

Is Well-being Measurable After All?

Anna Alexandrova*, Department of History and Philosophy of Science, University of Cambridge

*Corresponding author: Anna Alexandrova, Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK. Email: aa686@cam.ac.uk

In *Valuing Health*, Dan Hausman argues that well-being is not measurable, at least not in the way that science and policy would require. His argument depends on a demanding conception of well-being and on a pessimistic verdict upon the existing measures of subjective well-being. Neither of these reasons, I argue, warrant as much skepticism as Hausman professes.

Introduction

Dan Hausman's *Valuing Health* is a landmark text in philosophy of health, philosophy of social science and political philosophy. Its novelty and importance is to show the value of engaging all three of these projects at once, especially when this is done with Dan's characteristic depth and thoroughness. This formula of philosophical research is here to stay.

The individual arguments of this book are also very important. Hausman's main goal in the book is to challenge the reigning paradigm of measurement of health by its contribution to individual well-being, and to sketch an alternative. In the view he advocates, healthcare allocation is a matter of the public instead of the private value of health: public value requires a special public justification, a justification that goes beyond the benefits of health to the individual. This paradigm-shifting argument will deservedly receive the lion share of attention and likely set the agenda for philosophy of public health for years to come.

One premise of this argument is that well-being is not measurable. This premise, which is my focus here, is not actually essential to Hausman's main goal. Those who reject well-being as a legitimate political goal in the first place, will agree that access to healthcare should not depend on its importance for well-being, whether or not well-being is measurable. Yet, that is not the route Hausman takes. For him, non-measurability of well-being is one of the reasons why it should not be the basis of valuation of health. And even if this premise did not feature in his argument, it warrants attention. If well-being is not measurable, that is a big deal both for science and for politics. On the political side, if the property of well-being is not a proper quantity, then it is

harder to use it as a benchmark for evaluating policy proposals. On the scientific side, it spells doom for the ambitious and newly resurgent project of investigating causes and consequences of well-being. In my estimation, Hausman's challenge to measurement of well-being is the most compelling.¹ Moreover, it comes at the right time as medical professionals are also raising doubts about the well-being agenda that has entered into public health and clinical research in recent decades.² If Hausman and other skeptics are right, then this well-being agenda is due for retirement.

Hausman's argument, which he makes in chapters 6–11 of *Valuing Health* and which I reconstruct in the first half of this essay, is roughly that well-being is too person-relative to measure reliably. Less roughly, he relies first on some a priori considerations about the concept of well-being—it is a concept that calls for aggregation of goods in a person's life in a way that is duly sensitive to who this person is. Second, Hausman is skeptical that the existing measures of well-being—whether focused on feelings of happiness, or life satisfaction—aggregate all the well-being-relevant goods in a way that respects individuality, while also allowing for population-level generalizations. Comparing and ranking well-being states is possible, albeit hard, on an individual level, Hausman concedes, but becomes largely unrealistic on societal level.

My claim here is that this argument does not doom the project of measurement of well-being. In the second half of this essay, I explore two avenues for resistance. I start by noting that well-being as a concept has multiple meanings. Given this multiplicity, first, we could question whether Hausman's demanding sense of well-being that calls for a comprehensive aggregation of all goods over the course of one individual's life is the right notion

for science and policy. Second, we could put pressure on Hausman's critique of existing measures. To undermine these measures, Hausman gives intuitively plausible reasons why they should fail to capture well-being. But these intuitions can be very compelling and still fail to undermine a measurement tool if this tool systematically behaves in reliable ways consistent with the empirical knowledge of factors surrounding well-being. The latter is the conceit behind *construct validation*, the main approach to evaluating measures in social and medical sciences, which Hausman does not discuss and which, I expect, most practitioners will appeal to, in reply to him.

Although Hausman's skeptical verdict is not warranted on the basis of reasons he gives, his challenge is a deep one and unlikely to go away. At the end of the day, whether well-being is measurable depends on how flexible we are willing to be about what counts as well-being and what counts as valid measurement.

