Behavioural Public Policy (2025), **9**, 530–544
doi:10.1017/bpp.2024.53

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Standard rationality versus inclusive rationality: a critical assessment

Roberto Fumagalli

Department of Political Economy, King's College London, London, UK
Email: roberto.fumagalli@kcl.ac.uk

(Received 9 February 2024; Revised 16 October 2024; Accepted 22 October 2024)

Abstract

This paper critically assesses Rizzo and Whitman's theory of inclusive rationality in light of the ongoing cross-disciplinary debate about rationality, welfare analyses and policy evaluation. The paper aims to provide three main contributions to this debate. First, it explicates the relation between the consistency conditions presupposed by standard axiomatic conceptions of rationality and the standards of rationality presupposed by Rizzo and Whitman's theory of inclusive rationality. Second, it provides a qualified defence of the consistency conditions presupposed by standard axiomatic conceptions of rationality against the main criticisms put forward by Rizzo and Whitman. And third, it identifies and discusses specific strengths and weaknesses of Rizzo and Whitman's theory of inclusive rationality in the context of welfare analyses and policy evaluation.

Keywords: choices; consistency; policy evaluation; preferences; rationality; welfare analyses

Introduction

In a series of influential works, Rizzo and Whitman (henceforth, RW) put forward several theoretical and practical challenges to behavioural paternalists' attempts to design and implement welfare-enhancing paternalistic interventions in public policy (see, e.g., RW, 2009a, 2009b, 2018, 2020a and 2023; also WR, 2015 and 2021). RW's works provide a systematic critique of behavioural paternalism's conceptual foundations and of behavioural paternalists' attempts to enhance individuals' welfare. I lack the space here to engage with the many interesting themes examined in RW's works. In this short paper, I focus on one foundational theme that figures centrally in such works, namely the theory of *inclusive rationality* (henceforth, IR) that RW put forward 'as *an alternative*' to the standard axiomatic conceptions of rationality (henceforth, SR) 'shared by neoclassical and behavioral economics' (WR, 2021, 382, italics added; also RW, 2018, 217). According to RW, SR is 'excessively narrow [and] cannot capture the full depth

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

and complexity of human choice' (2020a, xi and 433). For its part, IR 'does not dictate the normative structure of preferences' and encompasses a wide range of choice patterns that 'do not fit nicely into the straitjacket of [SR]' (ibid., 17 and 26).[1]

The two main sections of this paper are structured as follows. The next section ('Standard rationality versus inclusive rationality') outlines the *consistency conditions* presupposed by SR and explicates the *alleged contrast* between RW's theory of IR and SR. The third section ('Inclusive rationality: a critical assessment') articulates and defends a *critical assessment* of RW's theory of IR. I shall argue for four main claims concerning such theory. First, the consistency conditions presupposed by SR can be defended against the main criticisms put forward by RW. Second, the proponents of SR can incorporate several insights provided by RW's theory of IR without having to relinquish their reliance on SR's consistency conditions. Third, RW's theory of IR faces substantial falsifiability concerns, which seem more widespread and pervasive than those faced by SR. And fourth, RW's theory of IR does not ground more informative and reliable evaluations of public policies' welfare implications than SR. These four claims do not detract from the many merits of RW's works. In particular, they do not bear against the main criticisms that RW articulate against behavioural paternalists' attempts to design and implement welfare-enhancing paternalistic interventions in public policy.[2] In this perspective, my critical assessment of RW's theory of IR can be seen as a constructive contribution to the ongoing discussion about RW's theory of IR (see, e.g., Cowen and Dold, 2021, on a dedicated special issue) and the broader cross-disciplinary debate about rationality, welfare analyses and policy evaluation (see, e.g., Oliver, 2023, for recent contributions to such debate).

Standard rationality versus inclusive rationality

SR explicates the notion of rationality in terms of specific *structural conditions* on preferences. More specifically, to qualify as SR-rational, an agent's preferences 'must satisfy [the axioms of] completeness and transitivity, as well as certain corollaries [such as] independence of irrelevant alternatives' (RW, 2020a, 16; also 80).[3] These axiomatic conditions constrain sequences of preferences, taken collectively, but place no substantive constraints on preferences, taken individually (see, e.g., Broome, 1993, 52; Sugden,

[1]RW occasionally characterize IR as a 'notion' and a 'research programme' rather than a specific 'theory' (see, e.g., RW, 2020b, WR, 2021, 385). I expand on these other characterizations of IR in the following sections. Also, RW (2020a) frequently use the expressions 'neoclassical rationality' and 'puppet rationality' to refer to SR. I stick to 'SR' for terminological clarity (see, e.g., Colander, 2000, on the heterogeneous senses ascribed to 'neoclassical rationality' in the economic literature).

[2]I endorsed many of RW's criticisms of behavioural paternalism in previous articles (see, e.g., Fumagalli, 2016a, on the knowledge problems that hamper behavioural paternalists' attempts to design and implement welfare-enhancing paternalistic interventions; Fumagalli, 2020a, on the risk that implementing moderate and seemingly justifiable paternalistic interventions leads policy makers to implement morally problematic or otherwise objectionable paternalistic interventions via slippery slope mechanisms).

[3]Completeness requires that, for any two options *A* and *B* in the choice set, the agent has definite preferences regarding such options. Transitivity requires that, for any options *A, B* and *C* in the choice set, if the agent prefers *A* to *B* and *B* to *C*, then the agent prefers *A* to *C*. Analogous formulations of these structural conditions can be provided for choices rather than preferences (see, e.g., Bhattacharyya *et al.*, 2011, 143).

