This is not to say that there is no reason to have people check your reasoning, especially if it is complex. The point, rather, is that the reasoning is the focus, not simply the end result.

[15] There is a more positive defense of the PP that involves a contractarian argument. Contractarians hold that rules are justified if they can command unanimous consent. Everyone can agree to a PI – some (those who prefer it) will advocate for it, no one will block it (since no one prefers the alternative) (Cudd 1996, 7; Sudgen 1984, 507). Everyone would agree to PP (and a bit more) behind veil of ignorance (Sudgen and Weale 1979, 113). In this sense, at least, rationality endorses PIs – they are what rational people can achieve.

inadequate conceptions of the good.[16] Figuring out ethics may not a job for economists, but this doesn't imply that economists can ignore philosophical insights.[17] It isn't obvious, at any rate, giving people what they want enhances their welfare. A number of popular views of the good (some of them quite defensible) imply that people often want the wrong things. Many religious views of the good, for instance, hold that people do not generally want good things. Buddhists have a problem with desire in general – as the Second Noble Truth has it, suffering is caused by attachment (e.g., Kupperman 2007, 23-42). Many Christians think that Original Sin leads to depraved desires – people pursue what is actually bad for themselves (e.g., Rea 2007, 323).[18] Deontological views of ethics (e.g., Kantian views) rule out certain desires as inappropriate (Kant 1997, 4: 424-428).[19] Even views that tie the good closely to human nature generally hold that it is difficult to determine what actions and attitudes are appropriate under various circumstances (e.g., Aristotle 1985, 1113a15-1113b2; Mill 1906, 14-15). The unlimited-welfare-gain principle is inconsistent with these common views of the good. Economists can defend their usual approach to welfare economics, then, only if they are willing to enter the debate about the good. They remain unwilling, however, to even address such philosophical issues (e.g., Hausman and McPherson 2006, 67, 119-20; Rescher 1979, 176).

Burden-of-proof issues aside, the controversial-views-of-the-good response is misdirected. As I draw it above, the distinction between preference satisfaction and welfare doesn't depend on any particular view of the good. The criticism isn't that there is one true view of the good such

[16] Even if we can't isolate the best possible sort of life for someone, it doesn't follow that we can't rule out lives that involve meth addiction.
[17] Arguably, economists have already committed to a philosophical position when by holding the unlimited-welfare-gain principle.
[18] A less radical take holds that Original Sin has merely (?) disordered human desires, making right desires impossible to achieve without divine aid (e.g., Rea 2007, 323-4).
[19] According to Kant, "philosophy is to manifest its purity as sustainer of its own laws …: that they expect nothing from the inclination of human beings but everything from the supremacy of the law and the respect owed it, or failing this, condemn the human being to contempt for himself and inner abhorrence" (1997, 4:425-426)

that people want things that are not in fact good. Rather, it holds that for any (remotely plausible) conception of the good, even someone who holds it will prefer some things that are inconsistent with that conception [20] There are a number of reasons why a person's preferences might come apart not only from what is objectively valuable (if there is such a thing) but also from her own deepest subjective values. As we saw before, perfectly normal reasoning can lead people to actually want things they wouldn't want if they knew the facts or were better reasoners. People have false beliefs, sometimes due to poor information and sometimes due to poor (e.g., non-Bayesian) information processing (Koszegi and Rabin 2008, 1827-8). Add in psychological biases and limited capacities and it is practically guaranteed that a person's actual preferences won't fully track even her own conception of well being.[21] The preference-versus-welfare criticism of the PP doesn't depend on any particular view of the good, much less a controversial one.

3. Response: Mitigation

Another response to the preference-versus-welfare criticism of the PP holds that the unlimited-welfare-gain principle is approximately true: the existence of a PI is *prima facie* evidence that someone has experienced a welfare gain without anyone's welfare being decreased. While the connections among choices, preferences, and welfare aren't certain, they are close

[20] To some, this might evoke the debate between objective and subjective views of the good. My argument, however, doesn't hinge on that discussion. It probably would be easier to make the case that people sometimes want harmful things if "there [were] things that are good in themselves for an individual independently of her desires and attitudes toward them" (Arneson 2000, 37). Still, it is possible to reason poorly about the good and so have misguided preferences even if "the things that are intrinsically good for an agent … acquire this status only in virtue of how she happens to regard them" (Arneson 2000, 37).