Hausman's Case against Measurability of Well-Being

Hausman puts forward his case in the context of exploring whether health should be valued by its contribution to well-being—a question essential for deciding how a community should allocate its scarce resources for healthcare. How bad is it to have a broken ankle? A natural answer is that a broken ankle is as bad as the resulting loss of well-being of that ankle's owner. This is the view that Hausman will eventually reject (see chapter 12 for a summary of reasons). One of his grounds is that a broken ankle can have a dramatically different impact on a person's well-being depending on who they are. In Hausman's own case, a broken ankle enabled him to write this book. Such fine-grained heterogeneity cannot be accommodated by any realistic population-level healthcare policy. This, among other reasons, is why communities need to look for an allocation rule that is not based on well-being. As we shall see shortly, independently of healthcare concerns, heterogeneity is the main obstacle to well-being measurement more generally.

To make an argument against measurability of well-being, Hausman first needs to say at least in broad terms what well-being is. Although he does not articulate a full theory, the outlines are clear enough. Hausman believes that any account of well-being should accommodate the following constraints (pp. 121–124)³:

(1) Well-being consists in several goods, not one.

- (2) 'What is good for me depends heavily on who I am' (p. 121), that is, an agent's well-being depends on the agent's goals, values and identity.
- (3) 'In assessing well-being we think primarily of whole lives, and our appraisal of how well someone's life is going during a limited period often depends on what their life is like before or after' (p. 122).
- (4) Well-being is holistic in that adding more of some valuable good does not necessarily improve the whole. It is the combination that matters (p. 124).

Hausman is well aware that this conception is not entirely uncontroversial. Hedonists, for instance, argue that there is only one non-instrumental good—positive mental states. If well-being is directly measurable, these states are the only things that need measurement. Hausman does not hide his rejection of hedonism, and indeed of the other major theory, subjectivism⁴—neither bare feelings nor the fulfillment of desires or goals captures what it is to live well. In chapter 11 he lays his cards on the table and backs a theory of well-being based on flourishing, aligning his views closely to those of Richard Kraut (2007):

A fundamental evaluation of the value of some property or state of affairs for an individual depends on how the things that make human lives good (such as friendship, happiness, health, or a sense of purpose) are integrated into the dynamic structure of that individual's life. What Kraut and I call 'flourishing' consists in the dynamic coherent integrations of objective goods into an identity. Well-being is flourishing. (p. 141)

To argue that flourishing is not measurable, Hausman relies on a plausible conception of measurement—existence and epistemic access to a numerical scale that enables comparisons of all well-being states across and within persons, and the distances between these states (Hausman's Section 4.3). This is known as an interval scale. Now it is easy to see how the case against measurability would go. Different people's flourishing is made up of different goods that combine in unique ways depending on their place in people's lives. No single scale applicable to all persons can capture the success of such unique combinations, and so, comparisons, let alone on an interval scale, are hopeless.

Hausman rehearses this argument but then recoils from its extreme skepticism. It is clear that sometimes comparisons are possible and very compelling—it is better not to die very young and not to become a child soldier, he agrees (p. 125).

It is equally true that sometimes feelings and preferences are decent guides to how well we are doing. Adherents of idealized subjectivism hold that, were an agent to form desires in light of full knowledge and with no mistakes in reasoning, these desires would be authoritative about what is good for this agent. Hausman is not a subjectivist, but he helps himself to the idea that some preferences are more authoritative than others. In particular, preferences can reliably indicate flourishing when these preferences are *laundered* in the right way (Section 10.4). Laundered preferences are those held by individuals who are '(i) self-interested, (ii) well-informed, (iii) evaluatively competent, and (iv) free of deliberative defects, and if (v) they have complete and transitive preferences among all alternatives' (p. 132). When cases are sufficiently clear-cut or when preferences are sufficiently laundered, comparisons, even measurement, are possible.

This allowance notwithstanding, Hausman still ends up with a skeptical conclusion, albeit less extreme. The reality in science and in public policy is that hard cases abound: who should get the scarce resources—the ones with broken ankles or the mildly depressed? Plus, the indirect measures of well-being—happiness-based, or preference-based—are very poor at their task.

Hausman comes down especially hard against subjective well-being measures so popular in today's psychology and behavioral economics. On this methodology, people are either asked to judge their overall satisfaction with life—an invitation to express their satisfaction with all things important to them and then integrate these considerations into a summary judgment. Or alternatively, psychologists gauge subjects' positive and negative affect, calculating the net affect using some averaging rule. About life satisfaction, Hausman complains that people's reports are unreliable and sensitive to irrelevant details (9.2 and p. 129). On net affect, Hausman notes that these measures count all emotions equally. He asks who decides how net affect is calculated—why should my sadness at reading international news count for as much as my sadness at losing a grandparent? 'Heterogeneity goes all the way down to feelings' (p. 129), he insists. Together, these considerations show in his view that relying on subjective evaluation is too risky, because the precise impact of the quality of subjective experience for overall well-being is a personal matter.