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

1991, 760). As noted by RW, such axiomatic conditions 'provide a logical foundation for […] the existence of utility functions [and make] economic models mathematically tractable' (RW, 2020a, 53 and 81; also WR, 2015, 409 and 416). The idea, encapsulated in so-called *representation theorems*, is that if an agent's preferences satisfy specific axiomatic requirements, then this agent's choices can be represented as if the agent maximizes expected utility (see, e.g., Von Neumann and Morgenstern, 1947, on situations of risk; Savage, 1954, on situations of uncertainty). SR models do not aim to provide accurate characterizations of the neuro-psychological substrates of choices and are typically agnostic about such substrates (see, e.g., Fumagalli, 2013). In particular, the preferences figuring in representation theorems are commonly regarded as indexes of choices, but SR does not commit choice modellers to regarding preferences in general as reducible to or identical with choices (see, e.g., Beck, 2024, on different conceptions of preferences).[4]

RW's theory of IR draws on interrelated descriptive, normative and prudential criticisms of SR. These criticisms can be explicated as follows. The consistency conditions presupposed by SR occasionally 'provide a reasonable approximation of how people really behave' (WR, 2021, 382). However, systematic *descriptive violations* of these conditions have been observed (see, e.g., RW, 2020a, ch. 1, WR, 2015). Moreover, an individual's preferences may 'violate the axioms of completeness and transitivity [without the individual being] irrational in any *normatively significant* sense' (RW, 2020a, 75, italics added; also WR, 2015, 420). For one may be 'discovering [or forming her] preferences […] during the process of choice' (RW, 2020a, 58 and 81; also WR, 2015, 418). And in many cases, 'the costs of completely rationalizing [one's] preferences exceed the benefits of doing so' (RW, 2020a, 81; also RW, 2018, 202). Therefore, an inclusively rational individual '*will not*, and *should not*, have complete and transitive preferences' (WR, 2015, 419, italics added; also RW, 2020a, 239, WR, 2021, 383).[5]

As to *prudential* considerations, RW hold that SR builds on consistency as the main *welfare criterion* and assumes that if individuals' preferences violate completeness and transitivity, then such preferences fail to reliably track individuals' welfare (see, e.g., RW, 2020a, ch. 6–7, WR, 2015). However, in their view, SR's consistency conditions provide 'no basis at all' for determining which preferences track welfare (RW, 2020a, 18). For individuals 'may have mutable preferences […] or no relevant preferences' (ibid., 28). And individuals' inconsistencies 'can typically be resolved in more than one way' (RW, 2023, 202; also WR, 2015, on the difficulties inherent in identifying welfare-optimal rates of saving and intertemporal discounting in specific choice settings). Moreover, abiding by SR's prescriptions does not guarantee individuals to make

[4]Some SR models do rest on empirical assumptions concerning the neuro-psychological substrates of choices (see, e.g., Glimcher, 2011, ch. 6–8, on 'hard' expected utility theory). Still, nothing in SR requires or implies that SR models accurately (or even approximately) represent such substrates (see, e.g., Ross, 2014). In fact, many SR models do not rest on any empirical assumption about either neural or psychological substrates (see, e.g., Fumagalli, 2020b).

[5]RW (2020a, ch. 5) also target the rationality of beliefs and hold that, contrary to what many choice modellers presuppose, following Bayes' rule does not constitute the uniquely reasonable way to form/update beliefs. I mention this issue in passing since my evaluation primarily concerns preferences rather than beliefs. For a critical evaluation of RW's claims about the rationality of beliefs, see, e.g., Grüne-Yanoff, 2021, 294–295.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

welfare-optimal choices. For choices that violate SR's axioms may be 'adaptive to the circumstances [and] can increase the agent's welfare' (Rizzo, 2018, 193).[6]

RW's theory of IR draws on these interrelated descriptive, normative and prudential criticisms of SR to provide '*an alternative*' to SR (WR, 2021, 382, italics added; also RW, 2018, 217). I shall expand on the specific tenets of RW's theory of IR in the next section. For now, I note that contrary to SR, IR 'does not dictate the normative structure of preferences' and encompasses a wide range of choice patterns that 'do not fit nicely into the straitjacket of [SR]' (RW, 2020a, 17 and 26). In particular, IR allows that the set of 'rational' preferences may include preferences that are 'inchoate, incomplete, inconsistent, mutable, and dependent on context' (ibid., 26; also RW, 2018, 205).

Inclusive rationality: a critical assessment

In this section, I articulate and defend a critical assessment of RW's theory of IR. I shall argue for four main claims concerning such theory, which respectively concern: the defensibility of the *consistency conditions* presupposed by SR against the main criticisms put forward by RW; the possibility of *incorporating* into SR several insights provided by RW's theory of IR without having to relinquish SR's reliance on consistency conditions; the *falsifiability concerns* faced by RW's theory of IR; and the applicability of RW's theory of IR to evaluating public policies' *welfare implications*.

SR's consistency conditions

RW correctly note that SR's consistency conditions are violated in several choice settings (see, e.g., Gilboa *et al.*, 2009, on violations of completeness; Sugden, 1991, on violations of transitivity) and that various authors challenge the normative validity of such conditions (see, e.g., Aumann, 1962, on completeness; Anand, 1993, on transitivity). However, SR's consistency conditions can be defended against the main criticisms put forward by RW. Below I provide three replies to RW's descriptive and normative criticisms.[7]

First, the reported *descriptive violations* of SR's consistency conditions tend to *significantly decrease* in presence of experienced decision makers (see, e.g., List, 2003; Choi *et al.*, 2014) and in situations where individuals are given time and incentives to learn about the choice problems they face (see, e.g., Plott and Smith, 2008; Oprea, 2020). Moreover, in recent decades the proponents of SR have developed several SR models which *modify* specific axioms so as to increase SR's descriptive fit with the available empirical findings (see, e.g., Machina, 2008, and Starmer, 2000, for reviews). In

[6]RW (2020a, ch. 2) explicitly draw on the notion of ecological rationality, which posits that individuals frequently rely on fast and frugal heuristics and regards heuristics as ecologically rational 'to the degree that [they are] adapted to the structure of [individuals'] environment' (Gigerenzer, 2021, 3548). However, IR differs from ecological rationality in various respects (see, e.g., RW, 2020a, 27, holding that their arguments for IR 'draw heavily on [Gigerenzer's] notion of ecological rationality, but […] do not limit [themselves] to it'). I shall comment briefly on the notion of ecological rationality in the next section. For a critical comparison of different approaches to ecological rationality, see, e.g., Dekker and Remic, 2019.

