J. Benefit Cost Anal. 2016; 7(1):196–219
doi:10.1017/bca.2016.4

Matthew D. Adler

Behavioral Economics, Happiness Surveys, and Public Policy[1]

Abstract: Two important developments in recent policy analysis are behavioral economics and subjective-well-being (SWB) surveys. What is the connection between them? Some have suggested that behavioral economics strengthens the case for SWB surveys as a central policy tool, e.g., in the form of SWB-based cost-benefit analysis. This article reaches a different conclusion. Behavioral economics shows that individuals in their day-to-day, "System 1" behavior are not expected utility (EU-) rational – that they often fail to comply with the norms of rationality set forth by EU theory. Consider now that the standard preference-based view of individual well-being looks to individuals' rational preferences. If the findings of behavioral economics are correct, an individual's answer to a question such as "How satisfied are you with your life?" is not going to tell us much about her rational (EU-compliant) preferences. Behavioral economics, by highlighting widespread failures of EU rationality, might actually argue for an objective-good (non-preference-based) view of well-being. However (except in the limiting case of an objective-good view positing a single mentalistic good, happiness), SWB surveys will not be strong evidence of well-being in the objective-good sense. In short, SWB surveys are no "magic cure" for the genuine difficulties in inferring rational preferences and measuring well-being underscored by behavioral economics.

Keywords: behavioral; behavioral economics; expected utility; happiness; rationality; subjective well-being; theory.

JEL classifications: D03; D61; D69.

1 Introduction

Two major developments in policy studies, in recent years, are subjective-well-being ("SWB") surveys and behavioral economics. A SWB survey asks the respondent to quantify her happiness, life satisfaction, sense of purpose or meaning, or

[1] Richard A. Horvitz, Professor of Law and Professor of Economics, Philosophy, and Public Policy, Duke University. Many thanks to Ryan Bubb and to two anonymous referees for helpful comments.

Matthew D. Adler: Duke Law School, Durham, NC 27708, United States,
e-mail: adler@law.duke.edu

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

some other such (affective or valuational) mental state. A substantial literature examines correlations between the respondent's stated SWB and other attributes, such as the respondent's income, employment status, health, the level of environmental amenities (e.g., the amount of air pollution) in her vicinity, and so forth (see Adler, 2013, citing literature; Graham, 2016). Recently, building on this literature, some have suggested that SWB surveys be used as a central tool for policy analysis (Bronsteen, Buccafusco & Masur, 2015; Fujiwara & Dolan, 2016; Adler, 2013, pp. 1514–1517, citing additional sources). In particular, SWB-based cost-benefit analysis would calculate individuals' compensating or equivalent variations for a policy change from the status quo in terms of SWB, rather than (as in traditional cost-benefit analysis) in terms of individuals' willingness-to-pay or -accept amounts.

Behavioral economics, of course, documents individual deviations from the model of expected utility ("EU") maximization central to neoclassical economics (Camerer, Loewenstein & Rabin, 2004; Cartwright, 2011; DellaVigna, 2009; Kahneman, 2011; Kahneman & Tversky, 2000), and many have now explored the implications of such deviations for policy design (Chetty, 2015; Congdon, Kling & Mullainathan, 2011; Madrian, 2014; Robinson & Hammitt, 2011; Thaler & Sunstein 2009).

In this article, I will examine the connection between these two developments. Specifically: Does behavioral economics support the use of SWB surveys as a central tool for policy analysis – as in SWB-based cost-benefit analysis, or in other formats where SWB surveys function as the metric of individual well-being, e.g., a SWB-based social welfare function, or "gross national happiness" calculated using SWB surveys?

In his 2015 Richard T. Ely lecture, the prominent economist Raj Chetty argues as follows: behavioral economics demonstrates that individuals regularly make choices that fail to maximize their well-being, and thus policymakers should consider using non-choice-based measures of well-being, in particular SWB surveys. (Chetty, 2015, pp. 22–25).[2] A similar suggestion had earlier been made by Dolan and Kahneman in an oft-cited article (Dolan & Kahneman, 2008).[3]

[2] To be clear, Chetty does not suggest that SWB surveys are the only viable measure of well-being. He also endorses well-constructed choice-based measures – for example, inferring well-being from choices in situations where the individuals can be presumed to be welfare-maximizing (Chetty, 2015, pp. 25–26).

[3] In both these articles, the authors point to a gap between "decision utility" and "experience utility" – presupposing the latter to be the measure of individual well-being. But whether "experience utility" (in one or another sense) does indeed measure well-being is open to question. This topic is discussed at length in Section 4 below. According to one view of well-being discussed there ("View A"), well-being consists in the realization of rational self-interested preferences formed under favorable deliberative conditions, with "self-interest" defined to permit individuals to have preferences for features of the world other than their own mental states. According to View A, a given person's well-being may consist

Indeed, it is natural to think that behavioral economics and SWB-based policy analysis "go together," since both are challenges to neoclassical[4] economics – respectively, to preference maximization as an explanatory model of individual choice, and to the existing methodologies (stated-preference and revealed-preference methodologies) that economists use to measure well-being for purposes of formulating policy advice.

However, I will argue here that the nexus between behavioral economics and SWB-based policy analysis (for short, "SBPA") is more complicated.[5] The issues are, in substantial part, normative. Proponents of SBPA make a normative claim: that we *ought* to use SWB surveys in designing policy, that doing so is *better* than traditional cost-benefit analysis. Behavioral economics buttresses SBPA insofar as it helps to make the normative case for the latter more *persuasive*.

To be sure, there will be important empirical questions that arise in evaluating the case for SBPA – indeed, this is where behavioral economics will prove relevant. My point is that a defense of SBPA cannot be *wholly* empirical and non-normative. Every economist who gives advice about policy choice, or about the framework for structuring policy choice, should make a mantra of Hume's saying: "No ought from is" (Atkinson, 2009).

The most fundamental normative commitment of welfare economics is Paretian welfarism – that the comparative ethical goodness of two outcomes depends upon their associated patterns of well-being; and (strong Pareto) that if everyone is at least as well off in outcome x as in outcome y, with at least one person strictly better off, x is an ethically better outcome (Adler, 2012). Proponents of SBPA do not tend to challenge Paretian welfarism, and I will take it as fixed here.[6] *Given* a normative commitment to Paretian welfarism, does behavioral economics buttress the normative case for SBPA? The aim of what follows is to grapple with this question in some detail, with special care taken to illustrate the mix of empirical *and* normative issues that must be addressed in answering it.

in *more* than her experiences (mental states). The same is true for View C, an objective-good view, except in the special case where all the goods are mentalistic.

[4] By "neoclassical" I mean the understanding of economics that dominated academic work for much of the 20th century, beginning in the 1930s, and that is summarized for example in Mas-Collel et al. (1995).

[5] See also the contributions to this symposium by Bernheim (2016) and Sunstein (2016). These works – like this article – are skeptical of SBPA.

[6] Alternatively, one might start with weak welfarism and see SBPA as capturing the well-being component of ethical assessment (Bronsteen et al., 2015). Weak welfarism endorses the ethical significance of individual well-being, but is open to the possibility that non-welfare considerations (such as individual rights or intrinsic environmental values) also have ethical weight.