Preferences, especially the laundered ones, would be in a better shape, *if* they were measured. But the fact is that standard economic methodology either infers preferences from choices people make (that is the *revealed* preference approach) or else from their responses to

questionnaires about what state of affairs they would prefer and at what rate (that is the *stated* preference approach). Neither approach makes an effort to select among these choices or stated preferences only the authoritative ones.

Here then is Hausman's tempered skeptical conclusion:

Our evaluative abilities are limited with respect to our own lives, and the limits to those abilities imply limits to the completeness of our rankings of alternatives. It will often be the case that the objective of enhancing people's well-being does not discriminate among alternatives. As a practical matter, policy-makers will need other ways of comparing alternatives, and as a theoretical matter, either one has to conclude that prudence and ethics are less discriminating than previously thought or that normative notions other than well-being must play a large role. (p. 142)

It is a tempered conclusion in that Hausman allows for uncontroversial comparisons of starkly different well-being states in individual cases—it is indeed better for him to lead the life he leads than to have become a child soldier. But, in general, these rankings will not be possible. We might formalize his argument as follows:

Premise 1: Well-being, being an inclusive good, allows for much heterogeneity in how component goods are integrated.

Premise 2: The existing measures are at best fallible indicators that do not respect heterogeneity of well-being.

Premise 3: Policy-relevant measures require a systematic population-level connection between well-being and the indicators.

Conclusion: Therefore, well-being is not measurable for policy purposes.

The power of Hausman's argument is that, his endorsement of flourishing notwithstanding, it does not actually depend on this precise theory of well-being and can be accepted by at least some hedonists and subjectivists.

Premise 1 can be read as concerning instrumental goods, and no theorists of well-being deny that there are several such goods nor that these goods can be good indicators of well-being.

Premise 2 is also open to hedonists. Consider the hedonist who believes only in one single non-instrumental good—enjoyment, but rejects the possibility of objective measurement of enjoyment as envisaged by Jeremy Bentham in the 18th century and revived by some social scientists today. Such a measure is usually represented by a graph where *x*-axis represents time, *y*-axis represents intensity. The total enjoyment is the area

under the curve composed of ratings of intensity at an instant.⁵ Rejecting this picture, Roger Crisp, a modern hedonist, argues that the trade-offs between different enjoyable experiences implied by this classic approach are incompatible with Millian's high-quality pleasures which, in his view, can and should be respected by hedonism (Crisp, 2006). That an agent can judge enjoyability of an experience in a way that diverges from the product of intensity and duration is Crisp's way of accommodating heterogeneity of agents that Hausman is so keen on. So Crisp's is an example of hedonism about well-being that is not committed to existing measures. Similarly, there could well be hedonists (or partial hedonists) who believe that shape of life matters in a way that makes it impossible to evaluate enjoyment at a time without considering the rest of the person's life. If enjoyability of an experience is time-dependent in a way that is sensitive to individuals' identities, a hedonist can again endorse the first two premises.⁶

Similarly it is open to subjectivists to share Hausman's concerns about current measures expressed in Premise 2. Subjectivists would presumably require that measures of well-being gauge the extent to which agents' most important priorities formed under the right conditions are fulfilled. This is a tall order. Existing measures do not get at considered (or laundered) preferences. Merely asking people what they prefer and at what rate, or merely observing their actual choices in the marketplace as most economists do, likely fails to detect the sort of authoritative judgments about deep values that subjectivists are after. Recently, economists have started exploring creative ways to measure considered preferences through judgments and choices people make in carefully selected circumstances that plausibly reveal their genuine priorities—for example, when medical students weigh the pros and cons of different residency programs and give reasons for their choices (Benjamin *et al.*, 2014). I suspect these scholars would argue they *are* measuring what Hausman calls laundered preferences. Hausman does not discuss such attempts, but I would anticipate him pointing out that health poses special challenges to preference-based measures—on what grounds would an agent make a laundered preference about the relative value of broken ankle versus mild depression? Like Hausman, subjectivists too may not be in a hurry to endorse these new measures: it is one thing to get people to form thoughtful preferences about residency programs, but whether these preferences are sufficiently close to the fully informed and the fully rational preferences that idealized subjectivists favor remains an open question.