[7]I expand on RW's prudential criticisms in sub-section *Welfare analyses*. My evaluation focuses on completeness and transitivity (rather than other axioms) since RW primarily target these axioms.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

this respect, it would be of limited import to object that the models involving such modifications are more plausibly regarded as IR (rather than SR) models (see, e.g., RW, 2020a, 31, holding that 'when behavioral economists invoke bounded rationality, they are in essence claiming that the bounds of [SR] models are not appropriate'). For despite modifying specific axioms, such models retain SR's reliance on axiomatic consistency conditions on preferences. That is to say, RW are correct that early SR models (e.g., expected utility theory) face descriptive criticisms and that some SR models merely accommodate (rather than predict) individuals' choices. Still, several modelling developments have occurred within SR over the last few decades, and various such developments are plausibly regarded as empirically progressive (see, e.g., Guala, 2005, and Starmer, 2005, for illustrations).[8]

Second, SR's consistency conditions can be given a plausible *normative defence*. For instance, completeness is less demanding than what RW appear to presuppose, since it requires individuals to be able to specify their preferences over the alternatives figuring in the examined decision problems rather than over all possible alternatives (see, e.g., Grüne-Yanoff, 2021, 291; Gustafsson, 2022, sec. 3). And transitivity can be defended by pointing to the losses that individuals tend to incur by violating it (see, e.g., Grüne-Yanoff, 2021, 294–295; Gustafsson, 2022, sec. 4, on monetary and other welfare-relevant losses) and to individuals' willingness to revise intransitive choices when they realize these choices' intransitivity (see, e.g., Hands, 2014, 401–402; Nielsen and Rehbeck, 2022, 2237–2239).[9] To be sure, the normative plausibility of completeness and transitivity may vary depending on what conception of preferences one presupposes (see, e.g., Mandler, 2005, 255–256, holding that completeness is more easily defended for individuals' behavioural preferences than for preferences that encapsulate individuals' judgements about their own welfare) and what SR models one examines (see, e.g., Sugden, 1991, 763, holding that the restrictions Savage's theory imposes on what factors can be included in the description of a consequence hamper the defensibility of transitivity within such theory). Still, these dependencies do not generally bear against the normative plausibility of SR's axioms. In this respect, it is telling that many of those who doubt the descriptive validity of SR's axioms retain those axioms as a normative standard (see, e.g., Angner, 2019, 203, on leading behavioural economists; RW, 2020a, 40, on leading behavioural paternalists).[10]

[8] In the recent literature, various authors debate as to whether the ongoing integration of empirical findings into SR models is predominantly neoclassical or behavioural in character (see, e.g., Chetty, 2015, 1, holding that 'behavioral economics represents a natural progression of […] neoclassical economic methods', versus Angner, 2019, 195, holding that 'the proposed synthesis represents the consummate conversion of neoclassical economists into behavioral ones'). I do not aim here to defend a specific position in this debate. For my purposes, it suffices to note that the consistency conditions presupposed by SR can be defended against the main criticisms put forward by the proponents of IR.

[9] Acting on intransitive preferences may not lead to losses in the absence of 'bookmakers ready to take advantage' (RW, 2020a, 68; also WR, 2015, 419–420). However, this falls short of implying that 'transitivity is irrelevant […] from a pragmatic perspective' (Rizzo, 2019, 84). For as noted in the main text, individuals who act on intransitive preferences often tend to obtain lower payoffs than they would obtain if they acted on transitive preferences. And these payoff differences can have great pragmatic relevance (see, e.g., sub-section *Welfare analyses* for discussion targeting individuals' welfare).

[10] This is not to suggest that choice modellers should stipulate that rationality generally consists in invariably abiding by SR's axioms. In fact, this stipulation would seemingly trivialize the debate about the normative

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

And third, many reported descriptive and normative violations of SR's axioms can be accommodated by modifying the *descriptions of choice options* presupposed by the purported counterexamples to such axioms (see, e.g., Broome, 1993; also Fumagalli, 2020c, for critical discussion). By way of illustration, consider an individual who, faced with pairwise comparisons between food items A (apple), B (banana) and C (cake), exhibits the intransitive preference pattern A $>$ B, B $>$ *C* and C $>$ A. One may accommodate this violation of transitivity by incorporating reference to *what options* are available to the individual into the description of each choice option. More specifically, let Ab indicate A when B is the other option available, Ba indicate B when A is the other option available, and so on. The individual's preference pattern can then be re-described as Ab $>$ Ba, Bc $>$ Cb and Ca $>$ Ac, which does not directly violate transitivity (see, e.g., Broome, 1993, 54). To be sure, RW are correct that 'allowing redescription whenever we encounter […] violations of the axioms' would render SR's axioms descriptively and/or normatively empty (2020a, 70; also Anand, 1993, 103; Bhattacharyya *et al.*, 2011, 146). However, the proffered re-descriptions of choice options are not *equally plausible*, and choice modellers are frequently able to *demarcate* whether specific factors (e.g., the price, spatial location and caloric content of specific food items) can be justifiably incorporated into the description of choice options by determining whether those factors do (or can plausibly) make a difference to individuals' preferences between such options (see, e.g., Dreier, 1996, and Fumagalli, 2020c, for illustrations).[11]

Compatibility of IR and SR

According to RW, the set of inclusively rational preferences may include preferences that are 'inchoate, incomplete, inconsistent, mutable, and dependent on context' (RW, 2020a, 26; also RW, 2018, 205). In their view, 'a good case can be made for [the] inclusive rationality' of several preference patterns that the proponents of SR regard as 'biases' (RW, 2020a, 17). However, SR appears to be more 'inclusive' than RW allege. For SR has the resources to accommodate a wide range of preference patterns that the proponents of IR deem to be incompatible with SR. In particular, the proponents of SR can incorporate several insights provided by RW's theory of IR without having to relinquish their

plausibility of SR's axioms by preventing choice modellers from distinguishing cases where individuals choose irrationally from cases where individuals do not abide by SR's axioms (see, e.g., Guala, 2000, 69).

[11] A proponent of IR may object that distinct choice modellers endorse dissimilar views as to which factors can be justifiably incorporated into the description of choice options and that it is up to the proponents of SR to provide a 'general answer to the question of how [choice options] should be described' (RW, 2020a, 70). However, it would be overly demanding to require the proponents of SR to provide such a general answer. For what set of factors can be justifiably incorporated into the description of choice options plausibly depends on a wide range of contextual elements (e.g., whether modellers have descriptive or normative purposes; what cognitive/computational abilities are possessed by the modelled agents). Moreover, adopting IR would not enable choice modellers to provide a general answer to the question of how choice options should be described. In fact, adopting IR may hamper choice modellers' ability to provide such answer. For IR seemingly presupposes that agents can rationally have any pattern of preferences (see, e.g., Rizzo, 2018, 208–211, on the purported inclusive rationality of several instances of wishful thinking). And this view imposes less informative constraints on how choice options should be described than the view (implicit in SR) which regards inconsistent patterns of preferences as irrational (see, e.g., Bradley, 2017, ch. 14; Dreier, 1996).

