

University Press Scholarship Online

Oxford Scholarship Online



A Philosophy for the Science of Well-Being

Anna Alexandrova

Print publication date: 2017

Print ISBN-13: 9780199300518

Published to Oxford Scholarship Online: August 2017

DOI: 10.1093/oso/9780199300518.001.0001

Is Well-Being Measurable?

Anna Alexandrova

DOI:10.1093/oso/9780199300518.003.0005

Abstract and Keywords

Among the scientists and policymakers, measurability of well-being is now almost uncontroversial—only the individual measures are. Philosopher Daniel Hausman, however, argues that well-being in the proper sense is too person-relative and hence heterogeneous. It is not a quantity that can be measured on the population level. This chapter agrees that there is a sense of well-being—the all-things-considered individual well-being—on which it is likely not measurable but disagrees that therefore well-being is not measurable in any sense. Crucial to measurement is the existence of generalisations between core components of well-being and observable indicators. Such generalisations are available if well-being is predicated of kinds of people, rather than of individual lives. Assuming this focus on well-being of kinds, validity of most existing measures of well-being is secured by the process of construct validation, whose logic relies on a plausible ideal of balancing all evidence.

Keywords: Hausman, heterogeneity of well-being, measurement, construct validation, individual well-being, well-being of kinds, indicators

The last remaining element in my story is measurement. To be properly about well-being, it is not enough that the science adopts well-being constructs based on the right theories, nor is it enough that these

constructs pass the objectivity test discussed in the last chapter. To be an object of science today, well-being also needs to be measurable and measured. In the Introduction I claimed that measurement is now a central pillar of the normal science of well-being. Gus O'Donnell's 2011 call to arms 'If you treasure it, measure it' was directed at policymakers and civil servants, but it was made possible by the comfort with measurability of well-being in the scientific community. Some of this comfort is very old indeed. Economics boast perhaps the most established tradition of measuring *welfare* as a function of wealth via various economic indicators such as gross domestic product, total surplus, household income, and consumption. But this is not the method that O'Donnell is celebrating and upon which the science of well-being stakes its novelty. Economic tradition of welfare measurement builds on the view that well-being consists in satisfaction of individual's preferences expressed in this individual's choices. My definition of the science of well-being is broad enough to encompass this approach. But the public face of the science of well-being typically comes with criticism of the economic definition and its replacement, or at least supplementation, with subjective well-being. 'Beyond Money: Towards an Economy of Well-Being' was the title of Diener and Seligman's seminal 2004 article. This is why more significant than the acceptance of the economic measurement of welfare is the growing acceptance of measures of happiness, life satisfaction, flourishing, and the many constructs in (p.107) medicine that include the patient's perspective. They are now part of official statistics in many countries. As I mentioned in the Introduction, the fundamental disagreements about whether well-being can be measured have been replaced by specific disagreements about how and what aspect of it to measure.

But sceptics remain and should be heard. In this chapter I analyse what is to my mind the most compelling and explicit such challenge, both to the current measures of well-being, mainly the subjective measures that so excited O'Donnell, and indeed to the very idea that well-being is measurable. The challenge is roughly that well-being is too person-relative to measure reliably. I encounter this and related concerns often enough among the critics of the science of well-being.¹ But the formulation I attend to is due to Dan Hausman, and it appears in his recent book *Valuing Health* (Hausman, 2015). It is unique in that his focus is explicitly on measurability of well-being, rather than of happiness. As a psychological state, happiness's measurability is a matter of access to and comparability of mental states (one's own and that of others). So in debates on measurement of happiness philosophers focus on the ability of, say, questionnaires to gauge how people truly feel and to compare mental states across people.²

Hausman, on the other hand, is concerned with well-being, which, as consensus has it, is a broader evaluative category, whose measurability consequently raises further problems in addition to the problems with measurement of happiness. In particular he argues that the concept of well-being calls for aggregation of goods in a person's life in a way that is duly sensitive to who this person is. The existing measures—whether focused on feelings of happiness, or life satisfaction, or quality of life—do not aggregate all the relevant goods in a way that respects individuality. So even if happiness is measurable, to the extent that other goods matter for well-being (p.108) in addition to happiness, well-being will turn out to be unmeasurable. Comparing and ranking well-being states is possible albeit hard on an individual level, Hausman concedes, but becomes largely unrealistic on the population level required by science and policy. For him this means that well-being should not be the yardstick on which healthcare is evaluated and distributed and instead different categories, for example opportunities, are better suited for this task.