All this is to say that Hausman's argument against measurement is far-reaching even in its tempered version. It does not require an endorsement of flourishing and would appeal to anyone who believes that the existing (and possibly any conceivable) methods of measurement of well-being do a poor job at detecting well-being in a way that respects differences between individuals. It is thus to good, in reply to Hausman, to just defend hedonism, nor to criticize flourishing.

How then could we argue with Hausman?

Resisting Hausman

I will explore two replies. The first one is to Premise 1—must the impossibility of measuring well-being in its most demanding sense doom its measurement in all cases? The second response is aimed at Premise 2—it takes more than Hausman's appeals to implausibility of current efforts to undermine measures when these measures have gone through a validation procedure. I explain these replies in turn.

What Should We Mean By 'Well-being'?

Premise 1 embodies Hausman's conception of well-being. It is a demanding one—requiring a complete aggregation of all important goods in a way that respects the agent's history, character, talents, culture and values. Elsewhere, I have argued that, though this is one and perhaps the main sense of 'well-being', it is not the only one (Alexandrova, 2013). A long-term personal therapist, a close friend, or an obituary writer would typically focus on this sort of all-things-considered evaluation either of a life or of a period of life. This is also the sense on which philosophers have traditionally focused.

But sometimes 'well-being' connotes a less demanding evaluation. A doctor might be interested in well-being of her patients with a chronic disease, a teacher—in well-being of students from foster homes, a social worker—in well-being of single-parent families and a development economist—in well-being of people in a refugee camp. In each of these cases, well-being is predicated of a particular *kind* of people in a specific *type* of circumstances. This sort of evaluation is at once narrower than Hausman's—not all goods are taken into account, but only those shared by a particular group of people in particular situations. It is also broader in that it considers a kind of person rather than an individual.

This—let us call it *contextual*—sense of well-being is sometimes all that we mean by well-being. It is also the

sense most suitable to science, policy and social services in contemporary bureaucracies. A good social worker knows a lot about how to help families in a specific type of trouble, and a good child psychologist often knows exactly the needs of the kids she is looking after. This is ‘well-being’ as studied by researchers in social and medical sciences and about which implicit knowledge of teachers, therapists and social workers is accumulated. Whether measured formally by indicators or questionnaires, or eyeballed by an experienced specialist, this sense of ‘well-being’ depends on modest generalizations. Here are some examples: recently adopted children benefit from a period of intense bonding with no one other than their parents; caregivers of chronic patients are at risk for health and well-being even with social support; early learning difficulties impact later well-being more than other causes.

Of course, if we focus on individuals we might find exceptions: recently adopted toddlers who can go to nursery right away, caregivers who are just fine, and successful adults that get over early learning problems. My claim is not that focusing on specific populations avoids all the problems that Hausman finds with well-being measurement (not here anyway). Rather, I claim that if it makes sense to predicate well-being of kinds and not merely of individuals, then general claims about what is good for a given kind will be possible too. This is because kinds are identified by the generalizations they support—that is one common definition of kinds anyhow (Boyd, 1991)—there will thus be generalizations about how members of this kind function in such and such circumstances.

To the extent that such knowledge is possible and to the extent that this knowledge is about well-being *in a sense*, we have one reply to Hausman. He selected the most demanding and the least epistemically accessible notion of well-being and showed an impossibility of measurement for this notion based on intuitive impossibility of making generalizations about well-being. But this is too easy. Whether such generalizations are on the whole reliable is an empirical question that needs more attention than I can give. But given that teachers, social workers and medical professionals routinely make well-being judgments on the basis of such generalizations, Hausman owes us a case against these practices. Hausman could retort that contextual well-being is not true well-being. It is perhaps quality of life, or performance according to one indicator, but not proper well-being. But at this point the argument has shifted into an unhelpful territory about who is entitled to the term ‘well-being’. Erring on the side of liberality, it is at least permissible to maintain, *pace* Hausman, that there

is more to evaluation than judging individual lives, all things considered.