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

reliance on SR's consistency conditions. To illustrate this, consider three putative inclusively rational 'biases' examined by RW, namely preference change, framing effects and self-regulation.[12]

Preference change. Individuals' preferences frequently change across time and choice settings (see, e.g., RW, 2020a, ch. 3). According to RW, preference change is incompatible with SR since SR models 'typically [assume that] the agent has preferences that remain the same over time' (ibid., 78). Moreover, RW hold, many instances of preference change are inclusively rational (ibid., ch. 3). However, the proponents of SR have developed several models of preference change (see, e.g., Dietrich and List, 2011, and Strohmaier and Messerli, 2024, for reviews) and are not committed to regarding preference change as irrational. To illustrate this, consider the so-called endowment effect. According to RW, the endowment effect 'is not consistent with [SR] if switching costs are low and the value of the [involved goods] is small relative to the chooser's wealth' (2020a, 13). Moreover, RW hold, many instances of the endowment effect are plausibly regarded as inclusively rational since 'possession or ownership [of a good] may reflect important human values' (ibid., 110). However, the proponents of SR may accommodate such instances of the endowment effect. For if possession or ownership of a good reflects individuals' values, then SR allows choice modellers to include reference to these values (and those values' influence on individuals' preferences) into the description of choice problems (see, e.g., Fumagalli, 2020c). To be sure, one may point to various cases where choice modellers lack reliable epistemic access to individuals' values (see, e.g., RW, 2020a, ch. 6–7). Yet, the existence of such cases does not selectively support IR over SR. For choice modellers frequently have reliable epistemic access to individuals' values (see, e.g., Bradley, 2017, ch. 1). And in cases where choice modellers lack reliable epistemic access to individuals' values, adopting IR does not *per se* yield more informative descriptive and/or normative insights about such values compared to SR.[13]

Framing effects. Individuals' preferences are frequently sensitive to framing effects (see, e.g., WR, 2015). According to RW, the sensitivity of individuals' preferences to framing effects is incompatible with SR, but is often plausibly regarded as inclusively rational (see, e.g., RW, 2020a, ch. 1). However, the proponents of SR are not committed to regarding the sensitivity of individuals' preferences to framing effects as irrational.

[12] A proponent of IR may object that RW's theory of IR does not aim to entirely displace SR and that RW acknowledge that SR models can 'serve a useful [descriptive] function' (WR, 2021, 382; also RW, 2020a, 35). However, as illustrated in the previous sections, RW explicitly present IR as 'an alternative' to SR and repeatedly juxtapose IR and SR on both descriptive and normative grounds (see also Cowen and Dold, 2021, 216, on RW's, 2020a, ch. 10, recommendation to 'replace' SR with IR). The illustrations provided in this section can be seen as a response to such juxtapositions.

[13] A proponent of IR may object that SR models positing preference change are empirically underconstrained since 'virtually any change in behavior can be rationalized as resulting from changing preferences' (RW, 2020a, 79). However, as noted in sub-section *SR's consistency conditions*, the posited instances of preference change are not equally plausible, and choice modellers can frequently assess the descriptive and/or normative plausibility of such instances of preference change. Moreover, when choice modellers cannot reliably assess the descriptive and/or normative plausibility of the posited instances of preference change, adopting IR does not *per se* yield more informative descriptive and/or normative insights about such instances of preference change compared to SR.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

To illustrate this, consider the case of defaults. As noted by RW (2023, 206), several defaults reduce individuals' decision-making costs and are regarded by individuals as recommendations from trusted sources. Still, RW's claim that SR-rational individuals 'would be affected by these clearly relevant factors' (ibid., 206) does not bear against SR. For if some defaults reduce individuals' decision-making costs and are regarded by individuals as recommendations from trusted sources, then SR allows that such defaults may rationally influence individuals' decisions (see, e.g., Oliver, 2013). To be sure, one may point to various cases where choice modellers disagree as to whether specific defaults may rationally influence individuals' decisions (see, e.g., RW, 2023, 206). Yet, the existence of such contested cases does not selectively support IR over SR. For several cases are not contested (see, e.g., Sunstein, 2015, on various defaults concerning dietary and financial decisions). And in contested cases, adopting IR does not *per se* yield more informative descriptive and/or normative insights about the examined defaults compared to SR.

Self-regulation. Individuals frequently rely on self-regulation across a variety of choice settings (see, e.g., RW, 2009a). According to RW, individuals' reliance on self-regulation is incompatible with SR, since SR-rational individuals 'will simply select the best option from those available' (2020a, 3). Moreover, RW hold, individuals' reliance on self-regulation is often plausibly regarded as inclusively rational (ibid., ch. 1). However, the proponents of SR have developed various models to accommodate individuals' reliance on self-regulation (see, e.g., Gul and Pesendorfer, 2001, on various models of temptation; Ross, 2011, on various multiple-self models). Moreover, the proponents of SR are not committed to regarding individuals' reliance on self-regulation as irrational. For nothing in SR excludes the possibility that, in a given decision problem, relying on self-regulation may be 'the best' option available to individuals. And although one may point to various cases where choice modellers disagree as to whether self-regulation is 'the best' option available to individuals (see, e.g., RW, 2009a), the existence of such contested cases does not selectively support IR over SR. For several cases are not contested (see, e.g., Fumagalli, 2024, on various cases involving harmful addiction). And in contested cases, adopting IR does not *per se* yield more informative descriptive and/or normative insights about the examined instances of self-regulation compared to SR.