I start by reconstructing this argument and then argue that although it does not doom the project of measurement of well-being, it teaches a lesson for what sort of measurement should be expected and trusted. Given my variantist view—that well-being constructs can legitimately differ in substance and focus—one reply to Hausman is straightforward. Hausman selects the most demanding sense of well-being, one that calls for a comprehensive aggregation of all goods over the course of one individual's life. But this sense of well-being is not unique and may not be the right sense for science and policy. Showing that well-being is not measurable in one sense—I largely agree with Hausman on that—does not show that it is not measurable in all senses. Is well-being measurable in *any* sense relevant to science and policy? In my view Hausman's scepticism does not generalise and there is still hope for measurement of well-being in other senses. This hope depends, however, on abandoning the project of capturing the all-things-considered well-being of individuals and focusing instead on its commonly valued components or on well-being of kinds that share features and circumstances.

Second, I put pressure on a common critique of existing measures, which Hausman exemplifies. To undermine these measures critics tend to give 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 empirical knowledge. 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 and other critics. So in addition to containing Hausman's scepticism, this chapter also presents an (p.109) interpretation of construct validation. Construct validation, I argue, follows a coherentist spirit according to which measures are valid to the extent that they cohere with theoretical and empirical knowledge about the states being measured. In this sense construct validation is similar to validation of measures in other sciences, and to the extent that its theoretical ideal is realised in practice, measures of well-being can be valid.

So the overall message of this chapter is optimistic. Hausman's sceptical verdict is not warranted on the basis of reasons he gives and the science of well-being comes with plausible methods for checking that its measures are valid. Still my optimism depends on the science of well-being adopting realistic target concepts and implementing its ideal of measurement.

5.1. Hausman 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 to this ankle's owner. This is the view that Hausman rejects. 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 the book making this argument. 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.

In order 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 (p.110) enough. Hausman believes that any account of well-being should accommodate the following constraints:

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

2. 'What is good for me depends heavily on who I am' (Hausman, 2015, p. 121); that is, an agent's well-being depends on this 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 periods 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's 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 noninstrumental 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 subjectivism—neither bare feelings nor the fulfillment of desires or goals capture 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 developmentalism 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. (Hausman, 2015, 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 (p.111) and within persons and the distances between these states. 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, so comparisons, let alone on an interval scale, are hopeless.

Hausman rehearses this argument but then recoils from its extreme scepticism. 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. It is equally true that sometimes feelings and preferences are decent guides to well-being. Adherents of idealised

subjectivism, recall, 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. Laundered preferences are those held by individuals who are '(1) self-interested, (2) well-informed, (3) evaluatively competent, and (4) free of deliberative defects, and if (5) they have complete and transitive preferences among all alternatives' (Hausman, 2015, 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 sceptical 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 measures of subjective well-being. About life satisfaction Hausman (2015, p. 129) complains that reports summarising a great deal of information are unreliable and sensitive to irrelevant details. About measures of net affect, such as Kahneman's 'objective happiness', he notes that when the average affect at an instant is calculated all emotions get counted (p.112) equally. He asks who decides how net affect is determined—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', he insists (p. 129). 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 (on the revealed preference approach) or else from their responses to questionnaires about what state of affairs they would prefer and at what rate (on the stated preference approach). Neither approach, he claims, makes an effort to select only the authoritative ones among these choices and judgements.

Here, then, is Hausman's tempered sceptical 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. (2015, 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 in Madison Wisconsin than to have become a child soldier. But in general these rankings will not be possible.

The following is a formalisation of his argument:

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

(p.113)

Premise 2: The existing measures are at best fallible indicators of some well-being relevant goods, but they do not respect heterogeneity.

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 proponents of different theories and even by those sceptical about a single theory. In Chapter 2 we came across Scanlon's (1998) worry that though we may manage to agree on core components of well-being (experiential quality, quality of life, success), a recipe for how much weight to assign to each is unlikely. This is Hausman's point too, but his additional contribution is to argue that this blocks social measurement.