Still, even allowing for well-being in this contextual sense, what confidence should we have in the existing methods of its measurement? It is not enough to show that Hausman’s argument is premised on too demanding of a notion. That merely shows that there are other notions which apply to kinds and since kinds are based on generalizations, these contextual notions are better candidates for measurement. That secures *potential* measurability of contextual well-being. But to address Premise 2, we also need to show that measurement of contextual well-being is in actual fact realistic and defensible or more so than Hausman maintains.

Are Current Measures That Bad?

Evaluating the validity of existing measures of contextual well-being is a big job which I could not possibly complete in this essay. But I can offer reasons to withhold judgment on Hausman’s skeptical verdict. Against these measures, Hausman offers insufficient evidence—a mixture of appeals to intuition and unsystematic references to studies that expose problems in one or another questionnaire. But the field of social and medical measurement has elaborate procedures for validation, some of which I will describe below. These procedures despite being well established remain controversial, so my argument is not ‘had Hausman examined them, he would be more optimistic about measurement of well-being’. I myself am more optimistic but I cannot make a full case for such optimism here. Rather my claim is that evaluating current measures of well-being takes more than appeals to intuition.

If you asked social scientists why they use a given measure of well-being, their answer would probably invoke a process known as psychometric validation.⁷ The psychometric tradition in the social sciences has traditionally specialized in developing tests and questionnaires for detecting unobservable attributes (called ‘constructs’) such as intelligence, personality and lately well-being. Some of the measurement tools and their use in research on race, gender and class, especially in the early 20th century, attracted much controversy. The eugenic roots of this work are dutifully and solemnly acknowledged in the introductory courses to psychometrics.⁸ But for virtually all researchers who measure an attribute on the basis of people’s reports or performances in tests, psychometric validation remains the obligatory procedure.⁹ Large swaths of the science of well-being in particular have embraced questionnaires and with that psychometric validation.¹⁰ Since measures of

life satisfaction are one of Hausman's targets, I will describe their validation further on.

The famous Satisfaction With Life Scale (SWLS) is a five-item questionnaire which elicits judgments about life as a whole (for example, item three is 'I am satisfied with my life', 1 = strongly disagree, 7 = strongly agree). The process of its validation is described in the much-cited article by Ed Diener and his colleagues (Diener *et al.*, 1985). They started with 48 items all inviting reports of satisfaction with life, and positive and negative affect. Questions about affect were eliminated first because life satisfaction, researchers judged, is a judgment about one's life as a whole, not a report of emotions. This left them with 10 items, 5 of which were eliminated because of 'semantic similarity'. During the validation stage, 176 undergraduate psychologists in University of Illinois took the test twice with an interval of 2 months. This produced the data that allowed Diener and his colleagues to judge, among other things, the ability of each item to predict the overall score (a property called internal consistency). This exercise showed that questions about life satisfaction elicit reliable responses. But do these responses have anything to do with well-being? This is when *construct validation* enters the picture.

Construct validation purports to ensure that the measure of the construct in question behaves exactly as it should, given researchers' background knowledge. On the original proposal formulated in the classic paper in 1955 by Cronbach and Meehl, construct validation consists in testing the nomological network of hypotheses in the neighborhood of the construct (Cronbach and Meehl, 1955). To measure X, we need to know how X behaves in relation to other properties and processes that are systematically connected with it by law-like regularities. Among psychologists, the consensus seems to be that construct validity is the true validity, encompassing all other validities such as criterion, prediction, discriminant and content (Strauss and Smith, 2009).

SWLS, according to its authors, earned construct validity when Diener and his colleagues compared responses on the SWLS to responses on other existing measures of subjective well-being and related constructs such as affect, happiness and domain satisfaction. The findings confirmed their expectation that SWLS scores correlate sufficiently well with those measures that also elicit a judgment on subjective well-being and less so with measures that focus only on affect or self-esteem or other related but distinct notions. In the final stage of testing, the SWLS scores of 53 elderly people from the Urbana-Champaign area were compared to ratings this same population received in a structured in-depth

interview that focused on 'the extent to which they remained active and were oriented toward self-directed learning' (Diener *et al.*, 1985: 73). Once the interview results were converted into a single score, this score was found to have 0.43 correlation with the SWLS, an 'adequate' correlation by the standards of the discipline.