A defense of SBPA grounded in weak rather than Paretian welfarism would be equally vulnerable (I believe) to the critical analysis presented in this article. Either defense must argue that SWB is a good measure of individual well-being – and that is what my analysis seeks to question.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

2 Rationality and well-being

Right at the outset, two key normative questions need to be flagged. The first concerns the nature of *rational choice*. The second concerns the nature of *well-being*.

Why think that these are normative issues? The word "rational," in its standard usage, is a term of approval, and "irrational" a term of criticism (Gibbard, 1990). It would be deeply odd to say to a decisionmaker: "choice a is not rational, and I advise you to make choice a." To characterize EU maximization as rational is to *recommend* (both to oneself and to other agents) behavior in conformity with the EU framework. Similarly, someone's well-being has normative force. Specifically, for Paretian welfarists, well-being is that feature of individual lives – whatever it might be – such that what ought to be done, ethically, depends solely on the pattern of well-being. Deciding what well-being means is, thus, one central aspect of ethical deliberation.

Let us consider, to begin, the views of neoclassical economists on these two normative issues, as well as related philosophical positions. (1) *Rational choice*. Rational choice has both static and dynamic aspects. (a) *Static rationality*. Norms of static (synchronic) rationality concern *contemporaneous* choice: how a decisionmaker should choose among the options in a choice situation that she is currently facing, given her current preferences and information. Neoclassical economics adopts EU maximization as the standard of static rationality (Gilboa, 2009; Kreps, 1988; Joyce, 1999; Mas-Collel, Whinston & Green, 1995, Chapter 6). An individual's choice in a given choice situation is rational if (i) the individual's preferences can be expressed as a complete and transitive ranking of all the possible outcomes of the choices, as well as a complete and transitive ranking of the choices; (ii) the individual's ranking of the choices and outcomes conforms to some cluster of axioms (for example, the Savage (1954), Anscombe and Aumann (1963), or Jeffrey (1983) axioms) such that the position of each choice in the choice ranking corresponds to its probabilistically expected utility (in light of some utility function assigning utilities to outcomes, and some probability measure assigning probabilities to outcomes conditional on each choice); and (iii) the choice selected by the individual is at the top of the choice ranking or, equivalently, has the greatest expected utility. (b) *Dynamic rationality*. Norms of dynamic (diachronic) rationality concern choice over time. How should a current choice cohere with prior choices or plans? Neoclassical economists and decision theorists tend to endorse Bayesian updating as the norm for dynamic rationality of beliefs (Joyce, 2004). An individual's beliefs at a given time should not only take the form of a well-behaved probability distribution over outcomes; in addition, that distribution should be the result of a prior distribution, updated using new information in accordance with

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

Bayes' rule. A different norm of dynamic rationality, also commonly adopted, is time consistency: if an individual at an earlier time adopts a rational plan to make some choice in a future choice situation, then (absent some unforeseen contingency) the individual should carry out the plan (Strotz, 1956).

(2) *Rationality and well-being*. The nature of individual well-being is a matter of ongoing philosophical debate. (For citations to the philosophical literature that is reviewed in the following paragraphs, see generally Adler (2012), Chapter 3.) One school that dates back to Aristotle, and that retains vibrancy in contemporary philosophical scholarship, adopts an "objective-good" view of well-being. According to this view, an individual's welfare is enhanced by various goods that are "objective" in the sense that an individual's attainment of each good (or at least her overall attainment with respect to the balance of goods) is defined independently of what the individual prefers.

However, a different philosophical tradition adopts a preference-based view of well-being. Philosophers in this school endorse some specific version of the following: someone is better off with choice *a* rather than choice *b* iff (i) she would rationally prefer *a* to *b* under favorable deliberative conditions; and (ii) this preference is self-interested. "Favorable deliberative conditions" means having good information, being in an emotional state conducive to rational choice, etc. Philosophers within the preferentialist school debate the specific content of "favorable deliberative conditions": does good information mean omniscience; all the information that the human brain (or the human brain plus external storage devices) can hold; or something less demanding? Is the best emotional state for choosing with respect to well-being one of calm detachment, or some degree of aroused engagement?

These questions are themselves normative. Philosophers in the preferentialist school share a basic normative commitment to the notion that a given person's well-being is equivalent to what she would rationally self-interestedly prefer under favorable deliberative conditions; but have a range of more fully specified normative views all of which are consistent with this basic commitment.

Philosophers in the preferentialist school tend to agree that welfare-relevant preferences need to be self-interested. Cases of self-sacrifice suggest the need for a self-interest screen. Intuitively, there is no contradiction in someone knowingly choosing to sacrifice her own welfare (for example, to jump on the grenade so as to save her comrades), but if well-being is equivalent to preference satisfaction without a self-interest screen, then knowing self-sacrifice is a contradiction.[7]

[7] An anonymous referee notes that altruism *can* be self-interested (in the sense that, by acting altruistically to increase others' well-being, I also increase my own). The point of the self-sacrifice example is that, in *some* (not necessarily all) cases, the actor increases others' well-being or advances various other ethical, legal, religious, etc. goals, but does not improve her own well-being. A self-interest screen on

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

To be sure, what it means for a preference to be “self-interested” is a matter for debate. Again, there is a range of normative views here consistent with the more basic normative commitment to equating well-being and preferences. This topic will be further discussed below.

Finally, note that my summary of the preference view of well-being, as advanced by philosophers, requires the preferences to be *rational*. Since to characterize a choice as “rational” is to *recommend* that choice (see above), and since enhancing someone’s well-being is (ceteris paribus) a good thing for Paretian welfarists – something they want to recommend – it would be deeply odd to define well-being in terms of preferences that need not be rational.[8]

How do these philosophical views relate to welfare economics? Neoclassical economists would (I suggest) *agree* to the normative claim that: someone’s well-being is what she would rationally self-interestedly prefer under favorable deliberative conditions. There is no dispute that neoclassical economics defines well-being in terms of rational preferences. Moreover, since the “favorable deliberative conditions” requirement has been left unspecified, this component is capacious enough to capture the range of views held by neoclassical economists.[9]

Finally, neoclassical economics seems to oscillate on the “self-interest” condition. One view seems to be that no such screen is needed: if someone has a rational, well-informed preference for *a* over *b*, then she is better off with *a*, full stop. (For example, if someone prefers *a* to *b* because she believes *a* to be more fair, then she has a “taste for fairness,” and she is better off if this taste is satisfied (Kaplow & Shavell, 2002).) On the other hand, in practice, neoclassical economists often do describe “self-interested” preferences as a particular category of preferences – seeing non-self-interested behavior as a possible, albeit non-standard case. (On this issue, see Congdon et al., 2011; DellaVigna, 2009.)

Since, again, the nature of “self-interest” has been left unspecified (at the limit, one *could* take the position that there is never a difference between self-interested and all-things-considered preferences), the definition of well-being in terms of rational, self-interested preferences formed under favorable deliberative conditions *does* (I believe) capture the views of neoclassical economics.

preferences is needed for *any* such cases to be possible. To be sure, specifying the screen – and thus delineating between self-interested and non-self-interested altruism – is difficult and controversial. The philosophical literature disagrees on how to do so (see Adler, 2013), but not on the need for some such screen.

[8] For a Paretian welfarist to define well-being in terms of preferences that may fall short of rationality would mean that she adopts an internally conflicted normative stance, one that both recommends a choice that increases well-being (qua well-being), yet may recommend not taking the choice (qua lacking rationality).