Even some hedonists and subjectivists will agree. Premise 1 can be read as concerning instrumental goods, and no theorists of well-being denies that there are several such goods, nor that these goods can be good indicators of well-being. Premise 2 is also open to hedonists, for example, Roger Crisp, whom we encountered in Chapter 2. Recall that

he takes well-being to consist only in enjoyment but rejects the possibility of objective measurement of enjoyment. If the x-axis represents time, the y-axis represents intensity, and a curve is formed from ratings of affect at an instant, then according to the hedonist tradition in psychology the total enjoyment is the area under the curve. Rejecting this picture, Crisp (2006) argues that an agent may well judge a given experience to be so high on enjoyment as to not be tradeable against longer experiences that are lower on enjoyment. This is a way of accommodating Millian high-quality pleasures and indeed Hausman's heterogeneity of agents. So this 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 (p.114) a way that's sensitive to individuals' identities, a hedonist can again endorse the first two premises.

Similarly it is open to subjectivists to share Hausman's concerns about current measures expressed in Premise 2. Subjectivists, as we have seen, call for measures of well-being to gauge the extent to which agents' most important priorities formed under the right conditions are fulfilled. This is a tall order. Existing measures are unlikely to tap into laundered preferences. Merely asking people what they prefer and at what rate, or merely observing their actual choices in the marketplace, is a far cry from detecting the sort of authoritative judgements about deep values that subjectivists are after. I already mentioned that there are efforts under way to measure considered preferences through judgements 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 he might point 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? Evaluations of health states differ strikingly depending on whether they are offered by those who have experienced a condition or those who merely imagine it.³ 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 idealised subjectivists favour 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 (p.115) differences between individuals. It is thus no good in reply to Hausman to just defend hedonism or subjectivism nor to attack flourishing. How then could we argue with Hausman?

5.2. Is Well-Being Heterogeneous?

I start by questioning Premise 1—is well-being so heterogeneous as to doom measurement? Heterogeneity for Hausman means that there is no stable contribution a good makes toward all agents' well-being. Rather this contribution depends on the agent's history, values, and so on. It is the combination of goods that makes for well-being, not the presence of any specific good at any specific level. This appeal to uniqueness of individuals—no person is likely to have the same recipe for their well-being stew as any other—can be challenged. Are there really no commonalities to human flourishing? Tolstoy after all taught that all happy families are alike.

Consider happiness. In his *Pursuit of Unhappiness*, Dan Haybron (2008) bets that so long as this psychological state is understood in a sufficiently rich way, it is a good proxy for well-being. Happiness for him is no mere experience of pleasure, nor mere approval of one's life. Neither of these are deep or central enough to our psyche. Rather happiness is an emotional state that disposes us to certain ways of reacting to the world and to ourselves. Happiness for Haybron is the opposite of depression—the negative emotional state underlying all affects—not the opposite of superficial and peripheral affects such as pain or sadness. Three dimensions make up the happy emotional state: to feel at home with oneself and one's world (which Haybron calls attunement), to be engaged, and only lastly to endorse one's life. Think of the peaceful and awesome Santiago in Hemingway's *Old Man and Sea*. Call this Haybron-happiness.

Haybron contends that though well-being without this kind of happiness is conceivable, it is vastly unlikely—Wittgenstein famously claimed he had a great life, but who is inclined to wish such misery on a newborn? We are sure all unique, as Hausman says, but it is hard to imagine that we do not all need Haybron-happiness, at least to a minimal degree. (p.116) Depression is not compatible with well-being

no matter what stew of goods uniquely suits each of us. Well-being must have something to do with ‘living in accordance with our emotional nature’ (Haybron, 2008, p. 193). Not all current measures of subjective well-being get at happiness so understood, but the emotional state in question is not inherently unmeasurable, and some measures come close enough (Haybron, 2016). Attunement, engagement, and endorsement are characteristics that, for the purposes of measurement, are similar to personality traits. If those are measurable—and it is hard to find more consensus in psychology than that traits of character are real and detectible—then so is Haybron-happiness. Should such a measure be available, it would plausibly tap into the most important and central aspect of well-being, so well-being would end up being measurable by a proxy.