Since 1985, SWLS has been and continues to be scrutinized for its agreement with the growing data about subjective well-being. Individual judgments of life satisfaction have been checked against the reports of informants close to the subjects (Schneider and Schimmack, 2009). Proponents of SWLS argue that it exhibits a plausible relationship with money, relationships, suicide and satisfaction with various domains of life, such as work and living conditions.¹¹

Of course, scientists readily admit that validation is a continuous process, that it is never strictly speaking over and that measures need to be revalidated for each new environment in which they are used. Validation of the SWLS described above did not stop many skeptics raising questions about the relation between life satisfaction and actual subjective well-being. Critics were keen to show that life satisfaction judgments are ad hoc constructions that sway with arbitrary changes in the environment, such as finding a coin right before the test (Schwarz and Strack, 1999). These are the criticisms that Hausman invokes against life satisfaction (p. 110). He neglects to mention, however, the lengths to which psychologists have gone to check whether life satisfaction judgments are quite as fragile. It turns out that they are not, and today SWLS continues to be popular partly because these judgments are more robust than its critics alleged (Oishi *et al.*, 2003, Lucas, 2013).

The story of validation of SWLS is fairly typical. All the questionnaire-based measurement of health-related quality of life, flourishing and emotional state go through a similar process.¹² I recount this story in detail to emphasize that most measures of well-being are not selected purely on the basis of their intuitive appeal to the scholars who put them forward.¹³ They are validated according to a principle that forms the backbone of psychometric methodology: 'To determine whether a measure is useful, one must conduct empirical tests that examine whether the measure behaves as would be expected given the theory of the underlying construct' (Diener *et al.*, 2008: 67). This is a broadly coherentist philosophy according to which a good measure of a phenomenon is validated against other fallible measures of it or against fallible measures of related phenomena. In this sense, it is a familiar story for historians and philosophers of measurement.¹⁴

I suspect that Hausman would welcome attention to psychometric validation. But does this methodology offer a solution to his heterogeneity problem, which to remind, stems from the fact that different people place different values on the goods picked out in well-being questionnaires? Potentially yes. If a questionnaire really does agree with all of the relevant background knowledge as construct validation aspires to ensure, then the mere intuition that this questionnaire *could* go wrong in some individual case remains just that, an intuition. I imagine that believers in construct validation would reply to Hausman's worries in just this way—'we will only worry about heterogeneity if the data indicate that our questionnaires do not behave as they should'.

This is a fine response as far as it goes. But in actual fact, construct validation does not live up to its great ambition to check questionnaires against *all* the relevant knowledge. Often questionnaires are generated and validated in a mechanical way that ignores the meaning and normative roles of concepts such as happiness and well-being (Alexandrova and Haybron, forthcoming). For example, construct validation can fail to take into account the phenomenon of *response shift*.¹⁵ Response shift occurs when the meaning of subject's evaluation changes due to a change in their values or adjustment to new personal circumstances. This is a case of heterogeneity, albeit intra-personal rather than inter-personal. Medical researchers are concerned about response shift because it is documented in subjects whose health status changes, for example, with a sudden onset of disability (Schwartz and Spranger, 1999; Schwartz *et al.*, 2007). When patients reconceive their lives in response to new obstacles or opportunities, a well-being questionnaire validated by the standard methods will not necessarily pick out this change. McClimans and coauthors argue that this can be a failure of validity (McClimans *et al.*, 2013).

So Hausman's concern about heterogeneity might well reappear even when the focus is on contextual well-being and even when measures have passed the conventional tests of psychometric validation. Nevertheless, a sweeping skeptical verdict of the kind he reaches in this book is not warranted on the basis of reasons he provides.¹⁶ Heterogeneity is a fact of life, and a well-known one to the researchers, but whether it shuts the door on measurement is a far more complicated question than Hausman allows. My own bet is that there will be pockets of valid measurement provided careful enough application of existing methods.

Taking Stock

I have argued that well-being could be measurable if we focus on contextual rather than general well-being and if our measures behave in a way that coheres with all the available evidence. Does my case make a serious dent in Hausman's argument? Yes and no.

No, because I have not said anything to undermine Hausman's main contention that well-being in its most expansive sense is not measurable with the current (nor possibly any) tools. It is also plausible that this general sense of well-being is the most central and significant to human life. My invocation of a different sense of well-being—the contextual sense—may come across as lowering the bar in a way that makes the concept lose its unifying force in human life. If such a redefinition serves only the goal of making measurement possible, that seems like putting the scientific cart before the philosophical horse.

