DOI: 10.1017/psa.2025.6

This is a manuscript accepted for publication in *Philosophy of Science*.

This version may be subject to change during the production process.

A Defence of Informed Preference Satisfaction Theories of Welfare

Dr. Roberto Fumagalli

Senior Lecturer, King's College London, UK**,** Research Associate, London School of Economics, UK, Visiting Scholar, University of Pennsylvania, US

Email: roberto.fumagalli@kcl.ac.uk; R.Fumagalli@lse.ac.uk

Abstract

This article defends informed preference satisfaction theories of welfare against the most influential objections put forward in the economic and philosophy of science literatures. The article explicates and addresses in turn: the objection from inner rational agents; the objection from unfeasible preference reconstruction; the objection from dubious normative commitments; the objection from conceptual ambiguity; and the objection from conceptual replacement. My defence does not exclude that preference satisfaction theories of welfare face significant conceptual and practical challenges. Still, if correct, it demonstrates that philosophers/welfare economists are justified in relying on specific versions of such theories, namely informed preference satisfaction theories.

Keywords: Preferences; Choices; Welfare; Rationality; Policy Evaluation.

1. Introduction

Theories of welfare are commonly classified into *mental state* theories of welfare, which hold that individuals are well-off to the extent that they experience specific kinds of mental states (e.g., Clark et al. 2018, on happiness; Feldman 1997, on pleasure), *preference satisfaction* theories of welfare (henceforth, PSTW), which hold that individuals are well-off to the extent that their own preferences are satisfied (e.g., Bernheim 2009, Ferreira 2023), and *objective list* theories of welfare, which hold that individuals are well-off to the extent that they have certain goods/experiences (e.g., health, education, friendship) irrespective of whether they experience specific kinds of

mental states or satisfy their preferences (e.g., Fletcher 2013, Nussbaum 2000).[1] Philosophers/welfare economists frequently assume that individuals' preferences can be reliably inferred from their choices and often rely on PSTW (e.g., Adler and Fleurbaey 2016, Angner 2016, Sobel 1998, Sumner 1996, ch.5). In recent years, however, several prominent authors have argued that PSTW fail to provide a plausible theory of welfare. In this article, I develop and support a qualified defence of PSTW against the most influential objections put forward in the economic and philosophy of science literatures. My main claim is that although PSTW face significant conceptual and practical challenges, specific versions of PSTW - namely, informed PSTW - are more plausible than their critics maintain, and philosophers/welfare economists are justified in relying on such versions of PSTW.[2]

The article is structured as follows. Section 2 outlines the main tenets of PSTW and distinguishes between the three main proffered versions of PSTW, namely actual PSTW, informed PSTW and ideal PSTW. Sections 3-7 defend informed PSTW against five prominent objections. More specifically, I explicate and address in turn: (1) the objection from *inner rational agents* (e.g., Infante et al. 2016a, Sugden 2018, ch.4-5); (2) the objection from *unfeasible preference reconstruction* (e.g., Infante et al. 2016b, Rizzo and Whitman 2020, ch.6-7); (3) the objection from *dubious normative commitments* (e.g., Hausman and McPherson 2009, Kraut 2007, part II); (4) the objection from *conceptual ambiguity* (e.g., Hausman 2024, Lecouteux 2022); and (5) the objection from *conceptual replacement* (e.g., Levy and Glimcher 2012, Thoma 2021a).

[1] I use the terms 'welfare' and 'well-being' interchangeably to indicate prudential value, i.e. what is non-instrumentally good for individuals (e.g., Griffin 1986, ch.1-3, Sumner 1996, 20-25). Also, I speak of 'theories' of welfare to refer to explanatory theories of welfare (rather than enumerative theories of welfare), i.e. I take such theories to specify both which goods/experiences are non-instrumentally good for individuals and in virtue of what properties or features these goods/experiences are non-instrumentally good for individuals (e.g., Crisp 2006, ch.4, Woodard 2013).

[2] I speak of 'welfare economists' broadly so as to include policy advisors and policy makers involved in normative welfare evaluations. In the philosophical literature, various authors contrast mental state theories and objective list theories with desire fulfilment theories rather than preference satisfaction theories (e.g., Heathwood 2016, Parfit 1984, 493-502). I focus on preference satisfaction theories for the purpose of this article. For further discussion concerning the relation between desire fulfilment theories and preference satisfaction theories, e.g., Griffin 1986, ch.1-3, Sobel 2009. For further discussion concerning the tripartite classification of theories of welfare highlighted in the main text, e.g., Adler 2012, 159-170, Scanlon 1998, ch.3.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

Before proceeding, the following three preliminary remarks are worth making. First, the objections I address target the main tenets of informed PSTW, which concern the *existence* of informed preferences, the possibility of *reconstructing* these preferences and the *normative/evaluative significance* of such preferences. These objections do not exhaust the set of critical issues faced by informed PSTW (e.g., Sugden 2008, Whitman and Rizzo 2015, holding that adopting informed PSTW leads philosophers/welfare economists to endorse unjustified paternalistic interventions; Fumagalli 2024, for a reply). Still, as I illustrate below, those objections target the main respects in which the plausibility of informed PSTW has been called into question in the recent economic and philosophy of science literatures. I take such objections to be especially interesting to philosophers/welfare economists since they encompass the major bones of contention between the proponents and the critics of informed PSTW and highlight the most pressing challenges faced by PSTW more generally.

Second, PSTW are not *the only* approach that aims to ground reliable and informative welfare evaluations on information concerning individuals' preferences. Still, PSTW differ from several other preference-based approaches to welfare evaluations in that PSTW take preference satisfaction to *constitute* welfare rather than merely provide *evidence* for welfare (e.g., Hausman and McPherson 2009, 1-2, on the evidential account of welfare, which holds that "if individuals seek to benefit themselves and are good judges of what is good for them, then […] their preferences will be reliable indicators of what is good for them"). I do not aim in this article to assess the comparative strengths and limitations of different preference-based approaches to welfare evaluations. However, I shall expand on preference-based approaches other than informed PSTW when consideration of such approaches directly bears on my defence of informed PSTW (e.g., Section 5 on the evidential account of welfare).[3]

[3] Some welfare economists hold that they do not posit any substantive relation between preference satisfaction and welfare, and claim to regard welfare merely as a technical term representing individuals' preference rankings (e.g., Mas-Colell et al. 1995, ch.16 and 21). I mention this view in passing for the purpose of my defence of informed PSTW.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

And third, my defence of informed PSTW primarily focuses on *individual* welfare evaluations rather than *social* welfare evaluations since social welfare evaluations raise additional complications that are orthogonal to the merits of informed PSTW (e.g., Adler 2012, ch.3, Fleurbaey 2012, on various epistemic challenges faced by attempts to ground interpersonal comparisons of welfare; Adler 2019, ch.3-4, Fleurbaey and Maniquet 2011, ch.2-4, on various normative challenges faced by attempts to determine what weights should be ascribed to different individuals' welfare in interpersonal aggregations of welfare). Still, I shall comment in various places on the applicability of informed PSTW to social welfare evaluations (e.g., Sections 3-7 on the evaluation of policies' welfare implications).[4]

2. Preference Satisfaction Theories of Welfare

According to PSTW, the satisfaction of individuals' preferences *constitutes* individuals' welfare, i.e. makes individuals better off than they would be in otherwise identical situations where their preferences are not satisfied (e.g., Fumagalli 2021, Hausman 2012, ch.7-8). In particular, an individual's preferences for some state of affairs count as *satisfied* if such state of affairs occurs (e.g., Hausman 2010, 326; also Griffin 1986, ch.1, Sumner 1996, ch.1). Individuals may derive feelings of pleasure or satisfaction from knowing that their preferences are satisfied. Still, on PSTW, preference satisfaction does not have to involve any feelings of pleasure or satisfaction (e.g., Hausman and McPherson 2009, 13, Kraut 2007, 98-99) and may constitute welfare irrespective of whether it involves such feelings (e.g., Hausman and McPherson 2009, 10, holding that "there is only [a] contingent connection between the satisfaction of a

[4] My defence of informed PSTW does not purport to demonstrate that informed PSTW are the only plausible theory of welfare. In particular, it allows that what theories of welfare are most aptly adopted in specific contexts may depend on theoretical and pragmatic factors besides these theories' plausibility (e.g., Angner 2011, Fumagalli 2022, Van der Deijl 2017, on measurability considerations). I do not expand on the relative importance of these factors since my defence of informed PSTW does not directly rest on what view one advocates about the relative importance of such factors.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

preference and the satisfaction [felt by] a person"; also Arneson 1999, 123, Rabinowicz and Osterberg 1996, 2).[5]

Three main versions of PSTW have been articulated in the specialized literature. *Actual* PSTW take individuals' welfare to be constituted by the satisfaction of their actual preferences, i.e. the preferences individuals happen to have in the examined choice settings (e.g., Gul and Pesendorfer 2008, 24). For their part, *informed* PSTW take individuals' welfare to be constituted by the satisfaction of their informed preferences, i.e. the preferences individuals are able to form on the basis of accurate information and considerate judgments concerning their choice options/circumstances (e.g., Griffin 1986, ch.1-2). Still differently, *ideal* PSTW take individuals' welfare to be constituted by the satisfaction of their ideal preferences, i.e. the preferences individuals would counterfactually have "if they had complete information [concerning their choice options/circumstances and] unlimited cognitive abilities" (Sunstein and Thaler 2003, 1162; also Harsanyi 1982, 55, on the preferences individuals would have if they "had all the relevant factual information [and] were in a state of mind most conducive to rational choice").[6]

The distinction between actual PSTW, informed PSTW and ideal PSTW categorizes the proffered versions of PSTW into three exclusive sets. These versions of PSTW require preferences to meet dissimilar conditions to qualify as

[5] This does not exclude the possibility that preference satisfaction may constitute welfare through the contingent link between the satisfaction of an individual's preferences and the sense of satisfaction that the individual may derive from knowing that such preferences are satisfied (e.g., Rizzo 2025, 10, holding that "the notion of satisfaction [presupposed by PSTW] does not imply that there is in fact no associated psychological state. […] It just means that [PSTW] are silent about it"). I expand on this possibility in Section 5.