[9] For example, some neoclassical economists require welfare-relevant preferences to be well-informed (Kaplow & Shavell, 2002, p. 410), while others might look to actual preferences.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

3 Behavioral economics and rationality

How might behavioral economics undercut the normative positions just discussed? Consider first (1), the neoclassical position regarding *rational choice*: EU maximization plus neoclassical norms of dynamic rationality (in particular dynamic consistency and Bayesian updating).

"Behavioral economics" is a body of scholarship comprised, in part, of empirical findings that demonstrate beyond a reasonable doubt that individuals often do not comply with the norms of neoclassical rationality (Camerer et al., 2004; Cartwright, 2011; DellaVigna, 2009; Kahneman, 2011; Kahneman & Tversky, 2000). With respect to static choice, individuals depart from EU maximization. They frame choices as losses or gains from a reference point, rather than as probability distributions over final outcomes. Their expressed beliefs in propositions often violate the probability calculus, as in the famous "Linda" example (assigning a greater degree of belief to Linda being a bank teller active in the feminist movement than to Linda being a bank teller); and sometimes no probability distribution can explain individual choices even leaving aside loss–gain framing (as in Ellsberg choice (Machina, 2014)). With respect to intertemporal choice, individuals are often dynamically inconsistent (preferring at time T_0 *not* to engage in some activity at time T_1 that has immediate benefits but long-run costs, yet when T_1 comes "losing their will power" and going for the immediate gratification); and individuals stubbornly refuse to revise their "priors" in light of new information, as Bayesianism requires.

Behavioral economics also gives us parsimonious models of non-neoclassical behavior that help to explain these empirical findings: for example, prospect theory as a non-EU model of static choice; and hyperbolic discounting as a dynamically inconsistent model of intertemporal choice.

But how do these empirical findings and accompanying predictive models bear upon the *normative* issue at hand: whether *rational* choice, the kind of choice that we wish to *recommend*, conforms to neoclassical norms? *If* behavioral economics somehow demonstrated that it was impossible for individuals to satisfy neoclassical criteria of rationality – that ordinary humans are simply unable to regiment their choices in accordance with those criteria, just as ordinary humans cannot hear dog whistles or see in five dimensions – then behavioral economics *would* directly undercut a normative commitment to neoclassical criteria. Let's add Kant's dictum, "ought implies can," to Hume's "no ought from is."

Admittedly, the dynamic aspect of neoclassical rationality (especially the requirement that an individual's current probability assignments be derived via updating from initial assignments at some canonical moment – the onset of adulthood?)

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

is very demanding, and *is* likely beyond what humans can feasibly achieve. But nothing in behavioral economics (or in our commonsense understanding of human capacities) would seem to show that EU maximization, the *static* component of neoclassical rationality, is too difficult for ordinary humans.

One component of the scholarly literature on "decision analysis" is to develop procedures and tools to ensure that choice conforms to EU maximization (Keeney & Raiffa, 1993; von Winterfeldt & Edwards, 1986). The ordinary person can be trained to use these tools, or at least trained to recognize experts who can guide her in using them.

We do not think algebra is beyond ordinary human competence; high schoolers are given classes in the subject, and then go on to solve algebra problems themselves, or at least to know that there is a discipline, mathematics, with teachers, textbooks and websites that can help with algebra. Analogous points are true of EU maximization. When someone fails to behave in an EU fashion, we can point that out to her, and she can revise her choice. These mistakes do not necessarily undercut EU theory as the standard for *good* choice, any more than the pervasiveness of mathematical errors by ordinary folks somehow transmutes these errors into mathematical truth.

In short, someone who has read the findings of behavioral economics *might* react as follows: "I stick by my normative commitment to EU maximization, plus feasible norms of dynamic choice, as the criteria of rationality. These are *possible* norms of choice, and thus there is nothing contradictory in me endorsing them. To the extent that behavioral economics demonstrates that individuals often depart from EU behavior, this scholarship simply shows that individuals are – on my view of rationality – behaving irrationally."

What, in fact, is the position of behavioral economists about the norms of rational choice? Much work by behavioral economists does not engage the topic: after all, their expertise lies in psychological or social science, not in giving normative advice. Among those who do engage the topic, some do seem to view "behavioral" findings as instances of departures from rationality. (See generally Beshears, Choi, Laibson & Madrian, 2008, distinguishing between revealed and "normative" preferences; see also Viscusi & Gayer, 2016, in this symposium.) This was the stated position of Tversky and Kahneman in some of their early work (Tversky & Kahneman, 1988, 1992). Summarizing this work, Kahneman writes:

> [Tversky] called the theorists who tried to rationalize violations of utility theory "lawyers for the misguided." We went in another direction. We retained utility theory as a logic of rational choice, but abandoned the idea that people are perfectly rational choosers. We took on the task of developing a psychological theory that would describe the choices people make, regardless of whether they are rational. (2011, p. 314)

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

More recently, in his magnum opus *Thinking, Fast and Slow*, Kahneman distinguishes between the automatic, unconscious processing of "System 1" (subsuming the observed departures from neoclassical rationality), and the effortful, conscious deliberation of "System 2." Although Kahneman in this book oscillates somewhat in his normative position, he certainly at some points seems to endorse System 2, *not* System 1 thinking as the gold standard for rational choice. For example, he writes:

> How can we improve judgments and decisions, both our own and those of the institutions that we serve and that serve us? The short answer is that little can be achieved without a considerable investment of effort. As I know from experience, System 1 is not readily educable. Except for some effects that I attribute mostly to age, my intuitive thinking is just as prone to overconfidence, extreme predictions, and the planning fallacy as it was before I made a study of these issues. I have improved only in my ability to recognize situations in which errors are likely.
>
> The way to block errors that originate in System 1 is simple in principle: recognize the signs that you are in a cognitive minefield, slow down, and ask for reinforcement from System 2. (2011, p. 416)

Other behavioral economists reject the neoclassical view of rationality. The research agenda pursued in the joint work of Bernheim and Rangel is to develop norms for rational choice without reference to an underlying preference ranking over outcomes (Bernheim & Rangel, 2009; see also Bernheim, 2009, 2016). Some scholarship on Ellsberg choice (the so-called literature on "ambiguity") also seems to adopt the normative view that the absence of precise probability assignments is sometimes quite rational (Gilboa, Postlewaite & Schmeidler, 2012). Outside of behavioral economics, neoclassical rationality is challenged by philosophers who reject transitivity (Temkin, 2012) or, more deeply, consequentialism (Anderson, 1993).

In short, the view that rational choice is EU maximization plus feasible dynamic rationality is a normative position that cannot be shown to be "true" by virtue of empirical findings; reciprocally, someone who finds this position appealing can hold firm to it notwithstanding the findings of behavioral economics.

For the remainder of this article, given space limitations (and because it represents my own commitments!), I take as given the view of rationality just stated. For short, I will refer to this as an "EU" account of rationality. Surely, a normatively plausible account of rational choice will have *some* dynamic component. EU choice at each point in time (in each choice situation that the individual faces) will not be *sufficient* for rationality. However, on the view of rationality just stated, EU choice at each point in time is *necessary* for rationality. Moreover, my analysis

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

below does not depend on the specific content of dynamic rationality norms, and instead will focus on the departures from static, EU rationality that behavioral economics so richly evidences. With these caveats, I use the "EU" shorthand for the view of rationality as EU maximization in each choice situation plus compliance with feasible dynamic rationality constraints of some sort.