On the other hand, such a redefinition happens all the time and for a good reason. One of the lessons of recent work on history and philosophy of measurement is that the theory of the phenomenon and its measurement *co-evolve*. To quote Bas van Fraassen: 'The questions *What counts as a measurement of (physical quantity) X?* and *What is (that physical quantity) X?* cannot be answered independently of each other' (van Fraassen, 2008: 116, author's italics preserved). To apply this insight to our case, it is no good to decide ahead of time from a philosopher's pedestal what well-being is and then declare that no measure can do justice to this notion. The practicalities of measurement, the need for common reliable standards that enable comparisons and scientific communication, should all naturally inform the shape of concepts we posit. This mutual correction of scientific requirements and philosophical constraints (plus political and cultural ones) is the story of science. If well-being is to play a useful role in life of today's industrial bureaucracies which live by numbers, it may have to be *made* measurable even if it was not initially.

The big issue—and one on which I anticipate parting ways with Hausman—is how much flexibility well-being as a concept legitimately admits and how reliably construct validation can track differences among and within subjects. I hope it is flexible enough to apply to kinds of people rather than only to unique individuals, and flexible enough to be tractable, at least sometimes, by judiciously applied measures.

Notes

1. Recent discussions of the empirical study of well-being bypass Hausman's concerns. Angner (2013)

argues that data from self-reports of happiness are easily misinterpreted as data about well-being, without doubting that well-being itself is measurable. Haybron (2008) bets that happiness is a good enough proxy for well-being, so we should focus on improving measures of happiness. McClimans and Brown (2012) and Hunt (1997) attack quality-of-life measures in medicine, respectively, for treating well-being as an outcome rather than a process, and for not giving a clear definition of quality of life. These latter arguments perhaps echo Hausman's concerns.

2. England's Chief Medical Officer's 2014 report is the most recent example of the pushback Davies (2014). Older critiques are Hobart *et al.* (2007) and Hunt (1997).
3. For Hausman these constraints illustrate the contrast between well-being and health, but this is not critical for my purposes.
4. In Dale Dorsey's recent characterization, 'subjectivism about well-being holds that ϕ is intrinsically good for x if and only if, and to the extent that, ϕ is *valued*, under the proper conditions, by x ' (Dorsey, 2012, author's emphasis). Subjectivists then argue about the nature of these proper conditions and about what valuing should consist of.
5. An operationalization of this approach is described in Kahneman (1999) which he calls 'objective happiness' to distinguish it from happiness as judged by the individual retrospectively.
6. I thank Ben Bradley and Ben Bramble for clarifying this point for me.
7. The rest of this section uses some material from Alexandrova and Haybron (Forthcoming).
8. For a brief history of the first psychometrics laboratory, see <http://www.psychometrics.cam.ac.uk/about-us/our-history/first-psychometric-laboratory>.
9. DeVet *et al.* (2011) is a textbook for development and validation of measures in this tradition.
10. See Angner (2011) for the place of psychometric validation in the current science of well-being, especially vis-à-vis the welfare economics tradition.
11. See Diener *et al.* (2008: 74–93) for summary and references. Haybron (2008) presents a case against life satisfaction as a measure of well-being in chapter 5.
12. This is not, strictly speaking, true. There are many different approaches to validation (classical, item-response theory, Rasch, etc). But SWLS's story is largely typical for psychology of well-being. See Diener *et al.* (2010) for a recent example of validation of happiness and flourishing measures using roughly the same methods.
13. I say 'most' because the affect-based measures of happiness (see footnote 6) were not, to my knowledge, validated in this way. Rather, their justification is a mixture of critique of life satisfaction and derivation from axioms (see Kahneman (1999) for further references).
14. Chang (2004) describes the process of development of measures of temperature in coherentist terms; van Fraassen (2008) also emphasizes a co-evolution of measurement and theory.
15. I thank an anonymous referee for suggesting this example.
16. Michell (1999), another example of an overly sweeping verdict, argues that the very idea of quantitative scales in psychology is deeply flawed in part because the phenomena in question are qualities, not quantities. But most measurement scholars, even those who are critical of psychometric methodology, are less skeptical than Michell (Borsboom, 2005).

Acknowledgements

The author would like to thank two anonymous referees, James Wilson, Stephen John and the participants of the June 2015 Valuing Health Workshop at University College London.