[6] According to informed PSTW and ideal PSTW, what is good for one is not "what she would [prefer] for herself were she idealized" - as posited by various so-called 'full information' accounts of welfare - but rather "what, were she idealized, she would [prefer] for her actual, unidealized self" - as posited by various so-called 'ideal advisor' accounts of welfare (Heathwood 2016, 140; also Railton 1986, 16-17). Ideal advisor accounts' focus on the idealized agent's preferences for her actual, unidealized self is not without critics (e.g., Loeb 1995, 19-20). However, leading critics of ideal advisor accounts concur that such focus "neatly eschews the implausible identification of interests between our informed and our ordinary self" (Sobel 1994, 793) and is "a step in the right direction" (Sobel 2001, 229; also Rosati 1995).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

actual preferences, informed preferences and ideal preferences, respectively. I shall expand on these dissimilarities in Sections 3-7. For now, I note that - contrary to actual PSTW - both informed PSTW and ideal PSTW impose at least two conditions on individuals' preferences, namely *information* (epistemic) conditions, which concern the extent to which individuals' preferences are grounded on accurate information concerning individuals' choice options/circumstances, and *consistency* (structural rationality) conditions, which concern the extent to which individuals' preferences fit specific consistency requirements (e.g., transitivity). The idea is that only some of individuals' preferences are such that their satisfaction constitutes individuals' welfare and that only preferences which satisfy specific information conditions (e.g., accurate information about the available choice options) and specific consistency conditions (e.g., transitivity) are plausibly taken to belong to such set of welfare-relevant preferences (e.g., Griffin 1986, ch.1-2, Fumagalli 2025).[7]

More specifically, ideal PSTW impose rather demanding information and consistency conditions on preferences, in that they hold that only fully informed and consistent preferences are such that their satisfaction constitutes welfare (e.g., Harsanyi 1982, 55, Sunstein and Thaler 2003, 1162). For their part, informed PSTW impose less demanding information and consistency conditions on preferences, in that they allow that the satisfaction of incompletely informed and partly inconsistent preferences may constitute welfare (e.g., Bernheim 2021, 390-2, Fumagalli 2024, 89-91). Distinct versions of informed PSTW impose different information and consistency conditions on preferences (e.g., Griffin 1986, ch.1). These differences by no means justify taking any information and consistency conditions to reliably track individuals' welfare (e.g., Sections 3-4 on the inadequacy of various information and consistency conditions). Still, those differences are not inherently problematic for the proponents of informed PSTW.

[7] Informed PSTW and ideal PSTW impose both information conditions and consistency conditions on preferences since information conditions or consistency conditions alone do not "ensure that we prefer the option that is actually better for us" (Sobel 1994, 787; also Fumagalli 2025). Some authors speak of coherence (rather than consistency) conditions (e.g., Broome 2013, ch.7-8, Dorsey 2017, 203-6, Grill 2015, 708-9). I focus on consistency (rather than coherence) conditions since most versions of informed PSTW and ideal PSTW impose consistency (rather than coherence) conditions on preferences (e.g., Fumagalli 2019, Rizzo and Whitman 2020, ch.3).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

For both information and consistency are plausibly taken to admit of degrees, and the specification of the information and consistency conditions presupposed by informed PSTW may justifiably vary across choice settings (e.g., how much information individuals must possess for their preferences to qualify as informed may justifiably vary depending on what individuals are involved and what choices they face). I shall explicate such differences in Sections 3-7.[8]

According to the critics of PSTW, neither actual PSTW nor informed/ideal PSTW withstand scrutiny. The critics' case against actual PSTW can be explicated as follows. Let us call those actual preferences whose satisfaction is plausibly taken to constitute welfare (if any) *actual preferences**. Philosophers/welfare economists can identify actual preferences* *only if* individuals' actual preferences meet stringent information conditions (e.g., accurate information about the available choice options) and consistency conditions (e.g., transitivity). However, individuals' actual preferences *frequently fail* to meet these conditions (e.g., Sugden 1991, on violations of transitivity; Hausman 2011, on cases where individuals' actual preferences rest on inaccurate information about the available choice options). Moreover, individuals' actual preferences often track factors that appear to be prudentially *irrelevant* (e.g., Camerer and Loewenstein 2004, on cases where individuals' actual preferences depend on frames) or even *hamper* what most theories of welfare regard as individuals' welfare (e.g., Hausman and McPherson 2009, on cases where individuals prefer options that hamper their welfare because they mistakenly believe that such options enhance their welfare; Stoljar 2014, on cases where individuals prefer options that hamper their welfare as a result of adaptation). For these reasons, the critics of actual PSTW go, the

[8] Some authors propose to exclude several preferences from the set of informed/ideal preferences because of moral (besides epistemic and structural rationality) considerations (e.g., Harsanyi 1982, 56, calling to "exclude all clearly antisocial preferences, such as sadism, envy, resentment, and malice"). Others resist this proposal on the alleged ground that prudential value is conceptually distinct from moral value (e.g., Rosati 2006, 35, Sumner 1996, 20-25; also Bernheim 2016, 18, holding that "economists have no special expertise [concerning] moral considerations"). I do not expand on this issue since the proponents of informed PSTW may consistently advocate dissimilar positions about such issue (e.g., Griffin 1986, ch.2, Kagan 1992, Sobel 1998, Vromen 2022, for discussion).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

satisfaction of individuals' actual preferences cannot be plausibly taken to constitute individuals' welfare.[9]

The critics' case against informed/ideal PSTW can be explicated as follows.[10] Let us call those informed/ideal preferences whose satisfaction is plausibly taken to constitute welfare (if any) *informed/ideal preferences**. Informed/ideal PSTW can accommodate the fact that individuals' actual preferences are often inconsistent, ill-informed, and track factors that appear to be prudentially irrelevant/detrimental. For as leading critics of PSTW acknowledge, one may plausibly ascribe to individuals informed/ideal preferences* in several cases where individuals' actual preferences are inconsistent, ill-informed, and track factors that appear to be prudentially irrelevant/detrimental (e.g., Hausman and McPherson 2009, 11, holding that informed/ideal PSTW "resolve […] most of the difficulties facing the actual preference-satisfaction view" since a person's actual preferences often fail to be 'informed' "and so satisfying [such preferences] would not make the person better off"; also Loeb 1995, 1, Rosati 2006, 63, Sumner 1996, ch.5). However, it is difficult to determine *what notion* of informed/ideal preferences should ground welfare evaluations unless one makes substantive normative/evaluative assumptions (e.g., McQuillin and Sugden 2012, 560, claiming that the concepts of 'complete information' and 'unlimited cognition' figuring in ideal PSTW are "inescapably normative"). Moreover,

[9] Actual PSTW may be defended against some of the criticisms outlined in the main text. For instance, various alleged violations of the consistency conditions putatively required to identify individuals' actual preferences* may be accommodated by precisifying the description of the choice options faced by individuals (e.g., Broome 1993; also Dietrich and List 2016a, Fumagalli 2020, for recent discussion). Moreover, the proponents of actual PSTW are not committed to taking the satisfaction of any ill-informed actual preferences to enhance individuals' overall welfare (e.g., Heathwood 2005, 491-2, on cases where satisfying such preferences frustrates other and weightier actual preferences). I mention these defences of actual PSTW in passing since most philosophers/welfare economists concur that the criticisms outlined in the main text, taken together, cast serious doubt on actual PSTW (e.g., Hausman and McPherson 2009, 11, Hawkins 2019, 106-7, Sumner 1996, ch.5).

[10] The criticisms of PSTW outlined in the main text group informed PSTW and ideal PSTW together since the critics of PSTW frequently group informed PSTW and ideal PSTW together in arguing against PSTW (e.g., Fumagalli 2024). In the following sections, I focus on informed (rather than ideal) PSTW because I take informed PSTW to be more plausible than ideal PSTW (e.g., footnote no.11 on several criticisms put forward specifically against ideal, rather than informed, PSTW). In doing so, I retain references to ideal PSTW as helpful signposts to an extreme (and untenable) version of PSTW.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

different approaches have been developed to reconstruct individuals' informed/ideal preferences*, and different approaches classify different subsets of preferences as informed/ideal preferences* (e.g., Dold 2018, Whitman and Rizzo 2015, for illustrations). In fact, reconstructing some individuals' informed/ideal preferences does not *per se* enable philosophers/welfare economists to reliably assess these individuals' welfare (e.g., Sugden 2018, ch.4, on putative cases where individuals' choices reveal context-dependent informed/ideal preferences). For these reasons, the critics of informed/ideal PSTW go, the satisfaction of individuals' informed/ideal preferences cannot be plausibly taken to constitute individuals' welfare.[11]

In the following sections, I argue that the proffered criticisms of PSTW cast doubt on both actual PSTW and ideal PSTW, but fail to undermine informed PSTW. In particular, I shall defend informed PSTW against five prominent objections, namely: the objection from *inner rational agents* (Section 3); the objection from *unfeasible preference reconstruction* (Section 4); the objection from *dubious normative commitments* (Section 5); the objection from *conceptual ambiguity* (Section 6); and the objection from *conceptual replacement* (Section 7).[12]

3. Objection from Inner Rational Agents

The *objection from inner rational agents* holds that informed PSTW do not withstand scrutiny because individuals cannot be plausibly taken to have well-informed and consistent informed preferences*. The objection proceeds as

[11] Additional criticisms have been put forward specifically against ideal (rather than informed) PSTW (e.g., Loeb 1995, 15, holding that "a subject's [fully informed] counterpart would be so different from that subject that it is hard to see how his motivations - even his motivations for the subject - could be relevant to the subject's good"; Rosati 1995, 299, holding that "the 'fully informed' person […] may not be someone whose judgments [an actual person] would recognize as authoritative"; Sobel 1994, 808, holding that "the hope of [assessing welfare] by constructing a vantage point fully informed […] is misguided"; Sarch 2015, 143, holding that ideal PSTW "become unilluminating" if the information condition presupposed by ideal PSTW is "taken to involve knowledge of the true theory of welfare"; Rizzo 2025, 2, holding that "the relationship between the satisfaction of counterfactual preferences and the actual individual's […] welfare is tenuous"). I do not expand here on these criticisms since such criticisms do not directly bear against informed PSTW.

[12] I expand on my defence of informed PSTW in Sections 3-7 (rather than here) to make it clear in what respects exactly my position differs from the positions advocated by prominent authors concerning the objections I examine in each section.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

follows. Individuals' informed preferences can be regarded as either *actual* attitudes or merely *hypothetical* attitudes. Now, if individuals' informed preferences are regarded as actual attitudes, then the claim that individuals have 'inner rational agents' with well-informed and consistent informed preferences* "lacks *psychological foundations*" (Sugden 2018, 13, italics added). For "there is no general reason" to think that 'inner rational agents' with well-informed and consistent informed preferences* "exist at all" (Infante et al. 2016b, 34; also Infante et al. 2016a, 22). Conversely, if individuals' informed preferences are regarded as merely hypothetical attitudes, then these preferences lack sufficient connection to individuals' welfare to ground reliable and informative welfare evaluations. For what one would prefer under hypothetical circumstances may be rather uninformative about her welfare (e.g., Cowen 1993, 265, holding that "a self with radically different brain endowments and capacities […] cannot judge my welfare [because such self] is a different individual altogether").