4 SWB-based policy analysis and theories of well-being

Let us now focus on SBPA. To be precise, what I mean by SBPA is a methodology for policy analysis where someone's answer to an SWB survey is taken as good evidence of her *well-being*. If SWB survey responses *do* have a strong evidentiary nexus to well-being, then tools such as cost-benefit analysis with compensating and equivalent variations in terms of SWB, a social welfare function with SWB numbers as input, or "gross national happiness" calculated using SWB scores, would be well justified.

"Well-being," again, is a normative term, and thus I consider the evidentiary value of SWB surveys through the lens of three different normative positions regarding well-being (for short "View A," "View B" and "View C"): (A) someone's well-being is what she would rationally self-interestedly prefer under favorable deliberative conditions, with "self-interested" preferences *not* restricted to preferences for mental states; (B) someone's well-being is what she would rationally self-interestedly prefer under favorable deliberative conditions, with "self-interested" preferences *restricted* to preferences for mental states; and (C) someone's well-being is conceptually independent of what she would rationally self-interestedly prefer under favorable deliberative conditions.

For each of these views of well-being, we can ask: in light of this view, does behavioral economics strengthen or weaken the case for SBPA?

4.1 View A

Recall that we are holding fixed an EU account of rationality. Combining the EU account of rationality with View A of well-being, we have that individual i's rational self-interested preferences over a given set of choices, under favorable deliberative conditions, can be represented as maximizing the expected value of some utility function $u_i(\cdot)$, taking outcomes as arguments. $u_i(\cdot)$ is such that: if individual i is EU-compliant and under favorable deliberative conditions self-interestedly prefers

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

outcome x to outcome y, then $u_i(x) > u_i(y)$. If individual i is EU-compliant and under favorable deliberative conditions self-interestedly prefers action a to action b, then $\sum_x \pi_a(x)u_i(x) > \sum_x \pi_b(x)u_i(x)$, with $\pi(\cdot)$ some probability measure and $\pi_a(x)$ the probability of x given action a.

For short, let us refer to EU-compliant preferences formed under favorable deliberative conditions as "idealized" preferences. Thus, $u_i(\cdot)$ is individual i's idealized self-interested preference-utility function.

For convenience (as is quite common in neoclassical economics), we will assume that $u_i(\cdot)$ is temporally separable, so that $u_i(x) = \sum_{t=0}^{T \max} D_t v_i(x_t)$, with x_t the facts about outcome x at time t. $v_i(\cdot)$ is individual i's *momentary* idealized self-interested preference-utility function. D_t is the discount factor for time t, either decreasing with time, or constant (no discounting), typically assumed to take the form $D_t = \frac{1}{(1+r)^t}$, $r \geqslant 0$. The v_i values of time slices, summed (after discounting) over time, add up to the overall utility values u_i which in turn represent the self-interested preferences over outcomes that individual i would have, were she to deliberate rationally under favorable conditions.

On View A of well-being, an individual can have "self-interested" preferences for features of the world other than her own mental states (Adler, 2012, Chapter 3; Adler, 2013, 2014). As already mentioned, how to define a "self-interested" preference is a normatively contested topic. While the example of self-sacrifice (see above) suggests that we do need a non-trivial "self-interest" screen – that we cannot simply equate someone's well-being with her all-things-considered idealized preferences – Robert Nozick's famous "experience machine" hypothetical suggests that we may find it normatively unappealing to define someone's "self-interested" preferences as preferences regarding what occurs "in her head." Nozick writes:

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprograming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences .... Would you plug in? (Nozick, 1974, pp. 42–44).

The lesson of Nozick's hypothetical is *not* that someone well-being is *independent* of what occurs "in her head." That would be absurd. If I am in terrible pain then, ceteris paribus, I am worse off. Rather, the experience machine hypothetical crystallizes the normative case against *limiting* the sources of someone's

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

well-being to what occurs “in her head.” It suggests that someone’s well-being should be allowed to depend not only upon her mental states (cognitions, affects, perceptions, evaluations, memories, etc.), but also upon features of her life and the world that are not wholly mentalistic (such as her health, her relationships, the degree to which she has attained her goals, her standing in the community, political liberties, knowledge and education).[10]

In particular, preferentialists about well-being may plausibly arrive at the following position: We should not define a “self-interested” preference so narrowly that my “self-interested” preferences are *required* to be preferences regarding my own mental states, and nothing else. However precisely the concept of “self-interest” is defined, that definition should *permit* me to have a “self-interested” preference for health, liberty, knowledge, goal fulfillment, good relationships, political standing, and other features of my life that are not wholly mentalistic.

Consider again function $v_i(\cdot)$ – individual i’s momentary idealized self-interested preference-utility function. Let $M_{i,t}(x)$ be a vector of individual i’s mental states at time t (her feelings, cognitions, memories, perceptions, etc.) and $N_t(x)$ non-mental features of outcome x at time t. On the view of well-being under consideration, $v_i(\cdot)$ takes the form $v_i(x_t) = v_i(M_{i,t}(x), N_t(x))$. If self-interested preferences were *required* to be preferences for mental states, we could use a simpler functional form for $v_i(\cdot)$, namely $v_i(x_t) = v_i(M_{i,t}(x))$. But View A of well-being declines to adopt such a restrictive understanding of “self-interest.”[11]

[10] These goods are not *wholly* mentalistic in that they are hybrids of mental and non-mental components. For example, knowledge is justified true belief: what one believes is a feature of one’s mental states, whether the belief is true is a matter of the outside world. Similarly, the quality of April’s relationship is a complex mixture of her mental states and external facts. For example, April may prefer a loving, truthful marriage; this means both that her husband does not cheat on her – a feature of his behavior rather than of April’s beliefs – and that she feel affection, happiness, etc. The experience machine hypothetical powerfully supports the view that well-being is *partly* non-mentalistic, not that it is *wholly* non-mentalistic. For a fuller discussion, see Adler, 2013.

[11] One possible defense of SBPA is to endorse View A of well-being, but then make the empirical claim that individuals’ self-interested preferences are in fact limited to preferences for their own mental states. In other words, individuals are permitted by View A to have a utility function of the form $v_i(M_{i,t}(x), N_t(x))$, but in fact generally have a utility function of the form $v_i(M_{i,t}(x))$.

Two weaknesses in this line of argument should be noted. (1) An emerging literature examines precisely this empirical issue – the extent to which individuals prefer SWB, as opposed to the non-SWB aspects of their lives (Adler et al., 2015; Benjamin, Heffetz, Kimball & Rees-Jones, 2012, 2014; Clark, Senik & Yamada, 2015; Perez-Truglia, 2015). The literature is small, and the results mixed. Thus it is far from established, as an empirical matter, that most individuals do in fact have a utility function of the form $v_i(M_{i,t}(x))$ rather than $v_i(M_{i,t}(x), N_t(x))$. (2) If this *were* true, the upshot would be to establish an empirical equivalence between View A and View B. According to View B, self-interest is *defined* so that an individual’s “self-interested” preferences are necessarily limited to preferences for her own mental states. But, as discussed below, even on View B of well-being, the case for SBPA is problematic.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

Given this view of well-being, is SBPA justified? Let $LS_i(x_t)$ be individual i's answer to a life-satisfaction survey at time t in outcome x.[12] Assume that $LS_i(x_t)$ is indeed good evidence of $v_i(x_t) = v_i(M_{i,t}(x), N_t(x))$, individual i's momentary idealized preference utility in outcome x. In expressing his current life satisfaction in a given outcome (x), the individual is expressing the preference utility of his current attributes in x. Adding those preference utilities over time, we get his overall (lifetime) preference utility in that outcome. If this were true – if it were true that $LS_i(x_t) \approx v_i(x_t)$ – then SWB (in particular, life-satisfaction) surveys *would* have a key role to play in policy analysis, given View A, as indicators of the basic measure of well-being ($u_i(\cdot)$ and $v_i(\cdot)$). Indeed, one suggestion in the SWB literature is that SWB surveys tell us about preference utility (Clark, Frijters & Shields, 2008).