References

- Angner, E. (2011). Current Trends in Welfare Measurement. In Davis, J. B. and Hands, D. W. (eds), *The Elgar Companion to Recent Economic Methodology*. Northampton, MA: Edward Elgar, pp. 121–154.
- Angner, E. (2013). Is it Possible to Measure Happiness? *European Journal for Philosophy of Science*, **3**, 221–240.
- Alexandrova, A. (2013). Doing Well in the Circumstances. *Journal of Moral Philosophy*, **10**, 307–328.
- Alexandrova A. and Haybron, D. (forthcoming). Is Construct Validity Valid? *Philosophy of Science*.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (2014). Can Marginal Rates of Substitution Be Inferred From Happiness Data? Evidence from Residency Choices. *The American Economic Review*, **104**, 3498–3528.

- Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.
- Boyd, R. (1991). Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies*, **61**, 127–148.
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
- Crisp, R. (2006). Hedonism Reconsidered. *Philosophy and Phenomenological Research*, **73**, 619–645.
- Cronbach, L. J. and Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, **52**, 281–302.
- Davies, S. C. (2014). *Annual Report of the Chief Medical Officer 2013: Public Mental Health Priorities—Investing in the Evidence*. London: Department of Health.
- Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, **49**, 71–75.
- Diener, E., Lucas, R., Schimmack, U., and Helliwell, J. (2008). *Well-being for Public Policy*. New York: Oxford University Press.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., and Biswas-Diener, R. (2010). New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research*, **97**, 143–156.
- De Vet, H. C., Terwee, C. B., Mokkink, L. B., and Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press.
- Dorsey, D. (2012). Subjectivism without Desire. *Philosophical Review*, **121**, 407–442.
- Haybron, D. M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. New York: Oxford University Press.
- Hobart, J. C., Cano, S. J., Zajicek, J. P., and Thompson, A. J. (2007). Rating Scales as Outcome Measures for Clinical Trials in Neurology: Problems, Solutions, and Recommendations. *The Lancet Neurology*, **6**, 1094–1105.
- Hunt, S. M. (1997). The Problem of Quality of Life. *Quality of Life Research*, **6**, 205–212.
- Kahneman, D. (1999). Objective Happiness. In Kahneman, D., Diener, E. and Schwarz, N. (eds), *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, pp. 3–25.
- Kraut, R. (2007). *What is Good and Why: The Ethics of Well-being*. Cambridge, MA: Harvard University Press.
- Lucas, R. E. (2013). Does Life Seem Better on a Sunny Day? Examining the Association between Daily Weather Conditions and Life Satisfaction Judgments. *Journal of Personality and Social Psychology*, **104**, 872–884.
- McClimans, L. and Browne, J. P. (2012). Quality of Life is a Process Not an Outcome. *Theoretical Medicine and Bioethics*, **33**, 279–292.
- McClimans, L., Bickenbach, J., Westerman, M., Carlson, L., Wasserman, D., and Schwartz, C. (2013). Philosophical Perspectives on Response Shift. *Quality of Life Research*, **22**, 1871–1878.
- Michell, J. (1999). *Measurement in Psychology*. Cambridge: Cambridge University Press.
- Oishi, S., Schimmack, U., and Colcombe, S. J. (2003). The Contextual and Systematic Nature of Life Satisfaction Judgments. *Journal of Experimental Social Psychology*, **39**, 232–247.
- Schneider, L. and Schimmack, U. (2009). Self-informant Agreement in Well-being Ratings: A Meta-analysis. *Social Indicators Research*, **94**, 363–376.
- Schwarz, N. and Strack, F. (1999). Reports of Subjective Well-being: Judgmental Processes and their Methodological Implications. In Kahneman, D., Diener, E. and Schwarz, N. (eds), *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, pp. 61–84.
- Schwartz, C. E. and Sprangers, M. A. G. (1999). Methodological Approaches for Assessing Response Shift in Longitudinal Health-related Quality-of-life Research. *Social Science and Medicine*, **48**, 1531–1548.
- Schwartz, C. E., Andresen, E. M., Nosek, M. A., Krahn, G. L., and RRTC Expert Panel on Health Status Measurement. (2007). Response Shift Theory: Important Implications for Measuring Quality of Life in People with Disability. *Archives of Physical Medicine and Rehabilitation*, **88**, 529–536.
- Strauss, M. E., and Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, **5**, 1–25.
- van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.