Much heterogeneity in Hausman’s sense will remain though, no matter how good our measures of happiness are. It is hard to function without Haybron-happiness, but how much health, relationships, fulfilling work, and so on we all need and in what proportion is likely to differ from person to person. Suppose there existed a measurement tool that allows individuals to judge (perhaps even authoritatively) the combination of goods uniquely suitable to them. The Organisation for Economic Co-Operation and Development (2013) Better Life Index has precisely such ambition when it invites people to rate various areas of life (work, housing, health, education, etc.) for their importance to them before evaluating their satisfaction with each. Whether such a tool will be practical and informative, especially for largescale science and policy, is another matter. Science and policy as practiced today require generalisations about components of well-being, and even if such generalisations exist, it is an open question whether they are sufficiently robust, scientifically and politically. They may hold the interest of scientists and indeed they do—hence the research on well-being at work, in cities, and so on. But policy robustness will be tricky—how important should a well-being component be for policymakers to be justified to invest in this good at the expense of others and with a danger of disadvantaging those individuals for whom this good plays a minor role? The comparability of the value of different bundles of goods will remain a problem.

(p.117) For these reasons I am inclined to agree with Hausman that the judgement, let alone measurement, of the overall well-being of an individual is likely to be fiendishly complex—there are too many variables in the question of how a given unique person is faring in life, too many ways in which various goods may combine or fail to combine. This is why I do not pursue this reply much further and concentrate on

a different challenge to heterogeneity—must the impossibility of measuring well-being in its most demanding sense doom its measurement in all cases?

5.3. Contextual Well-Being as Measurable

The first thing to recognise is that while general well-being may be unmeasurable, some of its components are likely measurable. If well-being is a wholistic good as Hausman argues, it will be made up of some goods that are common to most or many individual packages. Haybron-happiness is one such component; others are health, positive relationships, security and so on. Multi-indicator measures of these surely provide valuable knowledge related to well-being, even for those who balk at calling them well-being'. But this is hardly controversial, and I believe more can be said.

Premise 1 embodies a demanding conception of well-being—requiring a complete aggregation of all important goods in a way that respects the agent's history, character, talents, culture, and values. In Chapter 1, I called this sense of well-being the all-things-considered evaluation and distinguished it from well-being in contextual sense. Hausman follows the trend in philosophy to call 'well-being' only the sort of evaluation that is relevant to a long-term personal therapist, a close friend, or an obituary writer. I have argued that sometimes 'well-being' connotes a less all-encompassing evaluation, such as when well-being is judged by family doctors, teachers, social workers, aid workers, and so on. In Chapter 2 I argued that the nature of this latter contextual well-being is the purview of mid-level theories, proposing one for children in Chapter 3. It is mid-level theories that underlie, though not as (p.118) explicitly as I wish they did, the constructs in the sciences and the definitions of welfare in contemporary bureaucracies.

In 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 this group of people in these situations. It is also broader in that it considers a kind of person rather than an individual. A good social worker knows how to help families in crisis; a good child psychologist knows what troubled children need. A reader willing to grant this knowledge should also grant that it depends on modest generalisations about well-being, whether measured formally by indicators or questionnaires or eyeballed by an experienced specialist. There are examples in addition to those about child development in Chapter 3: recently adopted children benefit from a period of intense

bonding with no one other than their parents; caregivers of chronically ill patients are at risk of ill-being even with social support.

Of course, if we focus on individuals we might find exceptions: recently adopted toddlers who can go to nursery school right away and caregivers who are just fine. But that is true about most generalisations, and the issue is ultimately empirical/pragmatic—how willing are we to attribute generalisable knowledge about well-being to specialists? It seems far more plausible to admit such knowledge about contextual well-being of kinds than about all-things-considered well-being of individuals. As I write this, thousands of unaccompanied refugee children from Syria, Eritrea, Afghanistan, and other troubled parts of the continent are roaming Europe—scared, vulnerable, abandoned. It is not a big mystery what is needed for *a* well-being of these children—security, stability, sustenance, and care. I say ‘a well-being’ because these goods are far more minimal than the child well-being I discussed in Chapter 3. Still such minimalism is defensible in an emergency, and in this sense the well-being of unaccompanied refugee children can be measured by their access to these basic goods. A table of indicators can represent the levels of each good, and, depending on how these indicators are aggregated, some comparisons between the levels of well-being of different groups of these children may be possible. Each of these children is an individual with a complex history, and it is possible to speak of the all-things-considered well-being of each. No doubt this is the (p.119) sense their families worry about. As individuals, different children within this group will need different levels of security, stability, sustenance, and care, but a potential benefactor on a rescue mission is justified in ignoring these at least sometimes. Such a benefactor will speak about the well-being of these children as a group, and, because this group’s well-being depends on a fairly obvious set of goods, measurement is conceivable.