However, for various reasons I have discussed at length elsewhere (Adler, 2013), $LS_i(x_t)$ is *not* reliable evidence of $v_i(x_t)$. One reason has to do with the interpretation of the life-satisfaction question: individuals may "read" the question as asking about feelings of happiness, not levels of preference utility. Second, individual i's preference utility is a function of a range of mental and non-mental attributes: all the features of the world at time t that individual i under idealized conditions would self-interestedly care about. But individual i, in answering a life-satisfaction question, may be focused on the attributes that are currently salient (for whatever reason attributes become salient). Unless individual i at time t in a given outcome x is carefully paying attention to the full range of inputs into his preference-utility function – to the full range of attributes in $(M_{i,t}(x), N_t(x))$ – $LS_i(x_t)$ will be overly responsive to the currently salient subset of attributes in $(M_{i,t}(x), N_t(x))$. Third, if $v_i(\cdot)$ is estimated by administering life-satisfaction surveys to a group of respondents consisting of different individuals, or the same individual at different times, on the assumption that all respondents have the same preferences – this would be done to increase the number of survey observations that are evidence of $v_i(\cdot)$ – problems of scaling arise. Two individuals, or the same individual at different times, may use different numerical scales to express the very same preferences.[13]

What role does behavioral economics play in this analysis of the connection between life satisfaction and preference utility? I suggest that behavioral economics

[12] My discussion here (and of View B, below) focuses on life-satisfaction surveys. If the goal is to use SWB surveys as evidence of preferences, then life-satisfaction surveys are presumably better suited to that goal than other types of SWB surveys (e.g., those asking about happiness, affects, or a sense of purpose).

[13] To be sure, traditional stated-preference or revealed-preference studies also typically rely upon evidence from multiple respondents, and embed some assumption of common preferences among the respondents. But these studies do not rely on the *further* assumption of a common numerical scale used by the respondent to express the common preferences. For detailed discussion, see Adler, 2013.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

weakens the nexus between $LS_i(x_t)$ and $v_i(x_t)$. If behavioral economics is correct, then $LS_i(x_t)$ is even *poorer* evidence of $v_i(x_t)$ than it would be if individuals, in their actual behavior, behaved as predicted by neoclassical models. Why? $v_i(\cdot)$ represents individual i's rational self-interested preferences under favorable conditions: rational in the EU sense. We ask: how *would* individual i rank outcomes and choices, were she to have good information, be self-interested, and conform to the axioms of EU theory? But behavioral economics demonstrates that individuals pervasively depart from EU theory in their actual choices. EU choice is not a feature of the automatic, unconscious, System 1 processes that drive ordinary life. In order to come into compliance with EU axioms, individuals need to engage in effortful, conscious, System 2 choice and (very likely) training in probability theory and decision analysis (Kahneman, 2011).

One could, in theory, imagine a life-satisfaction survey administered only *after* "debiasing" steps designed to help respondents grasp and satisfy EU norms of rationality. But (as far as I am aware) most life-satisfaction surveys are not preceded by such coaching.

A life-satisfaction survey or some other SWB survey, without debiasing, can only provide evidence about the respondent's current (non-debiased) attitudes. But behavioral economics tells us that a preference-utility function satisfying EU theory is a *constructed* object. Individuals do not naturally have this object "in their heads"; they must engage in deliberate, effortful steps, so that their preferences focus on final outcomes (not changes), are complete and transitive, and otherwise measure up to the coherence conditions expressed by the EU axioms. Further, on View A of well-being, the measure of an individual's welfare is a preference-utility function that satisfies the EU axioms (as well as being self-interested and meeting the requirement of "favorable deliberative conditions"). Thus, if behavioral economics is true (as a scientific matter) and if View A of well-being is adopted (as a normative matter), someone's response to an SWB survey (without debiasing) can hardly be taken as the measure of her well-being.

The reader might object that $LS_i(x_t)$ needs only to be "good enough" evidence of $v_i(x_t)$ – rather than "good" or "reliable" evidence in some more robust sense – in order to be used by policymakers as the measure of $v_i(x_t)$. However, behavioral economics shows there to be a critical gap between the two. The idealized preference-utility function $v_i(\cdot)$ is a numerical representation of preferences that are *rational* in the sense of being EU-compliant, while – behavioral economics shows us – respondents to SWB surveys without debiasing should *not* be expected to be rational in that sense. Why think that someone's verbal evidence expressed in a condition falling short of rationality is "good enough" evidence of what she would want, if rational? Complaints about the "infeasibility" of EU maximization are of no avail, here. First, EU maximization (like algebra) *is* feasible, if not natural and easy, for ordinary humans.[14] Second, if this is untrue (if EU maximization is more like quantum physics than algebra) the upshot should be to drop EU-compliant preferences as part of our conception of well-being – to cease defining well-being in terms of the idealized preference-utility functions $u_i(\cdot)$ and $v_i(\cdot)$ – and *not* to take $LS_i(x_t)$ as "good enough" evidence of $v_i(x_t)$.

4.2 View B of well-being

Assume that we are not persuaded by Nozick's "experience machine," and find it normatively attractive, on balance, to define "self-interested" preferences as preferences for mental states. (On this possibility, see Adler, 2013; Sumner, 1996, Chapter 4.) On this view of "self-interest," the only inputs to a utility function representing i's self-interested preferences are i's cognitions, feelings, perceptions, etc. $v_i(\cdot)$ is the momentary utility function representing the preferences that individual i would have, were i to be EU-rational, deliberating under favorable conditions, and self-interested. Thus, on the mentalistic view of self-interest, $v_i(x_t)$ takes the form $v_i(M_{i,t}(x))$, not $v_i(M_{i,t}(x), N_t(x))$. We might (if we are careful) refer to $v_i(\cdot)$ as an "experience utility" function – but only with care, since this is "experience" utility *only* in the sense of being an idealized preference-utility function with an extra structural constraint, namely that it must take the preference-holder's mental states as its sole arguments.