Falsifiability concerns

RW hold that SR models 'do sometimes pass falsification tests', but often 'do not perform […] well' in terms of falsifiability (2020a, 38). The idea is that choice modellers are frequently unable to assess the rationality of individuals' choices since 'rationality and irrationality are defined relative to subjective preferences that are typically *unobserved* and often *unobservable*' (ibid., 408, italics added; also Dold and Rizzo, 2024, 8–10). These remarks aptly emphasize the limited falsifiability of SR as an abstract mathematical framework (see, e.g., Blaug, 1992, ch. 4; Fumagalli, 2020b), but do not cast doubt on the falsifiability of specific SR hypotheses – i.e. hypotheses stating that particular individuals' preferences satisfy SR's axioms – compared to hypotheses based on RW's theory of IR. For in primis, SR hypotheses are *more amenable* to empirical/experimental testing than the proponents of IR seem to presuppose. And second, RW's theory of

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

IR faces substantial *falsifiability concerns*, which seem more widespread and pervasive than those faced by SR. Let me expand on these two issues in turn.[14]

The empirical implications of SR models are typically conditional on several auxiliary assumptions (see, e.g., Cubitt *et al.*, 2001, on assumptions concerning the adequacy of individuals' incentives). Therefore, the hypothesis that particular individuals' preferences satisfy SR's axioms can rarely be tested independently of auxiliary assumptions (see, e.g., Bhattacharyya *et al.*, 2011, 142–143; Cubitt, 2005, 208). In this context, the availability of some findings contrary to the empirical implications of SR models does not *per se* imply that reliance on SR's axioms is unjustified. For in many cases, findings contrary to the empirical implications of SR models are more plausibly regarded as evidence against some of the auxiliary assumptions rather than evidence against SR's axioms (see, e.g., Hausman, 1992, ch. 12). This, however, by no means implies that SR hypotheses are unfalsifiable. For choice modellers can test the validity of specific auxiliary assumptions by performing a series of experimental reproductions (see, e.g., Plott and Smith, 2008). And these experimental reproductions may enable choice modellers to significantly reduce the set of factors that can be plausibly invoked to accommodate alleged violations of SR's axioms (see, e.g., Fumagalli, 2016b, on the adequacy of individuals' incentives). That is to say, if observed choices seem to contradict the hypothesis that individuals' preferences satisfy SR's axioms, then choice modellers should test auxiliary assumptions about 'procedures, payoffs, context, instructions, etc. […] rather than conclude that [the involved individuals] are irrational' (Smith, 2003, 471). Yet, if the alleged violations of SR's axioms persist once these auxiliary assumptions have been tested, then the hypothesis that individuals' preferences fail to satisfy SR's axioms is more plausible than the alleged failure of such auxiliary assumptions (see, e.g., Fumagalli, 2020c, 350; also Section 'Standard rationality versus inclusive rationality' on the reported violations of specific axioms and on SR models developed in response to such violations).[15]

As to the falsifiability concerns faced by RW's theory of IR, RW hold that IR does not 'function exclusively as a normative *concept* [but also] as a positive *research program* […] for generating testable hypotheses' (WR, 2021, 385, italics added). In their view, IR 'incorporates many subsidiary questions with *testable implications* [about] whether (and how much) people learn over time [and] whether (and how) people adopt regimes of self-[regulation]' (RW, 2020b, italics added). However, the proffered characterizations of the notion of IR are insufficiently specific to imply specific testable hypotheses

[14]Over the last few decades, much debate has taken place concerning Popper's (1962, ch. 1) view that falsifiability is a requirement for regarding hypotheses as scientific (see, e.g., Hands, 1985; Hansson, 2006). I am not concerned here with assessing such view. For my evaluation, I note that most proponents of SR and IR concur that falsifiability is an important desideratum for the hypotheses figuring in specific models of choice (see, e.g., Dietrich and List, 2016, 195; Fumagalli, 2020c, 349; WR, 2021, 382–383).

[15]A proponent of IR may object that many SR hypotheses are unfalsifiable on the alleged ground that choice modellers cannot 'conclude whether or not [an] agent satisfies [SR's axioms] without referring to the concerns of the agent [and that] there can be an infinite number of different […] concerns guiding the agent's choices' (Bhattacharyya *et al.*, 2011, 145–146; also Sen, 1993, 501–503). However, this objection significantly underestimates the degree of falsifiability of SR hypotheses. For appeals to agents' concerns are not equally plausible, and choice modellers can often assess the plausibility of different appeals to such concerns (see, e.g., sub-sections *SR's consistency conditions* and *Compatibility of IR and SR*).

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

about individuals' learning and self-regulation (e.g., how much learning is implied by IR in particular contexts? Which instances of self-regulation are compatible with IR?). In this respect, it would be of limited import to object that RW's theory of IR counsels choice modellers 'to have some [epistemic] humility […] rather than indulging the impulse to find fault' (RW, 2020c). For although such epistemic humility is commendable, addressing the falsifiability concerns faced by RW's theory of IR would require the proponents of IR to provide clear and informative criteria for demarcating which choice patterns are incompatible with IR. Regrettably, the proponents of IR have hitherto failed to provide such criteria. In fact, even authors sympathetic to IR question the falsifiability of RW's theory of IR (see, e.g., Rajagopalan, 2021, 269, holding that RW 'do not go far enough to explore […] difficult cases' and calling the proponents of IR to identify clear cases of choice patterns incompatible with IR).[16]

A proponent of IR may object that choice modellers 'should not be guided exclusively by the *falsifying goal* of finding exceptions [and] should also engage in the *confirming goal* of finding more varieties of inclusive rationality' (WR, 2021, 386, italics added). The idea would be that IR generates 'useful and sometimes successful hypotheses' (ibid., 385) and that 'there are *mountains of evidence* for [IR consisting] in all manner of self-regulatory behaviors [and] learning over time' (RW, 2020c, italics added). However, these claims appear to significantly overestimate the alleged empirical support for IR. For the proffered empirical evidence does not selectively support IR over SR (see, e.g., sub-sections *SR's consistency conditions* and *Compatibility of IR and SR*). In this perspective, much purported empirical support for IR may be plausibly regarded as an artefact of the vagueness of IR and of the ensuing unclarity concerning the putative implications of IR.

Welfare analyses

According to RW, SR's axioms are 'analytical assumptions that are *not welfare-relevant*' (2020a, 17, italics added) and 'consistency of choice [fails to provide] an *adequate basis*' for welfare analyses (Rizzo, 2024, 13, italics added). In their view, 'there is no valid and convincing basis' for determining which choices maximize welfare (RW, 2020a, 363). For although 'in principle, we can objectively define the choices that will maximize health, or lifespan, [or] wealth' (RW, 2020c), 'the correct weighting of [choices'] benefits and costs is *unavoidably subjective*' (RW, 2020a, 408, italics added). RW are correct that it is often difficult for choice modellers to identify which choices maximize welfare and that many proffered identifications of such choices are contested (see, e.g., Dold, 2018). However, the existence of these contested cases does not *per se* license general scepticism about SR's potential to ground informative and reliable evaluations of public policies' welfare implications (see, e.g., Fumagalli, 2021). In this respect, it would be