This objection correctly notes that some versions of informed PSTW presuppose (rather than show) that individuals have a set of well-informed and consistent informed preferences* (e.g., Infante et al. 2016a, for illustrations). Still, there are at least two reasons to doubt that the objection undermines informed PSTW. First, informed PSTW do *not* rest on the assumption that individuals have inner rational agents with informed preferences*. In particular, the proponents of informed PSTW may provide detailed specifications of the *conditions* under which individuals are plausibly ascribed informed preferences* *without* having to posit any inner rational agents having such preferences* (e.g., Hausman 2016). For informed PSTW's information and consistency conditions do not concern whether individuals' preferences are formed via any particular psychological process. And the proponents of informed PSTW can determine what preferences meet such conditions without having to posit any inner rational agents (e.g., Fumagalli 2024, Beck 2023).[13]

[13] Evidence about psychological processes may inform philosophers'/welfare economists' attempts to reconstruct informed preferences* and discriminate between competing reconstructions of informed preferences* (e.g., Manzini and Mariotti 2014, Rubinstein and Salant 2012, on so-called model-based approaches, which attempt to reconstruct informed preferences* by drawing on specific assumptions about the neuro-psychological processes generating individuals' choices). Still, the proponents of

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

And second, the proponents of informed PSTW can identify preferences that *both* meet the information and consistency conditions presupposed by informed PSTW *and* have sufficient connection to individuals' welfare to ground reliable and informative welfare evaluations (e.g., Bernheim 2021, Fumagalli 2025, on various sets of well-informed and transitive preferences). To be sure, the fact that an individual's preferences meet *some* information and consistency conditions does not *per se* imply that satisfying such preferences is plausibly taken to constitute the individual's welfare. For *not all* information and consistency conditions are plausibly taken to reliably track individuals' welfare (Section 2; also Fumagalli 2021). Still, the information and consistency conditions presupposed by leading versions of informed PSTW (e.g., transitivity, accurate information about the available choice options) provide a *reliable criterion* for reconstructing informed preferences*. For satisfying well-informed/consistent preferences tends to yield individuals higher welfare than satisfying ill-informed/inconsistent preferences (e.g., Beshears et al. 2008, on cases where satisfying ill-informed preferences prevents individuals from achieving their own welfare-related goals; Gustafsson 2022, sec.4, on cases where satisfying intransitive preferences makes individuals vulnerable to sure loss).[14]

A critic of informed PSTW may object that informed PSTW evaluate individuals' welfare "*relative* to the preferences that [individuals] *would* have revealed if not subject to reasoning imperfections", and so implicitly presuppose that individuals have well-informed and consistent *latent* preferences*, i.e. preferences "that are formed *within the minds* of individual[s and] do not correspond directly with

---

informed PSTW are not committed to making any specific assumptions about psychological processes (e.g., Bernheim and Rangel 2009, Salant and Rubinstein 2008, on so-called model-less approaches, which attempt to reconstruct informed preferences* without drawing on any specific assumptions about neuro-psychological processes).

[14] Satisfying ill-informed/inconsistent preferences does not invariably hamper individuals' welfare (e.g., Whitman and Rizzo 2015, 419-420). However, as noted in the main text, satisfying well-informed/consistent preferences tends to yield individuals higher welfare than satisfying ill-informed/inconsistent preferences. The information and consistency conditions presupposed by informed PSTW can be defended also by pointing to synchronic (rather than diachronic) considerations (e.g., Williamson 2024, on transitivity) and to individuals' willingness to revise their choices in accordance with such information/consistency conditions (e.g., Hands 2014, 401-2, Nielsen and Rehbeck 2022, 2237-9, on experimental evidence demonstrating individuals' willingness to revise intransitive choices when they realize these choices' intransitivity).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

objective properties of the external world" (Infante et al. 2016a, 7 and 9, italics added; also Infante et al. 2016b, 33). However, the proponents of informed PSTW can ascribe individuals well-informed and consistent informed preferences* without presupposing that individuals have well-informed and consistent latent preferences*. To illustrate this, consider Bernheim and Rangel's preference-based approach, which aims to reconstruct a range of informed preferences* in settings where individuals' choices depend on ancillary conditions, i.e. "feature[s] of the choice environment that may affect behaviour, but [are] not taken as relevant to [welfare]" (2009, 55).[15]

Bernheim and Rangel's approach relies on the notion of 'unambiguous choice' as its welfare criterion. The idea is that "one alternative is unambiguously superior to another if and only if the second is never chosen when the first is available" to individuals (Bernheim 2016, 15). Conversely, when individuals' choices between two options vary across ancillary conditions, one should regard it as indeterminate which option enhances individuals' welfare unless the observed choices result from demonstrable mistakes, i.e. are "predicated on a characterization of the available options […] that is inconsistent with the information available" to individuals and "there is some other option in the opportunity set that [individuals] would select [in the absence of such] characterization failure" (2016, 48). According to some critics, Bernheim and Rangel's approach presupposes that individuals have "a neoclassical agent deep inside that is struggling to surface" (Rizzo and Whitman 2020, 80; also Sugden 2018, 57). However, the approach does not assume "a context-independent objective function […] defined over a domain encompassing all the options of potential interest" (Bernheim 2021, 392). In particular, the approach does not define mistakes in terms of divergences between choices and latent preferences*, and "does not assume that error-free choices reveal" well-informed and consistent latent preferences* (ibid., 392). In fact, Bernheim explicitly claims that

[15] In recent works, Bernheim notes that he does "no longer find [himself] in complete agreement with all the positions" (2016, 13) advocated in Bernheim and Rangel (2009). Still, the differences between Bernheim's works have limited relevance for the illustration in the main text. For even in his later works, Bernheim emphasizes that "the Bernheim–Rangel apparatus can serve as the foundation for a practical and unified approach to [welfare evaluations]" (2016, 13; also Bernheim 2021).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

individuals frequently "aggregate the many diverse aspects of [their] experience only when called to [choose]" (2016, 20).[16]

4. Objection from Unfeasible Preference Reconstruction

The *objection from unfeasible preference reconstruction* holds that informed PSTW do not withstand scrutiny because philosophers/welfare economists cannot reliably reconstruct well-informed and consistent informed preferences*. The objection proceeds as follows. Suppose, for the sake of argument, that individuals can be plausibly taken to *have* well-informed and consistent informed preferences*. Even so, the assumption that philosophers/welfare economists "can *reconstruct* [these preferences*] is a *mirage*" (Sugden 2018, 14, italics added). For the information and consistency conditions presupposed by informed PSTW frequently allow for different (and sometimes contradictory) reconstructions of individuals' informed preferences* (e.g., Matson 2022). And apparent conflicts between preferences can typically be resolved in multiple ways (e.g., Whitman and Rizzo 2015, on the difficulty of identifying welfare-optimal rates of saving and intertemporal discounting). Hence, philosophers/welfare economists often "have no means of determining which of the conflicting preferences reflect [informed preferences*]" (Rizzo and Whitman 2020, 75; also Dold 2018, 161).

This objection correctly notes that philosophers'/welfare economists' attempts to reconstruct informed preferences* face significant epistemic and normative

[16] Bernheim's claim that individuals often construct their preferences when called to choose stands in tension with the assumption that individuals have well-informed and consistent latent preferences*, but is compatible with preference-based approaches. To be sure, some contend that Bernheim (2016) "implicitly abandons" preference-based approaches on the alleged ground that he characterizes individuals' welfare in terms of "attitudes that stand at the beginning of the reasoning process" and allows to "no longer defer to revealed preference in cases where we have […] good evidence that there has been a mistake" (Thoma 2021a, 356). However, these contentions do not undermine the plausibility of regarding Bernheim's approach as preference-based. For the welfare-relevant attitudes envisioned by Bernheim can be plausibly regarded as preferences. In fact, one may regard Bernheim's approach as a version of informed PSTW since such approach imposes information and consistency conditions on preferences that are formally analogous to the information and consistency conditions imposed by informed PSTW (e.g., Bernheim 2016, 58-59, and 2021, 395-6, imposing acyclicity and consistency with information concerning the available options).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

challenges (e.g., Pettigrew 2023, on the epistemic and normative assumptions required to establish whether correcting specific inconsistencies enhances individuals' welfare). Still, there are at least two reasons to doubt that the objection undermines informed PSTW. First, philosophers/welfare economists *can reconstruct* informed preferences* in several cases where the involved individuals *fail* to exhibit well-informed and consistent actual preferences (e.g., Bernheim and Rangel 2009, for reconstructions of informed preferences* in settings where choices are affected by ancillary conditions; Salant and Rubinstein 2008, for reconstructions of informed preferences* in settings where choices are affected by frames). To be sure, philosophers/welfare economists may be unable to reconstruct informed preferences* in presence of *widespread* choice inconsistencies (e.g., Sugden 2018, 58; also Bernheim 2016, 60, conceding that his approach "may not be very discerning […] in settings where choice inconsistencies are pervasive"). Yet, individuals' choice inconsistencies are *rarely* so widespread that they prevent philosophers/welfare economists from reconstructing informed preferences*. To illustrate this, consider situations where individuals make some intransitive choices. These choices complicate philosophers'/welfare economists' attempts to reconstruct individuals' informed preferences*, but do not generally prevent philosophers/welfare economists from reconstructing such preferences*. For philosophers/welfare economists are frequently able to reconstruct informed preferences* in presence of some intransitive choices based on the core of transitive choices made by the involved individuals (e.g., Nishimura 2018, 589-599, for reconstructions of informed risk/time preferences* based on the core of transitive choices made by individuals). And philosophers/welfare economists can often point to experimental evidence demonstrating that individuals tend to regard transitivity as normatively compelling and are willing to revise intransitive choices when they realize these choices' intransitivity (e.g., Hands 2014, 401-2, Nielsen and Rehbeck 2022, 2237-9).