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 generalisations they support—that is one common definition of kinds anyway (Boyd, 1991)—there will thus be generalisations 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 another 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 the intuitive impossibility of making generalisations about well-being. But this is too easy.

Uniqueness of earthquakes, avalanches and wars does not stop their scientific study, and contextual well-being is no different. 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 well-being proper. 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 as I did in Chapter 1, I maintain 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 that apply to kinds, and, since kinds are based on generalisations, 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, at least more so than Hausman maintains.

(p.120) 5.4. Construct Validation

So far I have said very little about what measurement and validity are. To make further progress on Hausman's challenge, we need to attend to why scientists take the existing measures to be valid. I argue in this section that 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. Evaluating current measures of well-being takes evaluating these procedures, not appeals to intuition.

If we asked scientists why they use a given measure of well-being, their answer would invoke *psychometric validation*. The psychometric tradition in the social sciences has traditionally specialised in developing tests and questionnaires for detecting 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 twentieth century, have an unsavoury history. The eugenic roots of this work are dutifully and solemnly acknowledged in the introductory courses to psychometrics.⁴ 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. Large swaths of the science of well-being in particular have embraced questionnaires and with that psychometric validation.

5.4.1. Textbook Procedure

Validation follows a typical pattern described in measurement textbooks and articles.⁵ First, researchers define the construct to be measured by elaborating its scope and limits. This is the conceptual stage in which meaning of terms is discussed, invoking anything from classical sources ('Aristotle said ...') to untutored intuitions, to dictionary definitions. For (p.121) example, the scope of happiness is often deemed to be the overall positive affect, while the scope of satisfaction with life is a cognitive judgement about one's conditions and goals. In the second stage, researchers choose a measurement method (a questionnaire, a test, or a task), select the items (what questions? what tasks?) and settle on the scoring method.

The Satisfaction with Life Scale (SWLS), which we have already encountered and which Hausman specifically mentions, is a popular five-item Likert scale for measuring the cognitive aspect of subjective well-being, that is, the extent to which subjects judge their life to be satisfactory. The process of its validation is described in the much-cited article by Ed Diener and his colleagues (Diener et al., 1985). The first two stages in this case consisted in analysing 48 items all in the conceptual neighbourhood of subjective well-being. The team eliminated questions about affect because they viewed life satisfaction as a cognitive judgement about one's life as a whole. This left them with 10 questions, a further five of which were eliminated because of 'semantic similarity'.

In the third and final stage, the instrument is tested for its construct validity, that is, its ability to capture the intended attribute.⁶ In the case of well-being measures, this stage frequently involves *factor analysis*: when hundreds of subjects fill out the same questionnaire, it is possible to observe the correlations between responses to different items. Unless the questionnaire in question is single-item, a rare occurrence in this field, much can be learned from these correlations. In particular they are used to show that there are one or more clusters of items called 'factors' that account for the total information. Scientists speak of factor analysis as extracting 'a manageable number of latent dimensions that explain the covariation among a larger set of manifest variables' (Simms, 2008, p. 421).⁷ Here 'explanation' (p.122) is evidently used in an entirely phenomenological sense as saving the phenomena (the phenomena being the total data generated by administering the questionnaire in question), rather than stating the causes or presenting a theory of the phenomena. For the SWLS, factor analysis identified all five items to be measuring the same latent

variable because a single factor accounted for 66% of the variance in the data (Diener et al., 1985). Other scales may turn out to gauge more than one dimension.

The next step of the testing stage is to check that the behaviour of these factors accords with other things scientists know about the object in question. In case of subjective well-being, this knowledge includes how people evaluate their lives and surroundings, what behaviour results from these evaluation, and what other people who know the subjects say about them. For example, the 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 intensity, happiness, and domain satisfaction. The findings confirmed their expectation that SWLS scores correlate highly with those measures that also elicit a judgement on subjective well-being and less so with measures that focus only on affect or self-esteem or other related but distinct notions. One piece of evidence in favour of the SWLS was that the scores of 53 elderly people from Illinois correlated well to the ratings this same population received in an extended interview about 'the extent to which they remained active and were oriented toward self-directed learning' (Diener et al., 1985, p. 73). The correlation, $r = 0.43$, was judged adequate by the standards of the discipline. Since 1985 the SWLS has continued to be scrutinised for its agreement with the growing data about subjective well-being. Individual judgements of life satisfaction have been checked against the reports of informants close to the subjects (Schneider & Schimmack, 2009). Proponents of the 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.⁸