[16]In recent works, the proponents of IR hold that 'clear-cut cases of mistake are […] conceptually possible' within IR (Rizzo, 2024, 22) and conjecture that 'some child and addict behaviors [and] some varieties of mental illness might' be incompatible with IR (RW, 2020c). However, the proponents of IR concede that 'in practice, outside observers often lack the evidence' to identify choice patterns incompatible with IR since 'there are simply too many subjective variables' to consider (RW, 2020a, 433; also ibid., 407, holding that 'there are [no] cases in which people are obviously making mistakes'). In light of these remarks and the remarks provided in the main text, the falsifiability of RW's theory of IR appears to be significantly constrained.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

implausible to hold that SR's axioms are '*completely inadequate* as a prescriptive standard' (Rizzo, 2018, 193, italics added; also Berg and Gigerenzer, 2010, 148, holding that 'almost no empirical evidence exists documenting that individuals who deviate from [SR's axioms] earn less money, live shorter lives, or are less happy'). For although abiding by SR's axioms does not guarantee that individuals make welfare-optimal choices (see, e.g., Gilboa *et al.*, 2009, on cases where individuals' consistent choices are based on inaccurate information about the available options), individuals who abide by SR's axioms often tend to obtain higher welfare-relevant payoffs than they would obtain if they failed to abide by such axioms (see, e.g., sub-section *SR's consistency conditions* on transitivity).

More generally, the point remains that RW's theory of IR seemingly 'lacks analytical clarity when it comes to concrete questions of [welfare] evaluation' (Dold, 2023, 6). And this lack of clarity, in turn, constrains this theory's applicability to evaluating public policies' welfare implications. To illustrate this, consider RW's claim that within IR '*the appropriate* standard of well-being is *the one* you would impose on yourself' (2020c, italics added) and that 'the desirability of acting [on SR's axioms depends] on showing that failure to do so will result in bad consequences to decision-makers from *their own* point of view' (RW, 2020a, 121, italics added). These remarks seemingly presuppose a radical subjectivist conception of welfare, according to which the extent to which an individual is well-off is a purely subjective matter, i.e. exclusively depends on the individual's subjective judgements and attitudes towards her life. Yet, such conception of welfare is vulnerable to severe objections (see, e.g., Kagan, 2009; Parfit, 1984, appendix I) and does not ground informative and reliable evaluations of public policies' welfare implications (see, e.g., Griffin, 1986, ch. 1–3, and Scanlon, 1996, on various goods/experiences that can affect individuals' welfare at least partly irrespective of individuals' subjective judgements and attitudes towards their lives).

A proponent of IR may object that IR can ground informative and reliable evaluations of public policies' welfare implications and that IR's focus on subjective considerations *enhances* (rather than hampers) IR's applicability to welfare analyses (see, e.g., Rizzo, 2024, 13, holding that in their welfare analyses the proponents of SR 'must admit that in back of choices are mental preferences'). The idea is that within IR '*the ultimate* standard by which individuals' behavior is evaluated is the degree of successful attainment of *goals* in the *actual environment* in which they find themselves' (RW, 2020a, 38, italics added). However, this evaluative standard does not *per se* ground informative and reliable evaluations of public policies' welfare implications (e.g. how should choice modellers identify individuals' goals? How is the degree of successful attainment of such goals measured? And are all individuals' goals such that their attainment directly contributes to individuals' welfare?). In fact, appealing to individuals' 'environment' may further hamper IR's applicability to evaluating public policies' welfare implications. For the proponents of IR rarely provide precise specifications of which factors are plausibly taken to belong to individuals' 'environment'. And this paucity of precise specifications hampers choice modellers' ability to ground informative and reliable evaluations of public policies' welfare implications on appeals to individuals' 'environment' (see, e.g., Hands, 2014, 407, for similar remarks targeting the generic characterizations of individuals' 'environment' presupposed by leading ecological rationality models).

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

A proponent of IR may further object that IR 'could allow' choice modellers to ground informative and reliable evaluations of public policies' welfare implications 'in a manner *unrelated* to the violation of consistency axioms […] by getting *inside people's heads* as much as is feasible' (WR, 2021, 385, italics added; also WR, 2018, 214, commenting on what 'neuronal and behavioral responses to prediction errors […] we should expect on the part of [inclusively] rational actors'). However, the proponents of IR currently lack a suitable basis to ground informative and reliable evaluations of public policies' welfare implications on empirical assumptions about neuro-psychological substrates. For many different (and often conflicting) models of the neuro-psychological substrates of choice have been advocated in the recent literature (see, e.g., Padoa-Schioppa and Schoenbaum, 2015). And despite the ongoing advances in neuro-psychological modelling, many prominent neuro-psychological models of choice are more plausibly regarded as 'as-if' models rather than accurate characterizations of the neuro-psychological substrates of individuals' choices (see, e.g., Moscati, 2024, targeting leading ecological rationality models). Moreover, it is dubious that choice modellers' evaluations of public policies' welfare implications should be grounded on empirical assumptions about neuro-psychological substrates. For severe difficulties plague the proffered attempts to build neuro-psychological indexes of welfare (see, e.g., Fumagalli, 2019, on influential neuro-psychological indexes' failure to track what many theories of welfare regard as individuals' welfare). And prominent proponents of neuro-psychological indexes sharply disagree as to which indexes should be adopted to evaluate public policies' welfare implications (see, e.g., Fumagalli, 2022, for illustrations).

Conclusion

In this paper, I have articulated and defended a critical assessment of RW's theory of IR in light of the ongoing cross-disciplinary debate about rationality, welfare analyses and policy evaluation. The paper aimed to provide three main contributions to this debate. First, it explicated the relation between the consistency conditions presupposed by SR and the standards of rationality presupposed by RW's theory of IR. Second, it provided a qualified defence of the consistency conditions presupposed by SR against the main criticisms put forward by RW. And third, it identified and discussed specific strengths and weaknesses of RW's theory of IR in the context of welfare analyses and policy evaluation.

In their influential works, RW provide valuable critical insights concerning the descriptive/normative validity of SR and SR's applicability to evaluating public policies' welfare implications. However, as it stands, RW's theory of IR is vulnerable to objections. These objections do not detract from the many merits of RW's works and do not bear against the main criticisms that RW articulate against behavioural paternalists' attempts to design and implement welfare-enhancing paternalistic interventions in public policy. Still, if correct, they challenge RW to qualify and/or better support their theory of IR.