[21] The contractarian case for the PP is undermined by the same point. Contractarianism assumes that what people will agree to what they should. Even contractual views allow that people can misunderstand or misforecast value assessments (Cudd 1996, 26; Sudgen & Weale 179, 119). The normative force of a social contract depends on avoiding such errors. Even contractarians, then, must distinguish between deals that people would actually make and deals they should make (by way of constraints such as the original position) (Sudgen & Weale 1979, 111, 113, 117). Absent such a distinction, a contract is, at best, a *modus vivendi*. Political sustainability is important, of course, but it is a different issue than morality (Atkinson 2001, 197, 199).

enough to validate most economic claims. Botond Koszegi's and Matthew Rabin's recent paper "Choices, situations, and happiness" (2008) is a prime example of this approach. Koszegi and Rabin note that on the traditional economic approach, "observed behavior is assumed to reflect fully rational maximization of utility, and … welfare is higher in one situation than another if it lets a person attain the outcome she seems most inclined to choose" (2008, 1821). This account, they acknowledge, has conceptual deficiencies. Koszegi and Rabin allow, for instance, that people make mistakes in the pursuit of their ends, e.g., they misunderstand the stock market, commit the gambler's fallacy, or make other cognitive errors (2008, 1827-8).[22] They emphasize a methodological lesson here: choices aren't always good evidence for a person's values because her actions might express her mistakes rather than her desires (Koszegi and Rabin 2008, 1828, 1829).[23] Despite its problems, Koszegi and Rabin think the standard economic approach to welfare is on the right track: "[w]hen doing so with sensible ancillary assumptions, inferring people's well being based on the presumption that observed choices are rational is in our view the best scientific program for studying well being yet formulated" (2008, 1821).[24] In particular, they stand by the substantive conclusions of standard welfare economics: "[d]espite conceptual problems …, in many cases it seems clear that both rationality and choice-set independence of

[22] Another part of their view is that preferences are more complicated than is usually assumed. Someone might prefer to have help quitting smoking, other things equal, but prefer to pass up an opportunity to get aid because she hates to ask for it (Koszegi and Rabin 2008, 1823-4; see also 1821-2). Preferences of this sort violate what Koszegi and Rabin call "Menu Independence of Welfare" (1824). Koszegi and Rabin are primarily interested in drawing a methodological lesson here: choice behavior alone cannot isolate complex preferences of this sort (2008, 1823-7). Different sets of preferences lead to the same sorts of choices (e.g., people continue to smoke for different reasons) so no set of behavioral observations can, by itself, tease out which preferences are operative. A pattern of choices can't, for example, exclude the possibility that someone would get more utility from a painful, unavoidable death than anything she actually chooses over death – she chooses only with respect to avoidable deaths (Koszegi and Rabin 2008, 1823). Common-sense psychology suggests, of course, that people don't want to suffer a painful, unavoidable death and this is good enough to exclude a preference for such outcomes. In order to fix preferences, then, choice evidence must be supplemented with some ancillary assumptions. Koszegi and Rabin admit that economics always relies on such choice-unobservable principles (2008, 1821, 1823).They see a role for psychology in uncovering new, perhaps less common-sense, assumptions of this sort (2008, 1827, 1830-1).

[23] Investigating mistakes provides another potential role for psychology within the general economic framework.

[24] While Koszegi and Rabin "make the case for supplementing and combining" the traditional approach with psychological research (1821), they are primarily interested in patching conceptual cracks.

preferences are good enough approximations that in fact familiar approaches are quite sufficient. … [R]evealed preference is too powerful a tool for studying well being, and the ancillary assumptions needed to render the tool effective are often too minimal and reasonable, to fret much about the conclusions economists are reaching except in cases where there are specific reasons to doubt these assumptions" (2008, 1823).

Even critics of the PP accept something like the foregoing mitigation line. Hausman and McPherson, for example, make the case that "[t]here are problems with endorsing all Pareto improvements (as the Pareto principle does)" (2006, 137). Still, they hold that "[t]he Pareto principle has some real ethical appeal because satisfying preferences surely has something to do with promoting well being" (2006, 138).[25] Sen, likewise, emphasizes "the unacceptability of the Pareto principle as a universal rule" (1976, 235). Nonetheless, he holds that "there is something very central in the idea that preferences unanimously held by members of a community cannot be rejected by that community. As Blau … puts it, 'I can see no case for an outside observer denying a unanimous choice'" (Sen 1976, 235-6).[26]

Standard welfare economics treats choices (or at least the instrumentally rational ones) as welfare maximizing (Koszegi and Rabin 2008, 1821). The mitigation response to criticism of the PP allows that choice and welfare can come apart. It is important to recognize, however, that the connection between choice and welfare involves two steps: the standard view equates choice with utility maximization and utility maximization with welfare maximization. While mitigationists such as Koszegi and Rabin reject "the debilitating tautology that everything people

[25] Again, there *is* a connection between well being and preferences – the latter reflects an agent's judgments about the former. This doesn't make the PP plausible, however, because peoples' judgments can be wrong.