Assuming $v_i(\cdot)$ takes the form $v_i(M_{i,t}(x))$ rather than $v_i(M_{i,t}(x), N_t(x))$, can we *now* take an individual i's answer to an SWB survey at time t in outcome x as good evidence of $v_i(x_t)$? In brief, no. To begin, $M_{i,t}$ is a *vector* of mental characteristics. An individual's complete "state of mind," at any point in time, is a multidimensional bundle, comprised of many types of psychological attributes: sensations of pain and pleasure, emotions of different sorts, valuations, cognitions, perceptions, memories, etc. (Bernheim, 2016, in his contribution to this symposium, makes a similar observation.) Individuals can have heterogeneous rankings of these multidimensional psychological bundles – giving more weight to one or another dimension. Sam might prefer a mental life with lots of pleasure, even if

[14] See above, pp. 7–8. Thus, "good enough" evidence of $v_i(x_t)$ will be some subset of survey or behavioral evidence in which individuals have been primed to engage in System 2 thinking and to construct EU-compliant preferences. I say "some subset" because $v_i(\cdot)$ is a numerical representation of preferences that are both rational *and* satisfy further criteria of "favorable deliberative conditions" and "self-interest"; and also because we need to be assured that the preference-elicitation setup is not undermined by problems of incentive compatibility (where the setup is such that individuals have an incentive to hide their preferences).

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

his thinking is fuzzy, his memories disconnected, and his perceptual experiences fairly monotonous. Sheila might care a lot about having crisp cognitions and a rich perceptual life, even at the cost of some pleasure. Griffin provides the following example to illustrate that there is not one dimension of psychological life that necessarily drives human preferences for psychological bundles: "At the very end of his life, Freud, ill and in pain, refused drugs except aspirin. 'I prefer,' he said, 'to think in torment than not to be able to think clearly.'" (Griffin, 1986, p. 8).

Understanding that $M_{i,t}$ is multidimensional, is an individual's answer to an SWB survey (in particular, a life-satisfaction survey) good evidence of $v_i(M_{i,t}(x))$ – of idealized momentary self-interested preference utility according to View B? There is admittedly *one* respect in which $LS_i(x_t)$ is better evidence of $v_i(M_{i,t}(x))$ as contrasted with $v_i(M_{i,t}(x), N_t(x))$. With the general form $v_i(M_{i,t}(x), N_t(x))$, we might worry that an individual will not be paying attention to the N_t inputs – to the non-mental attributes that (according to View A) are allowed to be a determinant of self-interested preference utility. On View B of well-being, a failure of attention with respect to non-mental attributes does not undermine the role of life-satisfaction surveys in evidencing $v_i(\cdot)$, since only mental attributes are inputs into $v_i(\cdot)$.

However, all the other problems discussed above in using a life-satisfaction survey as evidence of $v_i(\cdot)$ remain in place. In particular, the key point remains that $v_i(\cdot)$ is supposed to satisfy EU theory. But behavioral economics tells us, again, that individuals do not walk around with EU-compliant preference-utility functions already "in their heads." Rather, arriving at these functions typically involves effort, conscious thought, training, etc. $v_i(\cdot)$ would be generated, specifically, by coaching individuals in the norms of EU theory; otherwise providing favorable deliberative conditions; and instructing them to rank outcomes *by focusing on the psychological bundles they have in outcomes* (how View B defines "self-interest"). Behavioral economics should make us dubious that a life-satisfaction survey administered to non-debiased respondents (even a "tweaked" such survey focusing on psychological life) will tell us much about $v_i(\cdot)$.

4.3 View C of well-being

As mentioned above, a long philosophical tradition adopts an objective-good view of well-being (reviewed in Adler, 2012, Chapter 3). There is, to be sure, plenty of disagreement within this tradition. "Objective-good" theorists disagree about which goods are on the list. For example, Finnis' list is life, knowledge, play, aesthetic experience, sociability, practical reasonableness, and religion (Finnis, 1988, Chapter 4). Nussbaum's is life; bodily health; bodily integrity; the senses, imagination,

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

and thought; emotions; practical reason; affiliation; other species; play; and control over one's environment (Nussbaum, 2000, pp. 78–80). Sher's is moral goodness; rational activity; the development of one's abilities; having children and being a good parent; knowledge; and the awareness of true beauty (Sher, 1997, p. 201). Griffin's is accomplishment; autonomy; physical integrity; understanding; enjoyment; and deep personal relations (Griffin, 1997, pp. 29–30).

Objective-good theorists also offer divergent *rationales* for placing goods on the list. One rationale is in terms of the human essence: goods are the realization of those capacities that are essential to human beings. A second appeals to commonly shared judgments or intuitions of well-being: objective goods are those things which people, under the right conditions, normally perceive as good for human welfare. A third possibility is for the theorist to rely upon her own judgments or intuitions regarding human welfare – without claiming that these are necessarily widely shared.

Leaving aside these (significant) internal disputes, objective-good accounts are alike in *severing* the conceptual link between a particular person's welfare and her preferences.[15] The list of goods (whatever its precise content) is such that Jim *can* be on balance better off in y than x even though Jim prefers x – for that matter, even though Jim under favorable deliberative conditions rationally self-interestedly prefers x.[16] Note that all three of the rationales for objectivism allow for a divergence between someone's objective goods and her actual or idealized preferences. This is clear for the first (human essence) and third (theorist's own conception) rationale. As for the second rationale (shared human judgments), consider the case of idiosyncratic preferences. Jim might actually and/or ideally self-interestedly prefer x to y, even though most humans believe that Jim's life in y is the better one.

How does behavioral economics enter the analysis? It does so indirectly. By demonstrating the wide gap between someone's *actual* preferences and her *rational* preferences, behavioral economics may undercut the normative appeal of preference-based views of well-being. Preference views are traditionally defended

[15] To be clear, an objective-good account says that well-being is *conceptually* (definitionally) independent of preferences. It denies a necessary connection between the two. According to an objective-good account, it is *possible* for individual i to prefer outcome x to outcome y and yet to be better off in y. Moreover, this is so even if we require the preferences to be rational, formed under favorable deliberative conditions, and/or self-interested.

However, an objective-good view certainly need *not* say that well-being is *statistically* independent of what individuals prefer. That is, it need *not* deny an empirical correlation between preference satisfaction and the realization of the objective goods. The posited list of goods might be such that individuals are often motivated to attain them; if so, statistical independence would not hold true.

[16] Note that this definition allows for some goods to be defined in terms of the individual's actual or idealized preferences as long as the overall balance is conceptually independent of the individual's preferences.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

by appealing to notions of individual *sovereignty*. Shouldn't Sarah be able to decide for herself what makes her life go best? At least if Sarah is a normal adult human (with the capacity for autonomous choice), shouldn't *Sarah*'s well-being depend upon *Sarah's* perspective – what Sarah likes, wants, judges, thinks, endorses, etc.? Ceteris paribus, a normatively appealing feature of an account of well-being is that the account is non-paternalistic. The normative evaluation of Sarah's life defers to Sarah's own evaluation.

On the other hand, for reasons I adverted to earlier, a preference-based account of welfare should very plausibly be oriented to an individual's *rational* preferences. But *any* definition of well-being in terms of rational preferences will itself be paternalistic, to a certain extent – if rationality is understood in the EU sense. This is what behavioral economics shows us. Consider, in particular, the definition of *Sarah's* well-being in terms of *Sarah's* EU-rational preferences under favorable deliberative conditions and conditions of self-interest. *However* we define the latter two components, there will be a gap between Sarah's well-being and the non-EU-rational preferences that motivate her day-to-day behavior. Sarah's EU-rational perspective – what she wants when she regiments her preferences so as to comply with EU theory – is not the same as her day-to-day, System 1 perspective. Defining Sarah's well-being in terms of her EU-rational preferences *is* in one sense non-paternalistic (since it defers to the EU-rational perspective of *Sarah*) but, in another sense, is not (since it overrides her day-to-day perspective).