As to future developments in the broader cross-disciplinary debate about rationality, welfare analyses and policy evaluation, three lines of research seem especially worthy of

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

investigation, namely: assessing the prospects of a possible synthesis or partial convergence between SR and IR despite their several differences; further exploring the relation between SR, IR and other notions of rationality (e.g., ecological rationality) that figure prominently in the specialized cross-disciplinary literature; and probing the applicability of SR and IR to specific debates in welfare analyses and policy evaluation, as aptly showcased by RW's influential works concerning paternalistic interventions.

Acknowledgements. I thank Malte Dold, Mario Rizzo and an anonymous reviewer for their comments on earlier versions of this paper. I also received helpful feedback from audiences at King's College London, the University of Bolzano, the University of Pennsylvania, and New York University.

Funding statement. I acknowledge the support of the Foundations of the Market Economy Program, Department of Economics, New York University, for a brief research visit at New York University.

References

Anand, P. (1993), 'The philosophy of intransitive preference', *The Economic Journal*, **103**: 337–346.
Angner, E. (2019), 'We're all behavioral economists now', *Journal of Economic Methodology*, **26**: 195–207.
Aumann, R. (1962), 'Utility theory without the completeness axiom', *Econometrica*, **30**: 445–462.
Beck, L. (2024), 'Why we need to talk about preferences: economic experiments and the where-question', *Erkenntnis*, **89**: 1435–1455.
Berg, N. and G. Gigerenzer (2010), 'As-if behavioral economics: neoclassical economics in disguise?', *History of Economic Ideas*, **18**: 133–166.
Bhattacharyya, A., K. Pattanaik and Y. Xu (2011), 'Choice, internal consistency and rationality', *Economics and Philosophy*, **27**: 123–149.
Blaug, M. (1992), *The Methodology of Economics, or How Economists Explain*, 2nd edn, Cambridge: Cambridge University Press.
Bradley, R. (2017), *Decision Theory with a Human Face*, Cambridge: Cambridge University Press.
Broome, J. (1993), 'Can a Humean be Moderate?', in R. Frey and C. Morris (eds), *Value, Welfare and Morality*, Cambridge: Cambridge University Press, 51–73.
Chetty, R. (2015), 'Behavioral economics and public policy: a pragmatic perspective', *American Economic Review*, **105**: 1–33.
Choi, S., S. Kariv, W. Müller and D. Silverman (2014), 'Who is (more) rational?', *American Economic Review*, **104**: 1518–1550.
Colander, D. (2000), 'The death of neoclassical economics', *Journal of the History of Economic Thought*, **22**: 127–144.
Cowen, N. and M. Dold (2021), 'Introduction: symposium on Escaping Paternalism by Mario J. Rizzo and Glen Whitman', *Review of Behavioral Economics*, **8**: 213–220.
Cubitt, R. (2005), 'Experiments and the domain of economic theory', *Journal of Economic Methodology*, **12**: 197–210.
Cubitt, R., C. Starmer and R. Sugden (2001), 'Discovered preferences and the experimental evidence of violations of expected utility theory', *Journal of Economic Methodology*, **8**: 385–414.
Dekker, E. and B. Remic (2019), 'Two types of ecological rationality: or how to best combine psychology and economics', *Journal of Economic Methodology*, **26**: 291–306.
Dietrich, F. and C. List (2011), 'A model of non-informational preference change', *Journal of Theoretical Politics*, **23**: 145–164.
Dietrich, F. and C. List (2016), 'Reason-based choice and context dependence: an explanatory framework', *Economics and Philosophy*, **32**: 175–229.
Dold, M. (2018), 'Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics', *Journal of Economic Methodology*, **25**: 160–178.
Dold, M. (2023) 'Review of Escaping Paternalism: Rationality, Behavioral Economics and Public Policy. Rizzo, M. and Whitman, G. Cambridge University Press, 2020', *Behavioural Public Policy*, 1–8.
Dold, M. and M. Rizzo (2024), 'Hayekian psychological economics', *Behavioural Public Policy*, **8**: 773–788.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

Dreier, J. (1996), 'Rational preference: decision theory as a theory of practical rationality', *Theory and Decision*, **40**: 249–276.
Fumagalli, R. (2013), 'The futile search for true utility', *Economics and Philosophy*, **29**: 325–347.
Fumagalli, R. (2016a), 'Decision sciences and the new case for paternalism: three welfare-related justificatory challenges', *Social Choice & Welfare*, **47**: 459–480.
Fumagalli, R. (2016b), 'Economics, psychology and the unity of the decision sciences', *Philosophy of the Social Sciences*, **46**: 103–128.
Fumagalli, R. (2019), '(F)utility exposed', *Philosophy of Science*, **86**: 955–966.
Fumagalli, R. (2020a), 'Slipping on slippery slope arguments', *Bioethics*, **34**: 412–419.
Fumagalli, R. (2020b), 'How thin rational choice theory explains choices', *Studies in History and Philosophy of Science Part A*, **83**: 63–74.
Fumagalli, R. (2020c), 'On the individuation of choice options', *Philosophy of the Social Sciences*, **50**: 338–365.
Fumagalli, R. (2021), 'Theories of well-being and well-being policy: a view from methodology', *Journal of Economic Methodology*, **28**: 124–133.
Fumagalli, R. (2022), 'A reformed division of labor for the science of well-being', *Philosophy*, **97**: 509–543.
Fumagalli, R. (2024), 'Preferences versus opportunities: on the conceptual foundations of normative welfare economics', *Economics and Philosophy*, **40**: 77–101.
Gigerenzer, G. (2021), 'Axiomatic rationality and ecological rationality', *Synthese*, **198**: 3547–3564.
Gilboa, I., A. Postlewaite and D. Schmeidler (2009), 'Is it always rational to satisfy Savage's axioms?', *Economics and Philosophy*, **25**: 285–296.
Glimcher, P. (2011), *Foundations of Neuroeconomic Analysis*, New York: Oxford University Press.
Griffin, J. (1986), *Well-Being: Its Measure and Importance*, Oxford: Clarendon Press.
Grüne-Yanoff, T. (2021), 'Boosts: a remedy for Rizzo and Whitman's Panglossian fatalism', *Review of Behavioral Economics*, **8**: 285–303.
Guala, F. (2000), 'The logic of normative falsification: rationality and experiments in decision theory', *Journal of Economic Methodology*, **7**: 59–93.
Guala, F. (2005), 'Economics in the lab: completeness vs. testability', *Journal of Economic Methodology*, **12**: 185–196.
Gul, F. and W. Pesendorfer (2001), 'Temptation and self-control', *Econometrica*, **69**: 1403–1436.
Gustafsson, J. (2022), *Money-Pump Arguments*, Cambridge: Cambridge University Press.
Hands, W. (1985), 'Karl Popper and economic methodology: a new look', *Economics and Philosophy*, **1**: 83–100.
Hands, W. (2014), 'Normative ecological rationality: normative rationality in the fast-and-frugal-heuristics research program', *Journal of Economic Methodology*, **21**: 396–410.
Hansson, S. (2006), 'Falsificationism falsified', *Foundations of Science*, **11**: 275–286.
Hausman, D. (1992), *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.
Kagan, S. (2009), 'Well-being as enjoying the good', *Philosophical Perspectives*, **23**: 253–272.
List, J. (2003), 'Does market experience eliminate market anomalies?', *Quarterly Journal of Economics*, **118**: 41–71.
Machina, M. (2008), 'Non-expected utility theory', in S. Durlauf and L. Blume, *The New Palgrave Dictionary of Economics*, 2nd edn, New York: Palgrave Macmillan, 74–84.
Mandler, M. (2005), 'Incomplete preferences and rational intransitivity of choice', *Games and Economic Behavior*, **50**: 255–277.
Moscati, I. (2024), 'Behavioural and heuristic models are as-if models too - and that's ok', *Economics and Philosophy*, **40**: 279–309.
Nielsen, K. and J. Rehbeck (2022), 'When choices are mistakes', *American Economic Review*, **112**: 2237–2268.
Oliver, A. (2013), 'From nudging to budging: using behavioural economics to inform public sector policy', *Journal of Social Policy*, **42**: 685–700.
Oliver, A. (2023), *A Political Economy of Behavioral Public Policy*, Cambridge: Cambridge University Press.
Oprea, R. (2020), 'What makes a rule complex?', *American Economic Review*, **110**: 3913–3951.
Padoa-Schioppa, C. and G. Schoenbaum (2015), 'Dialogue on economic choice, learning theory, and neuronal representations', *Current Opinion in Behavioral Sciences*, **5**: 16–23.
Parfit, D. (1984), *Reasons and Persons*, Oxford: Clarendon Press.
Plott, C. and V. Smith (2008), *Handbook of Experimental Economics Results*. Vol.1, New York: Elsevier.
Popper, K. (1962), *Conjectures and Refutations*, New York: Basic Books.