And second, philosophers/welfare economists can frequently rely on *multiple sources* of evidence to reconstruct informed preferences*. In fact, philosophers/welfare economists have grounded several reconstructions of informed preferences* on both *choice-based* sources of evidence (e.g., Bernheim

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

and Taubinsky 2018, on information concerning individuals' hypothetical choices; Ferreira 2023, on information concerning the choices individuals would repeat at the time of welfare evaluation) and *non-choice-based* sources of evidence (e.g., Arieli et al. 2011, on eye-tracking data showing whether individuals attend to the available choice options; Bernheim 2016, on factual questions with objectively verifiable answers showing whether individuals understand the examined choice problems). To be sure, philosophers'/welfare economists' attempts to reconstruct informed preferences* typically *depend* on normative/evaluative presuppositions about the notion of welfare (e.g., Haybron and Tiberius 2015, 714-7). However, these dependences do not reflect limitations inherent in informed PSTW, but rather reflect the thickness of the notion of welfare, i.e. the fact that this notion involves both positive and normative/evaluative dimensions (e.g., Dold and Schubert 2018, 223-4) and that, as a result, welfare ascriptions typically rely on both positive and normative/evaluative presuppositions (e.g., Fletcher 2019, 703-4).

A critic of informed PSTW may object that distinct sources of evidence may ground *conflicting* reconstructions of informed preferences* and that the information and consistency conditions presupposed by informed PSTW do not enable philosophers/welfare economists to *discriminate* between such reconstructions, i.e. to determine which of the proffered reconstructions of informed preferences reliably track informed preferences* (e.g., Whitman and Rizzo 2015, 420-4). The idea is that philosophers/welfare economists frequently face substantial *normative ambiguity* and that the information and consistency conditions presupposed by informed PSTW do not enable philosophers/welfare economists to resolve such ambiguity (e.g., Berg and Gigerenzer 2010, 148-150, on putative cases where satisfying ill-informed and inconsistent preferences enhances individuals' welfare).

However, the objection significantly *overestimates* the extent of normative ambiguity inherent in individuals' preferences. For as noted in Section 3, satisfying well-informed/consistent preferences tends to yield individuals higher welfare than satisfying ill-informed/inconsistent preferences. Moreover, the information and consistency conditions presupposed by informed PSTW provide

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

a reliable (though fallible) basis to *resolve* the normative ambiguity inherent in individuals' preferences by discriminating between conflicting reconstructions of informed preferences*, or at least by narrowing down the set of plausible reconstructions of such preferences*. To illustrate this, consider situations where individuals exhibit varying willingness to pay for specific goods/experiences across multiple frames. This variability complicates philosophers'/welfare economists' attempts to reconstruct informed preferences*, but does not *per se* prevent philosophers/welfare economists from reconstructing a range of informed preferences* by demarcating precise and plausible bounds for minimal and maximal willingness to pay for the examined goods/experiences (e.g., Bernheim 2016, 60-64; also Abrahamson 2024, 24-26, for additional illustrations of philosophers'/welfare economists' ability to reconstruct informed preferences* in cases where the involved individuals exhibit context-dependent preferences).[17]

5. Objection from Dubious Normative Commitments

The *objection from dubious normative commitments* holds that informed PSTW do not withstand scrutiny because informed PSTW's normative assumption that the satisfaction of informed preferences constitutes welfare is implausible. The objection proceeds as follows. Suppose, for the sake of argument, that individuals can be plausibly taken to *have* well-informed and consistent informed preferences. Assume further that philosophers/welfare economists can reliably *reconstruct* these preferences. Even so, the satisfaction of such preferences is not plausibly taken to *constitute* individuals' welfare. For a given state of affairs is not good for one "*simply because* [one prefers] with proper information, and

[17] A critic of informed PSTW may object that the normative ambiguity inherent in many individuals' preferences frequently prevents reliable reconstructions of informed preferences* on the alleged ground that "the correct weighting" of the benefits and costs of individuals' choices "is unavoidably subjective" (Rizzo and Whitman 2020, 407-8). However, this objection seemingly presupposes (rather than supports) a radical subjectivist conception of welfare, according to which the extent to which individuals are well-off exclusively depends on individuals' subjective judgments/attitudes toward their lives. And such conception of welfare is vulnerable to serious objections (e.g., Arneson 1999, 141-2, Kagan 2009, 254-5, Lin 2017, 357-368, Parfit 1984, 501-2, Scanlon 1998, ch.3; also Heathwood 2014, 202-7, Hurka 2019, 453-9, Kagan 1992, 187-8, Keller 2009, 676-9, Wall and Sobel 2021, 2842-51, on various objectivist and hybrid conceptions of welfare).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

reflectively [such state of affairs] to occur" (Kraut 2007, 118, italics added; also Scanlon 1998, 115). In particular, "it is *one thing* to determine what people's [informed] preferences would be […] and it is *a different thing* to determine what is good for people" (Hausman 2016, 30, italics added; also Hausman and McPherson 2009, 12, holding that "the fact that [one] prefers *x* to *y* does not make it the case that *x* is better for [her] than *y*, no matter what conditions one imposes on [her] preferences").

This objection correctly notes that substantiating informed PSTW requires the proponents of informed PSTW to support informed PSTW's normative assumption that the satisfaction of informed preferences constitutes welfare. Still, there are at least two reasons to doubt that the objection undermines informed PSTW. First, supporting informed PSTW's normative assumption that the satisfaction of informed preferences constitutes welfare is *less demanding* than the objection seems to presuppose. To illustrate this, let us distinguish between *fundamental* constituents of welfare - i.e. non-instrumentally valuable goods/experiences whose constitutive relation with welfare grounds the constitutive relation (if any) between all other non-instrumentally valuable goods/experiences and welfare - and *intermediate* constituents of welfare - i.e. non-instrumentally valuable goods/experiences whose constitutive relation with welfare is grounded in the constitutive relation between fundamental constituents and welfare. Supporting informed PSTW's normative assumption that the satisfaction of informed preferences constitutes welfare does not require the proponents of informed PSTW to provide an exhaustive specification of all (fundamental and intermediate) constituents of welfare (e.g., Rabinowicz and Osterberg 1996, 8-12). In particular, one may consistently endorse informed PSTW and acknowledge the existence of multiple intermediate constituents of welfare. For the issue of whether a given good/experience is a constituent of welfare is conceptually distinct from the issue of whether the constitutive relation between this good/experience and welfare (if any) is grounded in the constitutive relation between some other goods/experiences and welfare. In fact, several versions of informed PSTW allow that the satisfaction of informed preferences may constitute welfare through dissimilar intermediate constituents of welfare in different contexts (e.g., footnote no.5 on the possibility that the satisfaction of

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

informed preferences may constitute welfare through the contingent link between the satisfaction of an individual's informed preferences and the sense of satisfaction that the individual may derive from knowing that such preferences are satisfied).

And second, the satisfaction of preferences that meet the information and consistency conditions presupposed by leading versions of informed PSTW (e.g., transitivity, accurate information about the available choice options) can be *plausibly* taken to constitute individuals' welfare (Sections 3-4). To be sure, one may point to several cases where philosophers/welfare economists *disagree* as to whether satisfying specific sets of informed preferences constitutes individuals' welfare (e.g., Griffin 1986, ch.1, Sumner 1996, ch.5, on cases where individuals are unaware that their informed preferences are satisfied; Parfit 1984, 494-5, Scanlon 1996, 111, on cases where individuals' informed preferences target states of affairs that seem unrelated to individuals' own welfare). However, the existence of contested cases does *not per se* license scepticism about informed PSTW. For many cases are not contested (e.g., individuals are often aware of whether their informed preferences are satisfied; individuals' informed preferences frequently target states of affairs related to what most theories of welfare regard as individuals' own welfare). And most contested cases are contested because of the normative/evaluative complexity inherent in such cases rather than because of alleged shortcomings inherent in informed PSTW (e.g., Fumagalli 2022, 532-3, Sunstein 2015, 518-9). That is to say, adopting theories of welfare other than informed PSTW does not *per se* enable philosophers/welfare economists to avoid contested cases (e.g., Griffin 1986, 17, Keller 2009, 656). And the proponents of informed PSTW may consistently endorse dissimilar positions concerning the proffered contested cases (e.g., Hawkins 2019, on recent debate about cases where individuals are unaware that their informed preferences are satisfied; Heathwood 2019, on recent debate about cases where individuals' informed preferences target states of affairs that seem unrelated to individuals' own welfare).[18]

[18] Various contested cases besides those cited in the main text have attracted philosophical debate, including: cases where individuals' informed preferences putatively target objectively worthless or objectively neutral states of affairs (e.g., Kagan

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

A critic of informed PSTW may object that informed PSTW rest on *unnecessary normative commitments* since philosophers/welfare economists can ground reliable and informative welfare evaluations on information concerning individuals' informed preferences *without* endorsing any theory of welfare (e.g., Hausman 2010, 341, Hausman and McPherson 2009, 16). The idea is that philosophers/welfare economists should retain informed PSTW's aim to ground reliable and informative welfare evaluations on information concerning individuals' informed preferences, but should relinquish informed PSTW's assumption that the satisfaction of informed preferences constitutes welfare because there is an *evidential* (rather than *constitutive*) connection between the satisfaction of informed preferences and welfare (e.g., Scanlon 1998, 116-8). However, it is dubious that appealing to this purported evidential connection undermines the justifiability of relying on informed PSTW. To illustrate this, consider the so-called evidential account of welfare, according to which "if individuals seek to *benefit themselves* and are *good judges* of what is good for them, then […] their preferences will be reliable indicators of what is good for them […] *regardless* of what theory of welfare one accepts" (Hausman and McPherson 2009, 1-2, italics added; also Hausman 2012, 89).