What behavioral economics show us is that an account cannot be non-paternalistic in *both* senses. And – here's the rub – by weakening the non-paternalistic credentials of a rational-preference view of welfare, behavioral economics might indirectly support an objective-good view of welfare. That is, the ethical deliberator might find herself at the following juncture in thinking about which view of well-being she wishes to endorse: "I find some appeal in an objective-good view of well-being. On the other hand, I am also moved by the thought that a view of well-being should be non-paternalistic. Behavioral economics now shows me that a preference-based well-being view cannot be *fully* non-paternalistic (at least if the view appeals to the EU view of rationality, which I also accept, and to *rational* preferences). It shows me that the considerations in favor of the preference view were weaker than I thought. Balancing the pros and cons of objective-good versus preference views of well-being, I now support an objective-good view."

While behavioral economics could in this way provide indirect, dialectical support for an objective-good view of well-being, we now need to ask about SBPA. On an objective-good view of well-being (View C), to what extent is an SWB survey good evidence of an individual's well-being? Well-being, here, means, the

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

overall quality of the individual's life in light of her attainment with respect to each of the objective goods specified by View C and in light of the appropriate balancing of such attainments (as this balancing is specified by View C).

One possibility is that View C includes some goods that are not mentalistic: goods that are not reducible to an individual's mental states. Nozick's experience machine argues powerfully in favor of a View C of this sort.

It is very hard to see how an SWB survey provides strong evidence regarding an individual's attainment of a good that is not mentalistic.[17] Consider the good of knowledge (a standard objective good – see the lists above). Someone's knowledge depends not merely on what she believes, but on whether those beliefs are *true*. It depends on her beliefs *and* the degree of correspondence between those beliefs and the outside world. Asking someone how happy she is, or how satisfied she is overall with her life, or specifically how satisfied she is with her knowledge of the world, is not going to track very well the extent to which she has *correct* beliefs.

Nor will an SWB survey tell us much about how an individual fares with respect to the overall balance of mentalistic and other goods. That balance will be determined (as per View C) by some criterion external to the person's own preferences and valuations (such as the human essence, commonly shared judgments of well-being, or some other objective criterion offered by the view).

A second possibility is that View C consists of a list of goods, all of which are mentalistic. For example, one good might be happiness; the second, the quality of someone's memories (Kahneman, 2011, part V). SWB surveys or other psychological surveys can be designed to tell us about how someone fares with respect to each good – but the problem of balancing the goods remains.

Psychological surveys would seem to offer strong evidence of well-being only according to a distinctive version of View C: one such that well-being consists in a single, mentalistic good. If the mentalistic good is the kind of affective or valuational state that SWB surveys are designed to measure – feeling happy, feeling satisfaction with life, having a sense of purpose – then someone's answer to an SWB survey can indeed be good evidence of his well-being as per this version of View C.

The supposition of an objective-good view of this sort is not ludicrous. Consider Benthamite hedonism, understood as follows. A person, at any moment in time, experiences some level of positive affect (pleasure) or negative affect (pain), measurable on a ratio scale with zero for the neutral level, and higher numbers indicating a more favorable affect (more pleasurable if above neutral, less painful if below neutral). This momentary affective level is an introspectible psychological

[17] To be clear about the terminology: goods that are not mentalistic may well be hybrids of an individual's mental attributes and features of the external world, while mentalistic goods depend just upon the individuals' mental attributes. See above note 10.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

magnitude: individuals can reliably detect whether they are feeling sensations of pain or pleasure, and how intense those sensations are. An individual's well-being in any given outcome is simply the lifetime sum of momentary affective levels (Bronsteen et al., 2015; see also Layard, 2011, defending monistic hedonism; on the measurement of affect, see Kahneman, Wakker & Sarin, 1997).

Benthamite hedonism (thus understood) is clearly mentalistic. Moreover, it is objective: maximizing the lifetime sum of momentary affective levels is posited as the well-being-maximizing course of action for any individual, regardless of whether the individual herself prefers to maximize this affective magnitude – regardless of whether she also cares about other dimensions of mental life (memories, perceptions, cognitions, valuations) or about non-mental attributes. Finally, Benthamite hedonism supports SBPA. Traditional happiness surveys offer some (albeit imperfect) evidence of an individual's average affective quality during some stretch of time prior to the survey; and even better surveys, using experience sampling or "day reconstruction" to estimate moment-to-moment affects, have been implemented by SWB scholars.

But note carefully the special structural features of Benthamite hedonism that enable SWB surveys to be good evidence of well-being. First, SWB surveys are informing us about the level of individual affects, not about the satisfaction of individual preferences. Second, affective quality is posited as the sole component of well-being. The normative deliberator who adopts this conception of welfare has not only rejected rational-preference accounts (View A and B) in favor of an objective-good account (View C). She, further, rejects both pluralism about goods and the existence of goods that are not mentalistic. Although this version of View C is certainly *possible*, it is far from obvious why we should find it normatively attractive.

5 Conclusion

We have reached generally skeptical conclusions about using someone's answer to an SWB survey as evidence of her well-being. Two families of views of well-being (A and B) identify well-being with the realization of idealized self-interested preferences – idealized in the sense of being rational and produced under favorable deliberative conditions. The views differ in how restrictively they define "self-interest." Regardless, on both these views, if "rational" means "EU-rational," the measure of someone's well-being is an idealized *preference-utility* function representing the preferences she would have if compliant with EU theory. Behavioral economics now tells us that individuals, without conscious, careful effort, do not

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

naturally comply with EU theory. Thus an individual's answer to an SWB survey (at least one elicited without initial interventions prompting the respondent to come into compliance with EU theory) does not tell us much about her idealized preference utility – about her level of well-being as per View A or B.

Frustrated with the difficulties of measuring idealized preference utility, the ethical deliberator might shift to an objective-good account of welfare. Now, the measure of someone's well-being is his overall balance of attainments with respect to the stipulated goods – as determined by some criterion external to the person's preferences. Except in the special case of a Benthamite objective-good account consisting of a single good, happiness, an SWB survey will not provide good evidence of that overall attainment.

Throughout this article, the question has been whether an SWB survey provides evidence of someone's *well-being* (idealized preference utility or the overall balance of goods, depending on the view). But, in closing, we should mention a different and more modest role for SWB surveys in guiding policy. Happiness or similar affective or valuational states could be a *component* of well-being. On a rational-preference account of well-being (View A or B), happiness could well be an *argument* in the preference-utility functions of most people. And, very plausibly, if well-being is taken to be objective then happiness is *one* of the objective goods.

SWB surveys, in turn, provide evidence about *this* component of well-being (not well-being all dimensions considered). Consider the analogy to evidence of an individual's health. Health is surely *one* thing that many individuals prefer (Adler, Dolan & Kavetsos, 2015). And health, plausibly, is *one* objective good. Self-assessments or physician reports are evidence of an individual's health, and can be expressed in numerical form (as with QALYs). But it would be a leap to redefine cost-benefit analysis in terms of compensating or equivalent variations with respect to health (rather than standard willingness-to-pay or -accept amounts), or to design policy to maximize social welfare with health utility (rather than preference utility) as the input. These methodologies would commit a part-whole fallacy – conflating someone's health with her all-dimensions-considered well-being.