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press

Rajagopalan, S. (2021), 'Inclusive rationality: struggle and aspiration', *Review of Behavioral Economics*, **8**: 259–283.
Rizzo, M. (2018), 'Rationality - what? Misconceptions of neoclassical and behavioral economics', in M. Henderson (ed), *The Cambridge Handbook of Classical Liberal Thought*, Cambridge: Cambridge University Press, 191–212.
Rizzo, M. (2019), 'Inconsistency is not pathological: a pragmatic perspective', *Mind & Society*, **18**: 77–85.
Rizzo, M. (2024), 'The problem of counterfactual preferences', *Social Philosophy & Policy* , In Press.
Rizzo, M. and G. Whitman (2009a), 'The knowledge problem of the new paternalism', *Brigham Young University Law Review*, 905–968.
Rizzo, M. and G. Whitman (2009b), 'Little brother is watching you: new paternalism on the slippery slopes', *Arizona Law Review*, **51**: 685–739.
Rizzo, M. and G. Whitman (2018), 'Rationality as a process', *Review of Behavioral Economics*, **5**: 201–219.
Rizzo, M. and G. Whitman (2020a), *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*, Cambridge: Cambridge University Press.
Rizzo, M. and G. Whitman (2020b), 'Escaping paternalism book club: Rizzo and Whitman response, part 2', *The Library of Economics and Liberty*. Available at: https://www.econlib.org/escaping-paternalism-book-club-rizzo-and-whitman-response-part-2/ (Accessed by 1 Janaury 2024).
Rizzo, M. and G. Whitman (2020c), 'Escaping paternalism book club: Rizzo and Whitman response, part 3', *The Library of Economics and Liberty*. Available at: https://www.econlib.org/escaping-paternalism-book-club-rizzo-and-whitman-response-part-3/ (Accessed by 1 Janaury 2024).
Rizzo, M. and G. Whitman (2023), 'The unsolved Hayekian knowledge problem in behavioral economics', *Behavioural Public Policy*, **7**: 199–211.
Ross, D. (2011), 'Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics', *Journal of Economic Methodology*, **18**: 217–231.
Ross, D. (2014), 'Psychological versus economic models of bounded rationality', *Journal of Economic Methodology*, **21**: 411–427.
Savage, L. (1954), *The Foundations of Statistics*, Hoboken, NJ: Wiley.
Scanlon, T. (1996), 'The status of well-being', *Tanner Lectures on Human Values*, **16**: 91–143.
Sen, A. (1993), 'Internal consistency of choice', *Econometrica*, **61**: 495–521.
Smith, V. (2003), 'Constructivist and ecological rationality in economics', *American Economic Review*, **93**: 465–508.
Starmer, C. (2000), 'Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk', *Journal of Economic Literature*, **38**: 332–382.
Starmer, C. (2005), 'Normative notions in descriptive dialogues', *Journal of Economic Methodology*, **12**: 277–289.
Strohmaier, D. and M. Messerli (2024), *Preference Change*, Cambridge: Cambridge University Press.
Sugden, R. (1991), 'Rational choice: a survey of contributions from economics and philosophy', *The Economic Journal*, **101**: 751–785.
Sunstein, C. (2015), 'Nudges, agency, and abstraction: a reply to critics', *Review of Philosophy and Psychology*, **6**: 511–529.
Von Neumann, J. and O. Morgenstern (1947), *Theory of Games and Economic Behavior*, 2nd edn, Princeton: Princeton University Press.
Whitman, G. and M. Rizzo (2015), 'The problematic welfare standards of behavioral paternalism', *Review of Philosophy and Psychology*, **6**: 409–425.
Whitman, G. and M. Rizzo (2021), 'Inclusive rationality and paternalism: responses to comments and criticism', *Review of Behavioral Economics*, **8**: 379–394.

Cite this article: Fumagalli R (2025), 'Standard rationality versus inclusive rationality: a critical assessment', *Behavioural Public Policy* **9**, 530–544. https://doi.org/10.1017/bpp.2024.53

https://doi.org/10.1017/bpp.2024.53 Published online by Cambridge University Press