The evidential account has gained significant prominence among the proponents of preference-based approaches in the recent economic and philosophy of science literatures (e.g., Beck 2023). Still, it is hard to establish whether the satisfaction of preferences that meet the conditions posited by the evidential account provides reliable evidence for welfare unless one relies on *specific theories* of welfare (e.g., Sarch 2015, 143-6). Moreover, the evidential account appears to have quite a *limited domain* of applicability (e.g., Hersch 2015, 282-3; also Hausman 2016, 29, acknowledging that the conditions posited by the evidential account, taken collectively, "are often not met"). These complications do not undermine the

2009, 254-5, Kraut 1994, 43-45, Sobel and Wall 2023, 7-8); cases where individuals' informed preferences allegedly target states of affairs that individuals are incapable of finding valuable or attractive (e.g., Rosati 1996, 297-9, Sarch 2011, 178-182, Wall and Sobel 2021, 2845-6); and cases where individuals purportedly have informed preferences to sacrifice their own welfare or simply be badly off (e.g., Bradley 2007, 45-47, Heathwood 2011, 18-19, Rosati 2009, 312-3). I do not expand on these contested cases since, as noted in the main text, the proponents of informed PSTW may consistently endorse dissimilar positions about such cases.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

justifiability of relying on the evidential account in choice settings where the conditions posited by such account hold (e.g., Hausman 2022a, 355-6), but constrain the evidential account's potential to ground reliable and informative welfare evaluations. In particular, they make it difficult to see on what basis philosophers/welfare economists may rely to establish whether the satisfaction of preferences provides reliable evidence for welfare in choice settings where philosophers/welfare economists are unable to determine whether the conditions posited by the evidential account are met and in choice settings where the conditions posited by the evidential account are not met (e.g., Fumagalli 2021, 126-8). In this respect, the evidential account's purported agnosticism concerning the correct theory (or theories) of welfare appears to significantly constrain the evidential account's potential to ground reliable and informative welfare evaluations.[19]

6. Objection from Conceptual Ambiguity

The *objection from conceptual ambiguity* holds that informed PSTW do not withstand scrutiny because informed PSTW are premised on dissimilar (and often conflicting) conceptions of preferences. The objection proceeds as follows. In the economic and philosophy of science literatures, *multiple conceptions* of preferences have been advocated, which rest on dissimilar (and often conflicting) presuppositions regarding the relationship between preferences and choices (e.g., Thoma 2021b, Vredenburgh 2024, on the relationship between preferences and actual or hypothetical choices), the causal bases of preferences (e.g., Guala 2019,

[19] Leading proponents of the evidential account concede that philosophers/welfare economists "need to [have] some notion of what is good for people" to justifiably regard the satisfaction of specific sets of preferences as evidence for welfare, but maintain that philosophers/welfare economists "do not have to wait for a satisfactory philosophical theory of welfare" (Hausman 2012, 92; also Hausman and McPherson 2009, 18). The idea is that "knowing that good health, happiness, enjoyment [...] generally contribute to welfare gives content to talk of welfare without defining the term" (Hausman 2010, 341) and that philosophers/welfare economists "can use that knowledge" to ground reliable and informative welfare evaluations (Hausman 2022b, 11). However, generic claims such as the claim that 'good health, happiness, enjoyment [...] generally contribute to welfare' do not ground reliable and informative welfare evaluations in the many cases where different theories of welfare disagree (e.g., Fumagalli 2022). And in such cases, grounding reliable and informative welfare evaluations would require philosophers/welfare economists to take a position concerning the merits of different theories (e.g., Kelman 2005, Sarch 2015).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

Ross 2011, on the issue of whether preferences are more plausibly regarded as mental attitudes, dispositions or behavioural patterns), and the nature of preferences (e.g., Broome 1993, Hausman 2012, ch.7-8, on the issue of whether preferences are more aptly characterized as judgments or feelings). However, substantiating informed PSTW requires the proponents of informed PSTW to specify *which conceptions* of preferences they endorse and put forward convincing *reasons/evidence* in favour of such conceptions. For the plausibility of informed PSTW critically depends on the merits of the conceptions of preferences on which informed PSTW are premised (e.g., Dietrich and List 2016b). Regrettably, the objection goes, the proponents of informed PSTW have hitherto failed to address these specification and justification challenges (e.g., Lecouteux 2022).[20]

This objection correctly notes that substantiating informed PSTW requires the proponents of informed PSTW to specify which conceptions of preferences they endorse and put forward convincing reasons/evidence in favour of such conceptions. Still, there are at least two reasons to doubt that the objection undermines informed PSTW. First, the information and consistency conditions presupposed by informed PSTW impose *several constraints* on admissible conceptions of preferences (e.g., Hausman 2011, 7-10, on how the consistency conditions presupposed by informed PSTW imply that preferences are inherently comparative). This does not *per se* enable the proponents of informed PSTW to univocally determine what conception of preferences philosophers/welfare economists should adopt in specific contexts (e.g., Mandler 2005, 255-6, on how the plausibility of various consistency conditions may itself vary depending on what conception of preferences one endorses). Still, it enables the proponents of informed PSTW to significantly narrow down the set of plausible conceptions of preferences (e.g., Cozic and Hill 2015, 297-9).

[20] Not all leading authors in the economic and philosophy of science literatures endorse a pluralistic view of the notion of preference (e.g., Hausman 2012, ch.7-8, arguing that preferences in welfare economics are most plausibly regarded as total subjective comparative evaluations). However, most leading authors doubt that a single conception can capture the different senses that the notion of preference may be plausibly ascribed in welfare economics (e.g., Sen 1977; also Angner 2018, Hausman 2023, for recent debate).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

And second, the proponents of informed PSTW may justify their reliance on informed PSTW *without* having to specify and support a *single general* conception of preferences. For the merits of different conceptions of preferences are plausibly taken to depend on the theoretical and pragmatic presuppositions of the models and the policy applications where preferences figure (e.g., Angner 2018, 675-9). And different conceptions of preferences may be suitable for distinct modelling and policy purposes (e.g., Beck 2024, 1444-50; also Vredenburgh 2021). To be sure, theoretical terms such as 'preference' may have specific *pre-theoretic connotations* (e.g., Hausman 1998, 197-8, Mäki 1998, 306, on folk psychological conceptions of preferences). Yet, these pre-theoretic connotations do not determine the meaning of the theoretical notion of 'preference' figuring in informed PSTW (e.g., Ross 2012, 182-4). Hence, the proponents of informed PSTW may consistently acknowledge the existence of such pre-theoretic connotations and advocate distinct conceptions of preferences (e.g., Guala 2012, 137-9).

A critic of informed PSTW may object that the proponents of informed PSTW should "*clarify* the concept of preferences [they endorse] rather than leaving preferences to be defined *implicitly* by formal conditions and [by their] explanatory and predictive practices" (Hausman 2024, 213, italics added). The idea is that although the proponents of informed PSTW "are free to reconceive of preferences in any way they wish [the proffered] reconceptualizations are not beyond criticism" (ibid., 224; also Sen 1973, 259). However, the proponents of informed PSTW can draw on *several considerations* to assess the comparative merits of different conceptions of preferences and support the specific conceptions they endorse. To illustrate this, consider the ongoing debate concerning the comparative merits of behaviorist conceptions of preferences, which regard preferences as indexes of choices (e.g., Gul and Pesendorfer 2008), and mentalist conceptions of preferences, which regard preferences as mental attitudes (e.g., Rubinstein and Salant 2008).

Behaviorist conceptions of preferences appear to be more general than mentalist conceptions of preferences since mentalist conceptions "limit the attribution of preferences to those with the requisite mental capacities" (Hausman 2024, 223).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

In particular, adopting a behaviorist conception of preferences allegedly enables philosophers/welfare economists to "black-box […] the psychological processes that lead to choice" and thereby avoid "controversial substantive commitments about psychological processes" (Thoma 2021b, 165). Conversely, mentalist conceptions of preferences purportedly provide a more informative evidential basis to assess individuals' welfare than behaviorist conceptions of preferences (e.g., Sumner 1996, ch.5). In particular, adopting a mentalist conception of preferences may usefully constrain philosophers'/welfare economists' attempts to reconstruct informed preferences (e.g., Hausman 2011, on how information concerning individuals' beliefs can help philosophers/welfare economists determine how individuals conceive of the choice options they face). These observations do not determine whether, in general, philosophers/welfare economists should adopt behaviorist or mentalist conceptions of preferences. For what conceptions of preferences philosophers/welfare economists should adopt may plausibly depend on various contextual elements such as individuals' cognitive/computational abilities (e.g., Okasha 2016) and whether philosophers/welfare economists aim to ground individual welfare evaluations or social welfare evaluations (e.g., Moscati 2021). Still, they nicely illustrate that the proponents of informed PSTW can draw on several considerations to assess the comparative merits of different conceptions of preferences and support the specific conceptions they endorse.[21]

7. Objection from Conceptual Replacement

The *objection from conceptual replacement* holds that informed PSTW do not withstand scrutiny because grounding reliable and informative welfare evaluations requires philosophers/welfare economists to replace preference-based approaches with non-preference-based approaches. The objection proceeds as follows. To ground reliable and informative welfare evaluations, philosophers/welfare economists should distinguish between *fundamental*

[21] Analogous remarks may be made concerning dispositionalist conceptions of preferences, which regard preferences as belief-dependent dispositions with multiply realizable causal bases (e.g., Guala 2019; also Beck 2024, 1446, holding that adopting a dispositionalist conception of preferences "avoids many of the pitfalls of [behaviorist and mentalist] conceptions", but faces the challenge to explicate "how exactly preferences [depend] on informational states").

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

attitudes that "are the starting point of deliberation [and] shouldn't be changed by the reasoning process" and *non-fundamental attitudes* that "may be formed in deliberation [and] can be described as mistaken" in light of the fundamental attitudes (e.g., Thoma 2021a, 355 and 361, on the putative contrast between "fundamental desires regarding features of the available options" and less fundamental preferences). Abiding by this distinction, however, would require philosophers/welfare economists to regard preferences as the outcome of reasoning processes that involve more fundamental attitudes than preferences and thereby abandon preference-based approaches (and, therefore, informed PSTW; e.g., Rizzo 2025, 10, holding that "in back of preferences is desire [and] what is prudentially good for the individual is what she desires").[22]

This objection correctly notes that non-preference-based approaches may enable philosophers/welfare economists to ground reliable and informative welfare evaluations. Still, there are at least two reasons to doubt that the objection undermines preference-based approaches (and, therefore, informed PSTW). First, philosophers/welfare economists can ground reliable and informative welfare evaluations by *refining* (rather than *replacing*) preference-based approaches. To illustrate this, consider again the challenges that apparent inconsistencies in individuals' actual preferences pose to philosophers'/welfare economists' attempts to ground reliable and informative welfare evaluations on information concerning individuals' preferences. The proponents of preference-based approaches have addressed various such challenges by distinguishing between different sets of preferences (e.g., Hausman 2012, 36-37, for a preference-based approach distinguishing 'basic preferences', which are independent of individuals' beliefs about "the character and consequences" of the available

[22] Various non-preference-based approaches to welfare evaluations have been developed in the literature besides desire-based approaches (e.g., Haybron and Tiberius 2015, who advocate grounding welfare evaluations on individuals' values rather than individuals' preferences; Kahneman et al. 1997, who advocate grounding welfare evaluations on measures of experienced utility rather than measures of preference satisfaction; Sugden 2018, ch.4-5, who advocates grounding welfare evaluations on measures of opportunities rather than measures of preference satisfaction). Below I focus on desire-based approaches since the debate about other non-preference-based approaches is already well-advanced in the specialized literature (e.g., Hersch 2020, for a critical appraisal of value-based approaches; Fumagalli 2019, for a critical appraisal of experienced utility approaches; Fumagalli 2024, for a critical appraisal of opportunity-based approaches).