[26] Sen rejects the PP in favor of "a *conditional* version …[:] If everyone in a community prefers *x* to *y* and wants that preference to count, then *x* must be socially preferred to *y* (conditional weak Pareto improvement …)" (1976, 236). Sen also discusses a conditional version of the strong PP (1976, 243). Conditionalizing the PP in this way doesn't help with the objection I am pressing. Sen wants to salvage the intuition that no outside observers can question a unanimous choice of the right sort; this is exactly the intuition I argue is unsupported.

do maximizes their utility" (1822), they still accept the even more problematic view that well being is a matter of utility maximization. Koszegi and Rabin, for example, only *argue* for the claim that behavior, supplemented by psychological assumptions, reveals preferences. The view that "rationality and choice set independence of preferences are good enough approximations" only supports the claim that we can read preferences off behavior, not the claim that looking at preferences "is a powerful tool for studying well-being" (Koszegi and Rabin 2008, 1823). They just take it for granted that there is a close connection between utility and welfare. Much the same point holds for Hausman and McPherson, as well as for Sen. While they recognize that they are concerned specifically with preferences on the one hand and welfare on the other, they simply appeal to their intuitions that there must be *something* that connects preference satisfaction and well being.

The existence of the non-desire-oriented views of the good canvassed above should be enough to cast at least some doubt on any intuition linking welfare and preference satisfaction. If the Second Noble Truth (or a Christian account of Original Sin or Kant's Categorical Imperative, etc.) were correct then the "ancillary assumptions" to which Koszegi and Rabin appeal, "minimal and reasonable" though they might be, wouldn't enough to vouchsafe "the conclusions economists are reaching" about the welfare properties of PIs (2008, 1823). The intuition that preference satisfaction has something to with welfare must be sensitive to the debate about the good. Mere appeal to the existence of PIs can't answer any normative questions.

More importantly, any intuition that actual, situation-specific preferences (as opposed to broader value commitments) at least approximate the good must yield when we recall the way in which welfare and utility come apart. The human capacity for error ensures that people will, with some regularity, prefer things that are worse for them by their own lights. Poor reasoning (in its many forms) will lead everyone to have proximate desires that don't make sense given their more fundamental values. Whatever conception of the good someone might hold, she can (and, as a practical matter, will) want something inconsistent with that conception. In order to separate mistaken from value-congruent preferences, we need to be explicit about the value standard at issue and use it to evaluate preferences. This, however, is exactly what standard normative economics tries to avoid – the PP is supposed to allow us to determine when we have a welfare gain without any need to look at the value judgments that preferences are based on. Once the value standards are required, there isn't much point in consulting preference satisfaction – we can look directly at whether behavior advances the relevant values. Again, taking a poll might provide correct answers to a math problem, but maybe not. The only way to tell is to either know the answer independently or to closely follow and evaluate the reasoning process of those polled. Once you have done or are doing the math, however, the existence of the poll isn't helping you find the answer. Likewise with the PP: the mere existence of the PI doesn't tell us about welfare; we need to look at the appropriate values for that.

Conclusion

The PP is false because the unlimited-welfare-gain claim is false: the mere existence of a PI neither implies nor provides good evidence that anyone has made a welfare gain. The problem isn't merely that PIs do not guarantee well being on many views of the good. Rather, no view of the good allows someone to draw conclusions about welfare (or any other facet of the good) from the existence of a PI (or any other pattern of preference satisfaction). The PP presupposes both a subjectivist account of welfare and a subjectivist account of what counts as progress toward welfare (i.e., it holds that what someone thinks will increase her welfare will count as a welfare increase). The first supposition is dubious, but even if it weren't, the second would still be completely untenable. However welfare is defined, there is a fact of the matter about how someone in a particular circumstance could best achieve it. A person's actual preferences, however, won't always track that optimal path. As a normative claim, then, the PP is worthless because it looks at the wrong thing.[27] Economists are therefore mistaken when they assert that "economic research and teaching does *now* (via our various notions of efficiency and welfare) reach strong conclusions about well-being" (Koszegi and Rabin 2008, 1831).

What, then, is the appropriate role for economics in the study of welfare? Economists, as such, have no particular insight about what makes people better (or worse) off. Even without its own characterization of well being, however, economic analysis can still help us understand features of the world that have been independently identified as relevant to well-being (Hausman and McPherson 2006, 129-33). Economics, for example, is the primary tool we have for studying the distribution of food, shelter, security, comfort items, etc. in a given society. While the connection is not straightforward, the distribution of such goods is (quite plausibly) a crucial determinant of human welfare. Economics has an important role to play in the study of well being but it is no part of that role to determine what counts as welfare.