Similarly, SWB surveys tell us a lot about whether things are feeling good or bad for people – about how life is going in terms of affects or a sense of satisfaction – but we should not make the normative leap from this useful information about one *aspect* of well-being, to SBPA.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

References

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

Adler, Matthew D. (2013). Happiness Surveys and Public Policy: What's the Use? *Duke Law Journal*, *62*, 1509–1601.

Adler, Matthew D. (2014). Extended Preferences and Interpersonal Comparisons. *Economics and Philosophy*, *30*, 123–162.

Adler, Matthew D., Dolan, Paul & Kavetsos, Georgios (2015). Would You Choose to Be Happy? Tradeoffs between Happiness and the Other Dimensions of Life in a Large Population Survey. CEP Discussion Paper No. 1366. http://cep.lse.ac.uk/pubs/download/dp1366.pdf.

Anderson, Elizabeth (1993). *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Anscombe, F. J. & Aumann, R. J. (1963). A Definition of Subjective Probability. *The Annals of Mathematical Statistics*, *34*, 199–205.

Atkinson, Anthony B. (2009). Economics as a Moral Science. *Economica*, *76*, 791–804.

Benjamin, Daniel J., Heffetz, Ori, Kimball, Miles S. & Rees-Jones, Alex (2012). What Do You Think Would Make You Happier? What Would You Choose? *American Economic Review*, *102*, 2083–2110.

Benjamin, Daniel J., Heffetz, Ori, Kimball, Miles S. & Rees-Jones, Alex (2014). Can Marginal Rates of Substitution be Inferred from Happiness Data? Evidence from Residency Choices. *American Economic Review*, *104*, 3498–3528.

Bernheim, B. Douglas (2009). Behavioral Welfare Economics. *Journal of the European Economic Association*, *7*, 267–319.

Bernheim, B. Douglas (2016). The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics. *Journal of Benefit-Cost Analysis*.

Bernheim, B. Douglas & Rangel, Antonio (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics*, *124*, 51–104.

Beshears, John, Choi, James J., Laibson, David & Madrian, Brigitte C. (2008). How are Preferences Revealed? *Journal of Public Economics*, *92*, 1787–1794.

Bronsteen, John, Buccafusco, Christopher & Masur, Jonathan S. (2015). *Happiness and the Law*. Chicago: University of Chicago Press.

Camerer, Colin F., Loewenstein, George & Rabin, Matthew (Eds.) (2004). *Advances in Behavioral Economics*. Princeton: Princeton University Press.

Cartwright, Edward (2011). *Behavioral Economics*. New York, NY: Routledge.

Chetty, Raj (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review: Papers & Proceedings*, *105*, 1–33.

Clark, Andrew E., Frijters, Paul & Shields, Michael (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, *46*, 95–144.

Clark, Andrew E., Senik, Claudia & Yamada, Katsunori (2015). When Experienced and Decision Utility Concur: The Case of Income Comparisons. IZA Discussion Paper No. 9189. http://hdl.handle.net/10419/114057.

Congdon, William J., Kling, Jeffrey R. & Mullainathan, Sendhil (2011). *Policy and Choice: Public Finance through the Lens of Behavioral Economics*. Washington, D.C.: Brookings Institution Press.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

DellaVigna, Steffano (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, *47*, 315–372.

Dolan, Paul & Kahneman, Daniel (2008). Interpretations of Utility and their Implications for the Valuation of Health. *Economic Journal*, *118*, 215–234.

Finnis, John (1988). *Natural Law and Natural Rights*. Oxford: Clarendon Press.

Fujiwara, Daniel & Dolan, Paul (2016). Happiness-Based Policy Analysis. In Matthew D. Adler & Marc Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy* (pp. 286–317). New York: Oxford University Press.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.

Gilboa, Itzhak (2009). *Theory of Decision Under Uncertainty*. Cambridge: Cambridge University Press.

Gilboa, Itzhak, Postlewaite, Andrew & Schmeidler, David (2012). Rationality of Belief: Or, Why Savage's Axioms are Neither Necessary nor Sufficient for Rationality. *Synthese*, *187*, 11–31.

Graham, Carol (2016). Subjective Well-Being in Economics. In Matthew D. Adler & Marc Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy* (pp. 424–450). New York: Oxford University Press.

Griffin, James (1986). *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Griffin, James (1997). *Value Judgment: Improving Our Ethical Beliefs*. Oxford: Oxford University Press.

Jeffrey, Richard C. (1983). *The Logic of Decision*. (2nd ed.). Chicago: University of Chicago Press.

Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Joyce, James M. (2004). Bayesianism. In Alfred R. Mele & Piers Rawling (Eds.), *The Oxford Handbook of Rationality* (pp. 132–155). Oxford: Oxford University Press.

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel & Tversky, Amos (Eds.) (2000). *Choices, Values, and Frames*. Cambridge: Cambridge University Press.

Kahneman, Daniel, Wakker, Peter, P. & Sarin, Rakesh (1997). Back to Bentham: Explorations of Experienced Utility. *Quarterly Journal of Economics*, *112*, 375–405.

Kaplow, Louis & Shavell, Steven (2002). *Fairness versus Welfare*. Cambridge, MA: Harvard University Press.

Keeney, Ralph L. & Raiffa, Howard (1993). *Decisions with Multiple Objectives*. Cambridge: Cambridge University Press.

Kreps, David M. (1988). *Notes on the Theory of Choice*. Boulder: Westview Press.

Layard, Richard (2011). *Happiness: Lessons from a New Science*. (Revised edition). London: Penguin Books.

Machina, Mark (2014). Ambiguity Aversion with Three or More Outcomes. *American Economic Review*, *104*, 3814–3840.

Madrian, Brigitte C. (2014). Applying Insights from Behavioral Economics to Policy Design. *Annual Review of Economics*, *6*, 663–688.

Mas-Collel, Andreu, Whinston, Michael D. & Green, Jerry R. (1995). *Microeconomic Theory*. New York: Oxford University Press.

Nozick, Robert (1974). *Anarchy, State and Utopia*. New York, NY: Basic Books.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press

Nussbaum, Martha (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Perez-Truglia, Ricardo (2015). A Samuelsonian Validation Test for Happiness Data. *Journal of Economic Psychology*, *49*, 74–83.

Robinson, Lisa A. & Hammitt, James K. (2011). Behavioral Economics and the Conduct of Cost-Benefit Analysis: Towards Principles and Standards. *Journal of Benefit-Cost Analysis*, *2*, 1–51.

Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: John Wiley and Sons.

Sher, George (1997). *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press.

Strotz, R. H. (1956). Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies*, *23*, 165–180.

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

Sunstein, Cass R. (2016). Cost-Benefit Analysis: Who's Your Daddy? *Journal of Benefit-Cost Analysis*.

Temkin, Larry S. (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Thaler, Richard & Sunstein, Cass (2009). *Nudge*. New York, NY: Penguin Books.

Tversky, Amos & Kahneman, Daniel (1988). Rational Choice and the Framing of Decisions. In David E. Bell, Howard Raiffa & Amos Tversky (Eds.), *Decision Making: Descriptive, Normative, and Prescriptive Interactions* (pp. 167–192). Cambridge: Cambridge University Press.

Tversky, Amos & Kahneman, Daniel (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Viscusi, W. Kip & Gayer, Ted (2016). Rational Benefit Assessment for an Irrational World. *Journal of Benefit-Cost Analysis*.

von Winterfeldt, D. & Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/bca.2016.4 Published online by Cambridge University Press