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

options, and 'non-basic preferences', which take into account these beliefs and may influence basic preferences in light of such beliefs; also Rubinstein and Salant 2012, 375, for a preference-based approach distinguishing 'observed preference orderings', which vary as the result of cognitive processes, and 'underlying preferences', which purportedly "reflect [individuals'] welfare").[23]

And second, replacing preference-based approaches with non-preference-based approaches does *not per se* enable philosophers/welfare economists to ground *more* reliable and informative welfare evaluations. For the proffered non-preference-based approaches face *major* conceptual and practical challenges and *radically diverge* on a number of foundational issues. To illustrate this, consider desire-based approaches. These approaches face major conceptual and practical challenges (e.g., Thoma 2021a, 360, on cases where desire-based welfare evaluations are indeterminate because the involved individuals' putatively fundamental desires are vague) and radically diverge on a number of foundational issues, including what notions of desire should ground welfare evaluations (e.g., actual versus informed versus ideal desires), on what grounds philosophers/welfare economists should differentiate between more or less allegedly fundamental desires (e.g., the mere fact that a desire happens to be 'the starting point of deliberation' falls short of implying that such desire 'shouldn't be changed by the reasoning process') and what exactly the connection between the posited desires and individuals' welfare is (e.g., constitutive versus evidential connection). These divergences do not exclude the possibility that specific desire-based approaches may ground reliable and informative welfare evaluations, but cast doubt on the claim that philosophers/welfare economists should replace preference-based approaches with desire-based approaches.

A critic of informed PSTW may object that the highlighted divergences between the proffered non-preference-based approaches point to *open problems* in these approaches, but do not *justify* philosophers'/welfare economists' reliance on

[23] Hausman (2012) advocates the evidential account rather than informed PSTW (Section 5). However, as noted in Section 1, both the evidential account and informed PSTW belong to the set of preference-based approaches in that both the evidential account and informed PSTW aim to ground reliable and informative welfare evaluations on information concerning individuals' preferences.

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

preference-based approaches. In particular, the critic may maintain that *neuro-psychological findings* may enable philosophers/welfare economists to *reduce* preference-based approaches to non-preference-based approaches grounded on empirical findings concerning the neuro-psychological substrates of choices (e.g., Glimcher 2011, ch.6-8). The idea is that neuro-psychological findings provide "a tool for measuring preferences neurobiologically" (Levy and Glimcher 2012, 1027) and enable policy makers to "design [policies] that maximize welfare" (Loewenstein and Haisley 2008, 238).

However, the great heterogeneity of the neuro-psychological substrates of welfare-enhancing choices *casts doubt* on the prospects of reductive non-preference-based approaches. For a given neuro-psychological process may contribute to generating choices having rather different welfare implications, and dissimilar sets of neuro-psychological processes may contribute to generating choices having similar welfare implications across choice settings (e.g., Ross 2014, Schulz 2024). Moreover, preference-based approaches frequently enable philosophers/welfare economists to ground reliable and informative welfare evaluations *without* having to draw on specific assumptions concerning neuro-psychological processes (e.g., Fumagalli 2019, on informed PSTW; also Section 3). These considerations do not exclude the possibility of grounding reliable and informative welfare evaluations on non-reductive non-preference-based approaches. However, together with the open problems faced by such approaches, they justify philosophers'/welfare economists' reliance on preference-based approaches.

8. Conclusion

Let us take stock. In recent years, several prominent authors have argued that PSTW fail to provide a plausible theory of welfare. In this article, I have explicated and addressed the most influential objections put forward against specific versions of PSTW, namely informed PSTW. In particular, I have argued that although PSTW face significant conceptual and practical challenges, the critics of PSTW have hitherto failed to substantiate convincing objections against informed PSTW. This result does not exclude the possibility that additional

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

objections may be put forward against informed PSTW. Still, as things stand, it demonstrates that philosophers/welfare economists are justified in relying on such versions of PSTW.

More generally, I take the considerations in this article to contribute to the ongoing cross-disciplinary debate about the plausibility of different theories of welfare in at least two respects of wide interest to philosophers/welfare economists. The first contribution concerns the *conceptual and practical import* of the objections put forward against specific theories of welfare. To illustrate this, consider again the objections put forward against informed PSTW. As argued in the previous sections, various objections share a tendency to misrepresent model-specific problems and particular contested cases as general conceptual and practical challenges to informed PSTW. This, however, by no means implies that the proffered objections are without merit. On the contrary, such objections provide valuable critical insights concerning philosophers'/welfare economists' ability to reliably reconstruct welfare-relevant preferences in specific choice settings (e.g., Section 4 on the constraints imposed by widespread choice inconsistencies), the descriptive/normative adequacy of specific information and consistency conditions (e.g., Section 3 on transitivity), and the alleged need to supplement these conditions with further conditions on welfare-relevant preferences (e.g., Section 2 on moral considerations).

The second contribution concerns the need to heed *cross-disciplinary differences* when assessing the plausibility of different theories of welfare. To illustrate this, consider again philosophers' and welfare economists' respective contributions to the debate concerning informed PSTW. On the one hand, philosophers frequently engage in this debate at a higher level of abstraction than welfare economists and occasionally seem to overlook that welfare economists' model specifications allow for more flexibility in the definition of preferences than many philosophers seek (e.g., Section 6). On the other hand, welfare economists often gloss over what many philosophers regard as significant normative/evaluative questions concerning individuals' welfare and occasionally seem to overlook philosophically motivated reasons to doubt that the satisfaction of empirically elicited preferences constitutes welfare (e.g., Section 5). In this context,

https://doi.org/10.1017/psa.2025.6 Published online by Cambridge University Press

philosophers' growing attention to welfare economists' modelling practices and welfare economists' deeper engagement with philosophers' normative/evaluative discussions can greatly advance the ongoing cross-disciplinary debate about the plausibility of different theories of welfare.

Competing Interests: The Author declares none.

Funding Information: The Author acknowledges the support of the Centre for the Study of Governance and Society (King's College London) and the John Templeton Foundation, 'The Political Economy of Knowledge and Ignorance', Grant #61823.

Acknowledgements: I wish to thank Lukas Beck, Dan Hausman, Gil Hersch, Johanna Thoma and Jack Vromen for their comments on previous versions of this article. I also received helpful feedback from audiences at the 5th Public Policy and Regulation Workshop (King's College London), Bicocca University (Milan), the 16th Biennial Meeting of the International Network for Economic Method (Venice), the workshop 'Adaptive Preferences, Structural Injustice and Moral Responsibility' (University of Zurich), the Grand Est 'Economics and Philosophy' webinar (University of Strasbourg and University of Lorraine), the European PPE Network 2nd Annual Conference (University of Warwick), the University of Bristol, Stanford University, and Rutgers University.

REFERENCES

Abrahamson, Mans. 2024. "Permissible Preference Purification". *Journal of Economic Methodology* 31:17-35. https://doi.org/10.1080/1350178X.2023.2257212

Adler, Matthew. 2012. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

Adler, Matthew. 2019. *Measuring Social Welfare: An Introduction*. New York: Oxford University Press.

Adler, Matthew, and Marc Fleurbaey, Eds. 2016. *Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press.

Angner, Erik. 2011. “Are subjective measures of well-being ‘direct’?” *Australasian Journal of Philosophy* 89:115-130. https://doi.org/10.1080/00048400903401665

Angner, Erik. 2016. “Well-being and economics”. In *The Routledge Handbook of Philosophy of Well-being*, Guy Fletcher (Ed.), 492-503. London: Routledge. https://doi.org/10.4324/9781315682266

Angner, Erik. 2018. “What preferences really are”. *Philosophy of Science* 85:660-681. https://doi.org/10.1086/699193

Arieli, Amos, Yaniv Ben-Ami, and Ariel Rubinstein. 2011. “Tracking Decision Makers under Uncertainty”. *American Economic Journal: Microeconomics* 3:68-76. https://doi.org/10.1257/mic.3.4.68

Arneson, Richard. 1999. “Human flourishing versus desire satisfaction”. *Social Philosophy and Policy* 16:113-142. https://doi.org/10.1017/S0265052500002272

Beck, Lukas. 2023. “The Econ within or the Econ above?” *Economics & Philosophy* 39:423-445. https://doi.org/10.1017/S0266267122000141

Beck, Lukas. 2024. “Why We Need to Talk About Preferences”. *Erkenntnis* 89:1435-1455. https://doi.org/10.1007/s10670-022-00590-2

Berg, Nathan, and Gerd Gigerenzer. 2010. “As-If Behavioral Economics: Neoclassical Economics in Disguise?” *History of Economic Ideas* 18:133-166. https://doi.org/10.2139/ssrn.1677168

Bernheim, Douglas. 2009. “Behavioral welfare economics”. *Journal of the European Economic Association* 7:267-319. https://doi.org/10.1162/JEEA.2009.7.2-3.267

Bernheim, Douglas. 2016. “The good, the bad, and the ugly: a unified approach to behavioral welfare economics”. *Journal of Benefit-Cost Analysis* 7:12-68. https://doi.org/10.1017/bca.2016.5

Bernheim, Douglas. 2021. “In defense of behavioral welfare economics”. *Journal of Economic Methodology* 28:385-400. https://doi.org/10.1080/1350178X.2021.1988133

Bernheim, Douglas, Andrey Fradkin, and Igor Popov. 2015. “The Welfare Economics of Default Options in 401(k) Plans”. *American Economic Review* 105:2798–2837. https://doi.org/10.1257/aer.20130907

Bernheim, Douglas, and Antonio Rangel. 2009. “Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics”. *Quarterly Journal of Economics* 124:51-104.
https://doi.org/10.1162/qjec.2009.124.1.51

Bernheim, Douglas, and Dmitry Taubinsky. 2018. “Behavioral public economics”. In *Handbook of Behavioural Economics*, Vol.1, Douglas Bernheim, Stefano DellaVigna, and David Laibson (Eds.), 381-516. Amsterdam: Elsevier. https://doi.org/10.1016/bs.hesbe.2018.07.002

Beshears, John, James Choi, David Laibson, and Brigitte Madrian. 2008. “How Are Preferences Revealed?” *Journal of Public Economics* 92:1787-1794. https://doi.org/10.1016/j.jpubeco.2008.04.010

Bradley, Ben. 2007. “A paradox for some theories of welfare”. *Philosophical Studies* 133:45-53.
https://doi.org/10.1007/s11098-006-9005-8

Broome, John. 1993. “Can a Humean be moderate?”. In *Value, Welfare and Morality*, Raymond Frey and Christopher Morris (Eds.), 51-73. Cambridge: Cambridge University Press.
https://doi.org/10.1017/CBO9780511625022

Broome, John. 2013. *Rationality through Reasoning*. Oxford: Wiley-Blackwell.