[27] There are some obvious practical benefits to unanimous consent, especially in the political realm. It might even be that agreement on a course of action has some evidentiary value – the wisdom of crowds and all that. I am willing to allow, then, that trying to achieve PIs might generally be a wise policy, despite the fact that the PP isn't true.

WORKS CITED

Abdulkadiroglu, A., Pathak, P., Roth, A., *forthcoming*. Strategy-proofness versus efficiency in matching with indifferences: redesigning the NYC high school match. American Economic Review.

Aristotle 1985. Nichomachean Ethics. Ed. T. Irwin. Hackett, Indianapolis.

Arneson, R. 2000. Perfectionism and politics. Ethics 111, 37-63.

Atkinson, A., 2001. The strange disappearance of welfare economics. Kyklos 54, 193-206.

Becker, G. 2009. The serious conflict in the modern conservative movement. The Becker-Posner Blog. May 10. <www.becker-posner-blog.com/archives/2009/05/the_serious_con.html>

Beshears, J., Choi, J., Laibison, D., Madrian, B., 2008. How are preferences revealed? Journal of Public Economics 92, 1787-1794.

Camerer, C., 1999. Behavioral economics: reunifying psychology and economics. Proceedings of the National Academy of Sciences USA 96, 10575-7.

Chang, H., 2000. A liberal theory of social welfare: fairness, utility, and the Pareto principle. Yale Law Journal 110, 173-235

Cudd, A., 1996. Is Pareto optimality a criterion of justice? Social Theory and Practice 22, 1-34.

Ellis, S. 2008. Market hegemony and economic theory. Philosophy of the Social Sciences 38, 513-32.

Feldstein, M. 1999. Reducing poverty, not inequality. Public Interest 139, 22-41.

Gloria-Palermo, S., Palermo, G., 2005. Austrian economics and value judgments: a critical comparison with neoclassical economics. Review of Political Economy 17, 63-78.

Gravel, N., 2001. On the difficulty of combining actual and potential criteria for an increase in social welfare. Economic Theory 17, 163-80.

Hausman, D., McPherson, M., 2006. Economic Analysis, Moral Philosophy, and Public Policy, 2nd ed. Cambridge UP, New York.

Hausman, D. 2005. Sympathy, commitment, and preference. Economics and Philosophy 21:33-50.

Hayden, G., Ellis, S., 2007. Law and economics after behavioral economics. University of Kansas Law Review 55, 629-675.

Kant, I., 1997. Groundwork of the Metaphysics of Morals. Ed. Gregor, M. Cambridge UP, New York.

Koszegi, B., Rabin, M., 2008. Choices, situations, and happiness. Journal of Public Economics 92, 1821-1832.

Kupperman, J., 2007.Classic Asian Philosophy, 2nd ed. Oxford UP, New York.

Mill, J.S., 1906. Utilitarianism, 13th ed. Chicago UP, Chicago.

Nelson, J., 1994. Business ethics in a competitive market. Journal of Business Ethics 13, 663-6.

Rea, M., The metaphysics of original sin. In: Van Inwagen, P., Zimmerman, D. (Eds.), Persons Human and Divine. Oxford, New York.

Rescher, N., 1979. Economics vs. moral philosophy: the Pareto principle as a case study of their divergent orientation. Theory and Decision 10, 169-79.

Scanlon, T., 1975. Preferences and urgency. Journal of Philosophy 72, 655-69.

Schick, F., 1991. Understanding Action. Cambridge UP, New York.

Sen, A., 1976. Liberty, unanimity and rights. Economica 43, 217-45.

Sen, A., 1979. Personal utilities and public judgments: or what's wrong with welfare economics. Economic Journal 89, 537-58.

Sen, A., 1985. The moral standing of the market. In: Paul, E., Miller, F., Paul, J. (Eds.), Ethics and Economics. Blackwell, Oxford.

Sen, A., 1992. Minimal liberty. Economica 59, 139-59.

Sudgen, R., 1984. Is fairness good? A critique of Varian's theory of fairness. Noûs 18, 505-11.

Sudgen, R., Weale, A., 1979. A contractual reformulation of certain aspects of welfare economics. Economica 46, 111-23.

Tetlock, P., Mellers, B., 2002. The great rationality debate. Psychological Science 13, 94-9.

Tungodden, B., 2003. The value of equality. Economics and Philosophy 19, 1-44.

Williams, W., 1996. The argument for free markets: morality vs. efficiency. Cato Journal 15, 179-189.