Camerer, Colin, and George Loewenstein. 2004. “Behavioral economics: past, present, future”. In *Advances in Behavioral Economics*, Colin Camerer, George Loewenstein, and

Matthew Rabin (Eds.), 3-51. Princeton: Princeton University Press. https://doi.org/10.2307/j.ctvcm4j8j

Clark, Andrew, Sarah Flèche, Richard Layard, Nattavudh Powdthavee, and George Ward. 2018. *The Origins of Happiness: The Science of Well-Being over the Life Course*. Princeton: Princeton University Press.

Cowen, Tyler. 1993. "The Scope and Limits of Preference Sovereignty". *Economics & Philosophy* 9:253-269. https://doi.org/10.1017/S0266267100001553

Cozic, Mikaël, and Brian Hill. 2015. "Representation theorems and the semantics of decision-theoretic concepts". *Journal of Economic Methodology* 22:292–311. https://doi.org/10.1080/1350178X.2015.1071503

Crisp, Roger. 2006. *Reasons and the Good*. Oxford: Clarendon Press.

Dietrich, Franz, and Christian List. 2016a. "Reason-based choice and context dependence: an explanatory framework". *Economics & Philosophy* 32:175-229. https://doi.org/10.1017/S0266267115000474

Dietrich, Franz, and Christian List. 2016b. "Mentalism versus behaviourism in economics: a philosophy-of-science perspective". *Economics & Philosophy* 32:249-281. https://doi.org/10.1017/S0266267115000462

Dold, Malte. 2018. "Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics". *Journal of Economic Methodology* 2:160-178. https://doi.org/10.1080/1350178X.2017.1421770

Dold, Malte, and Christian Schubert. 2018. "Toward a behavioral foundation of normative economics". *Review of Behavioral Economics* 5:221–241. http://doi.org/10.1561/105.00000097

Dorsey, Dale. 2017. "Idealization and the Heart of Subjectivism". *Noûs* 51:196–217. https://doi.org/10.1111/nous.12130

Feldman, Fred. 1997. “On the intrinsic value of pleasures”. *Ethics* 107:448-466. https://doi.org/10.1086/233744

Ferreira, João. 2023. “Which choices merit deference? A comparison of three behavioural proxies of subjective welfare”. *Economics & Philosophy* 39:124-151. https://doi.org/10.1017/S0266267121000365

Fletcher, Guy. 2013. “A Fresh Start for the Objective-List Theory of Well-Being”. *Utilitas* 25:206-220. https://doi.org/10.1017/S0953820812000453

Fletcher, Guy. 2019. “Against Contextualism about Prudential Discourse”. *Philosophical Quarterly* 69:699-720. https://doi.org/10.1093/pq/pqz023

Fleurbaey, Marc. 2012. “The importance of what people care about”. *Politics, Philosophy & Economics* 11:415-447. https://doi.org/10.1177/1470594X12447775

Fleurbaey, Marc, and Francois Maniquet. 2011. *A Theory of Fairness and Social Welfare*. New York: Cambridge University Press.

Fumagalli, Roberto. 2019. “(F)utility Exposed”. *Philosophy of Science* 86:955-966. https://doi.org/10.1086/705454

Fumagalli, Roberto. 2020. “On the Individuation of Choice Options”. *Philosophy of the Social Sciences* 50:338-365. https://doi.org/10.1177/0048393120917759

Fumagalli, Roberto. 2021. “Theories of well-being and well-being policy”. *Journal of Economic Methodology* 28:124-133. https://doi.org/10.1080/1350178X.2020.1868780

Fumagalli, Roberto. 2022. “A reformed division of labor for the science of well-being”. *Philosophy* 97:509-543. https://doi.org/10.1017/S0031819122000092

Fumagalli, Roberto. 2024. “Preferences versus Opportunities: On the Conceptual Foundations of Normative Welfare Economics”. *Economics & Philosophy* 40:77-101. https://doi.org/10.1017/S0266267122000323

Fumagalli, Roberto. 2025. “Standard Rationality versus Inclusive Rationality: A Critical Assessment”. *Behavioural Public Policy*. In Press.

Glimcher, Paul. 2011. *Foundations of Neuroeconomic Analysis*. New York: Oxford University Press.

Griffin, James. 1986. *Well-Being: Its Measure and Importance*. Oxford: Clarendon Press.

Grill, Kalle. 2015. “Respect for What? Choices, Actual Preferences, and True Preferences”. *Social Theory and Practice* 41:692-715. https://doi.org/10.5840/soctheorpract201541437

Guala, Francesco. 2012. “Are Preferences for Real?”. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, Aki Lehtinen, Jaakko Kuorikoski, and Petri Ylikoski, (Eds.), 137-155. New York: Routledge.

Guala, Francesco. 2019. “Preferences: neither behavioural nor mental”. *Economics & Philosophy* 35:383-401. https://doi.org/10.1017/S0266267118000512

Gul, Faruk, and Wolfgang Pesendorfer. 2008. “The case for mindless economics”. In *The Foundations of Positive and Normative Economics: A Handbook*, Andrew Caplin and Andrew Schotter (Eds.), 3-42. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328318.003.0001

Gustafsson, Johan. 2022. *Money-Pump Arguments*. Cambridge: Cambridge University Press.

Hands, Wade. 2014. “Normative ecological rationality”. *Journal of Economic Methodology* 21:396-410. https://doi.org/10.1080/1350178X.2014.965907

Harsanyi, John. 1982. “Morality and the theory of rational behaviour”. In *Utilitarianism and*

Beyond, Amartya Sen and Bernard Williams (Eds.), 39-62. Cambridge: Cambridge University Press.
https://doi.org/10.1017/CBO9780511611964

Hausman, Daniel. 1998. "Problems with Realism in Economics". *Economics & Philosophy* 14:185-213.
https://doi.org/10.1017/S0266267100003837

Hausman, Daniel. 2010. "Hedonism and welfare economics". *Economics & Philosophy* 26:321-344.
https://doi.org/10.1017/S0266267110000398

Hausman, Daniel. 2011. "Mistakes about preferences in the social sciences". *Philosophy of the Social Sciences* 41:3-25. https://doi.org/10.1177/0048393110387885

Hausman, Daniel. 2012. *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.

Hausman, Daniel. 2016. "On the Econ within". *Journal of Economic Methodology* 23:26-32.
https://doi.org/10.1080/1350178X.2015.1070525

Hausman, Daniel. 2022a. "Enhancing welfare without a theory of welfare". *Behavioural Public Policy* 6:342-357.
https://doi.org/10.1017/bpp.2019.34

Hausman, Daniel. 2022b. "Banishing the inner Econ and justifying paternalistic nudges". *Behavioural Public Policy* 1-12. https://doi.org/10.1017/bpp.2022.19

Hausman, Daniel. 2024. "Subjective total comparative evaluations". *Economics & Philosophy* 40:212-225.
https://doi.org/10.1017/S0266267122000311

Hausman, Daniel, and Michael McPherson. 2009. "Preference satisfaction and welfare economics". *Economics & Philosophy* 25:1-25. https://doi.org/10.1017/S0266267108002253

Hawkins, Jennifer. 2019. “Internalism and Prudential Value”. In *Oxford Studies in Metaethics*, Vol.14, Russ Shafer-Landau (Ed.), 95-120. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198841449.003.0005

Haybron, Daniel, and Valerie Tiberius. 2015. “Well-being policy: what standard of well-being?” *Journal of American Philosophical Association* 1:712-733. https://doi.org/10.1017/apa.2015.23

Heathwood, Chris. 2005. “The problem of defective desires”. *Australasian Journal of Philosophy* 83:487-504. https://doi.org/10.1080/00048400500338690

Heathwood, Chris. 2011. “Preferentism and Self-Sacrifice”. *Pacific Philosophical Quarterly* 92:18-38. https://doi.org/10.1111/j.1468-0114.2010.01384.x

Heathwood, Chris. 2014. “Subjective theories of well-being”. In *The Cambridge Companion to Utilitarianism*, Ben Eggleston and Dale Miller (Eds.), 199-219. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCO9781139096737.011

Heathwood, Chris. 2016. “Desire-Fulfillment Theory”. In *The Routledge Handbook of Philosophy of Well-Being*, Guy Fletcher (Ed.), 135–147. New York: Routledge. https://doi.org/10.4324/9781315682266

Heathwood, Chris. 2019. “Which Desires Are Relevant to Well-Being?” *Noûs* 53:664-688. https://doi.org/10.1111/nous.12232

Hersch, Gil. 2015. “Can an evidential account justify relying on preferences for well-being policy?” *Journal of Economic Methodology* 22:280-291. https://doi.org/10.1080/1350178X.2015.1071507

Hersch, Gil. 2020. “No theory-free lunches in well-being policy”. *Philosophical Quarterly* 70:43-64. https://doi.org/10.1093/pq/pqz029

Hurka, Thomas. 2019. “On ‘Hybrid’ Theories of Personal Good”. *Utilitas* 31:450-462. https://doi.org/10.1017/S0953820819000256

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016a. “Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics”. *Journal of Economic Methodology* 23:1-25. https://doi.org/10.1080/1350178X.2015.1070527

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016b. “On the Econ within: a reply to Daniel Hausman”. *Journal of Economic Methodology* 23:33-37. https://doi.org/10.1080/1350178X.2015.1070526

Kagan, Shelly. 1992. “The limits of well-being”. *Social Philosophy & Policy* 9:169-189. https://doi.org/10.1017/S0265052500001461

Kagan, Shelly. 2009. “Well-Being as Enjoying the Good”. *Philosophical Perspectives* 23:253–272. https://doi.org/10.1111/j.1520-8583.2009.00170.x

Kahneman, Daniel, Peter Wakker, and Rakesh Sarin. 1997. “Back to Bentham? Explorations of experienced utility”. *Quarterly Journal of Economics* 112:375-406. https://doi.org/10.1162/003355397555235

Keller, Simon. 2009. “Welfare as Success”. *Noûs* 43:656–683. https://doi.org/10.1111/j.1468-0068.2009.00723.x

Kelman, Mark. 2005. “Hedonic psychology and the ambiguities of ‘welfare’”. *Philosophy & Public Affairs* 33:391–412. https://doi.org/10.1111/j.1088-4963.2005.00038.x

Kraut, Richard. 1994. “Desire and the human good”. *Proceedings and Addresses of the American Philosophical Association* 68:39-54. https://doi.org/10.2307/3130590

Kraut, Richard. 2007. *What is Good and Why*. Cambridge, MA: Harvard University Press.

Lecouteux, Guilhem. 2022. “Reconciling normative and behavioural economics: the problem that cannot be solved”. In *The Positive and the Normative in Economic Thought*, Sina Badiei and Agnès Grivaux (Eds.), ch.8. London: Routledge. https://doi.org/10.4324/9781003247289

Levy, Dino, and Paul Glimcher. 2012. “The Root of All Value”. *Current Opinion in Neurobiology* 22:1027-1038. https://doi.org/10.1016/j.conb.2012.06.001

Lin, Eden. 2017. “Against welfare subjectivism”. *Noûs* 51:354-377. https://doi.org/10.1111/nous.12131

Loeb, Don. 1995. “Full-Information Theories of Individual Good”. *Social Theory and Practice* 21:1-30. https://doi.org/10.5840/soctheorpract199521116

Loewenstein, George, and Emily Haisley. 2008. “The economist as therapist”. In *The Foundations of Positive and Normative Economics: A Handbook*, Andrew Caplin and Andrew Schotter (Eds.), 210–245. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328318.003.0009

Mäki, Uskali. 1998. “Aspects of realism about economics”. *Theoria* 13:301-319.

Mandler, Michael. 2005. “Incomplete preferences and rational intransitivity of choice”. *Games and Economic Behavior* 50:255–277. https://doi.org/10.1016/j.geb.2004.02.007

Manzini, Paola, and Marco Mariotti. 2014. “Welfare economics and bounded rationality: the case for model-based approaches”. *Journal of Economic Methodology* 21:343-360. https://doi.org/10.1080/1350178X.2014.965909

Mas-Colell, Andreu, Michael Whinston, and Jerry Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Matson, Erik. 2022. “Our dynamic being within: Smithian challenges to the new paternalism”. *Journal of Economic Methodology* 29:309-325. https://doi.org/10.1080/1350178X.2022.2145338

McQuillin, Ben, and Robert Sugden. 2012. “Reconciling normative and behavioural economics: the problems to be solved”. *Social Choice and Welfare* 38:553–567. https://doi.org/10.1007/s00355-011-0627-1

Moscati, Ivan. 2021. “On the recent philosophy of decision theory”. *Journal of Economic Methodology* 28:98-106. https://doi.org/10.1080/1350178X.2020.1868777

Nielsen, Kirby, and John Rehbeck, J. 2022. “When Choices Are Mistakes”. *American Economic Review* 112:2237–2268. https://doi.org/10.1257/aer.20201550

Nishimura, Hiroki. 2018. “The transitive core: inference of welfare from nontransitive preference relations”. *Theoretical Economics* 13:579-606. https://doi.org/10.3982/TE1769

Nussbaum, Martha. 2000. *Women and Human Development.* New York: Cambridge University Press.

Okasha, Samir. 2016. “On the Interpretation of Decision Theory”. *Economics & Philosophy* 32:1-25. https://doi.org/10.1017/S0266267115000346

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pettigrew, Richard. 2023. “Nudging for changing selves”. *Synthese*. In Press. https://doi.org/10.1007/s11229-022-04020-2

Rabinowicz, Wlodek, and Jan Österberg. 1996. “On Two Interpretations of Preference Utilitarianism”. *Economics & Philosophy* 12:1-27. https://doi.org/10.1017/S0266267100003692

Railton, Peter. 1986. “Facts and values”. *Philosophical Topics* 24:5-31. https://doi.org/10.5840/philtopics19861421

Rizzo, Mario. 2025. “The problem of counterfactual preferences”. *Social Philosophy & Policy*. In Press.

Rizzo, Mario, and Douglas Whitman. 2020. *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*. Cambridge: Cambridge University Press.

Rosati, Connie. 1995. “Persons, perspectives, and full information accounts of the good”. *Ethics* 105:296-325. https://doi.org/10.1086/293702

Rosati, Connie. 1996. “Internalism and the good for a person”. *Ethics* 106:297-326. https://doi.org/10.1086/233619

Rosati, Connie. 2006. “Preference-Formation and Personal Good”. *Royal Institute of Philosophy Supplements* 59:33-64. https://doi.org/10.1017/S1358246106059030

Rosati, Connie. 2009. “Self-Interest and Self-Sacrifice”. *Proceedings of the Aristotelian Society* 109:311-325. https://doi.org/10.1111/j.1467-9264.2009.00269.x

Ross, Don. 2011. “Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics”. *Journal of Economic Methodology* 18:217-231. https://doi.org/10.1080/1350178X.2011.611024

Ross, Don. 2012. “Mäki’s Realism and the Scope of Economics”. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, Aki Lehtinen, Jaakko Kuorikoski, and Petri Ylikoski (Eds.), 181-202. New York: Routledge.

Ross, Don. 2014. “Psychological versus economic models of bounded rationality”. *Journal of Economic Methodology* 21:411-427. https://doi.org/10.1080/1350178X.2014.965910

Rubinstein, Ariel, and Yuval Salant. 2008. “Some Thoughts on the Principle of Revealed Preference”. In *The Foundations of Positive and Normative Economics: A Handbook*, Andrew Caplin and Andrew Schotter (Eds.), 116-124. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328318.003.0005

Rubinstein, Ariel, and Yuval Salant. 2012. “Eliciting welfare preferences from behavioral datasets”. *Review of Economic Studies* 79:375-387. https://doi.org/10.1093/restud/rdr024

Salant, Yuval, and Ariel Rubinstein. 2008. “(A, f): choice with frames”. *Review of Economic Studies* 75:1287-1296. https://doi.org/10.1111/j.1467-937X.2008.00510.x

Sarch, Alexander. 2011. “Internalism about a Person’s Good: Don’t Believe It”. *Philosophical Studies* 154:161–184.
https://doi.org/10.1007/s11098-010-9533-0

Sarch, Alexander. 2015. “Hausman and McPherson on welfare economics and preference satisfaction theories of welfare: a critical note”. *Economics & Philosophy* 31:141-159. https://doi.org/10.1017/S0266267114000431

Scanlon, Thomas. 1996. “The status of well-being”. *Tanner Lectures on Human Values* 16:91-143.

Scanlon, Thomas. 1998. *What we Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schulz, Armin. 2024. “Phylogenetic economics: animal models and the study of choice”. *Philosophy of Science* 91:811-830. https://doi.org/10.1017/psa.2024.4

Sen, Amartya. 1973. “Behaviour and the Concept of Preference”. *Economica* 40:241-259. https://doi.org/10.2307/2552796

Sen, Amartya. 1977. “Rational fools: a critique of the behavioral foundations of economic theory”. *Philosophy & Public Affairs* 6:317-344.

Sobel, David. 1994. “Full information accounts of well-being”. *Ethics* 104:784-810. https://doi.org/10.1086/293655

Sobel, David. 1998. “Well-Being as the Object of Moral Consideration”. *Economics & Philosophy* 14:249–281. https://doi.org/10.1017/S0266267100003850

Sobel, David. 2001. “Explanation, Internalism, and Reasons for Action”. *Social Philosophy and Policy* 18:218-235. https://doi.org/10.1017/S026505250000296X

Sobel, David. 2009. “Subjectivism and idealization”. *Ethics* 119:336-352. https://doi.org/10.1086/596459

Sobel, David, and Steven Wall. 2023. “The Objectivist Attempt to Appropriate Subjective Value”. In *Oxford Studies in Metaethics*, Vol.18, Russ Shafer-Landau (Ed.), 1-23. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198884699.003.0001

Stoljar, Natalie. 2014. “Autonomy and adaptive preference formation”. In *Autonomy, Oppression, and Gender*, Andrea Veltman and Mark Piper (Eds.), 227–252. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199969104.003.0011

Sugden, Robert. 1991. “Rational choice: a survey of contributions from economics and philosophy”. *The Economic Journal* 101:751-785. https://doi.org/10.2307/2233854

Sugden, Robert. 2008. “Why incoherent preferences do not justify paternalism”. *Constitutional Political Economy* 19:226-248. https://doi.org/10.1007/s10602-008-9043-7

Sugden, Robert. 2018. *The Community of Advantage*. Oxford: Oxford University Press.

Sumner, Leonard. 1996. *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.

Sunstein, Cass. 2015. “Nudges, Agency, and Abstraction”. *Review of Philosophy and Psychology* 6:511–529. https://psycnet.apa.org/doi/10.1007/s13164-015-0266-z

Sunstein, Cass, and Richard Thaler. 2003. “Libertarian paternalism is not an oxymoron”. *University of Chicago Law Review* 70:1159-1202. https://doi.org/10.2307/1600573

Thoma, Johanna. 2021a. “On the possibility of an anti-paternalist behavioural welfare economics”. *Journal of Economic Methodology* 28:350-363. https://doi.org/10.1080/1350178X.2021.1972128

Thoma, Johanna. 2021b. “In defence of revealed preference theory”. *Economics & Philosophy* 37:163-187. https://doi.org/10.1017/S0266267120000073

Van der Deijl, Willem. 2017. “Are Measures of Well-Being Philosophically Adequate?” *Philosophy of the Social Sciences* 47:209-234. https://doi.org/10.1177/0048393116683249

Vredenburgh, Kate. 2021. “The economic concept of a preference”. In *The Routledge Handbook of the Philosophy of Economics*, Conrad Heilmann and Julian Reiss (Eds.), ch.5. New York: Routledge. https://doi.org/10.4324/9781315739793

Vredenburgh, Kate. 2024. “Causal Explanation and Revealed Preferences”. *Philosophy of Science* 91:269-287. https://doi.org/10.1017/psa.2023.112

Vromen, Jack. 2022. “Does the inclusion of social preferences in economic models challenge the positive - normative distinction?”. In *Methodology and History of Economics*, Bruce Caldwell, John Davis, Uskali Mäki, and Esther-Mirjam Sent (Eds.), ch.11. London: Routledge. https://doi.org/10.4324/9781003266051

Wall, Steven, and David Sobel. 2021. “A robust hybrid theory of well-being”. *Philosophical Studies* 178:2829–2851. https://doi.org/10.1007/s11098-020-01586-w

Whitman, Douglas, and Mario Rizzo. 2015. “The problematic welfare standards of behavioral paternalism”. *Review of Philosophy and Psychology* 6:409-425. https://doi.org/10.1007/s13164-015-0244-5

Williamson, Timothy Luke. 2024. “A risky challenge for intransitive preferences”. *Noûs* 58:360–385. https://doi.org/10.1111/nous.12455

Woodard, Christopher. 2013. “Classifying Theories of Welfare”. *Philosophical Studies* 165:787–803. https://doi.org/10.1007/s11098-012-9978-4