(p.123) 5.4.2. Construct Validation Is Good, in Theory

The first step to evaluating these practices is to capture their rationale. Neither philosophers nor scientists themselves have done that, so I propose to start with the following schema that summarises all the grounds on which a measure can be declared valid in this particular tradition:

Implicit Logic: A measure M of a construct C is validated to the extent that M has been shown to behave in a way that respects three sources of evidence:

1. M is inspired by a plausible theory of C.

2. Subjects reveal M to track C through their questionnaire answering behaviour.
3. Other knowledge about C is consistent with variations in values of M across contexts.

The first condition captures the role of philosophising about the nature of C in the first stage of measure development. The second condition specifies the assumption behind factor analysis. It helps to reveal the structure of the construct as respondents see it (more on this in Chapter 6). The third acknowledges that scientists go beyond the merely internal analysis of the scale: a valid measure correlates with indicators that their background knowledge says it should and does not correlate with indicators that it should not.

There is much to be said on the foundations of this tradition. Philosophers of measurement roughly agree that goodness of a measure lies in its ability to mirror the behaviour of the target system. But how can such mirroring be established? Here there is no single story. Erik Angner (2009, 2011a) argues that there are at least two traditions—axiomatic and psychometric—corresponding respectively to economic and psychological approaches to the measurement of well-being.⁹ In economics the key to the measurement of well-being is a representation relation between preferences and behaviour contained in axioms of the fundamental utility theory, while in psychology and the clinical (p.124) sciences the key is a valid questionnaire. The leading account about how mirroring is accomplished is known as the representational theory of measurement. According to it a measure is validated if there is a demonstrated homomorphism between an empirical relational structure (say an ordered series of rods for measurement of length) and a numerical relational structure (relations between real numbers). Various systems of axioms establish the conditions for this homomorphism.¹⁰ However, the availability of a representational story for psychometric validation is far from obvious. Most scholars agree that the psychometric approach does not have the axiomatic basis characteristic of the representational approaches.¹¹ For example, there is no proof that an ordinal scale such as the SWLS can be treated as an interval scale, which nevertheless typically happens when it is used to rank, say, countries. In foundations of economics there are axiomatic structures to take us from preferences to utility functions and to choice behaviour but typically not so in psychology.¹²

My bet, though here I just state rather than defend it, is that the coherentism of Implicit Logic makes up for the lack of representational theorems. If a measure really behaves in accordance with background knowledge, then this alone is enough to secure its validity. Construct

validation as described earlier conceives of measurement as part of theory development and validation as part of theory testing. On the original proposal formulated in the classic 1955 paper by Lee Cronbach and Paul Meehl, construct validation consists in testing the nomological network of hypotheses in the neighbourhood of the construct in question. To measure x , we need to know how x behaves in relation to other properties and processes that are systematically connected with x by law-like regularities (Cronbach & Meehl, 1955). Something like this view is still the consensus among scientists of well-being: 'To determine whether a measure is useful, one must conduct empirical tests that examine whether the measure behaves as (p.125) would be expected given the theory of the underlying construct' (Diener et al., 2008, p. 67).

This vision of measure validation is *prima facie* defensible. Its spirit is remarkably similar to the coherentist vision that characterises recent work on measurement of physical quantities.¹³ The historians and philosophers of measurement emphasise that the outlines of the concept in question, be it temperature or time, and the procedure for detecting it, are settled not separately but iteratively, checking and correcting one against another. Similarly in our case, the initial philosophical judgement about the nature of happiness or quality of life is coordinated with other constraints such as the statistical features of the questionnaires and the background knowledge about behaviour, related indicators, and ratings of informants. The resulting measurement tools can be deemed valid to the extent that they accommodate all evidence.

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 and population. Validation of the SWLS described earlier did not stop many sceptics from raising questions about the relation between life satisfaction and actual subjective well-being. As we have already seen, critics accused life satisfaction judgements to be ad hoc constructions that sway with arbitrary changes in the environment. These are the criticisms that Hausman (2015, p. 110) invokes against life satisfaction. He neglects to mention, however, the lengths to which psychologists have gone to check whether life satisfaction judgements are quite as fragile. It turns out that they are not, and today the SWLS continues to be popular partly because these judgements are more robust than its critics alleged (Lucas, 2013; Oishi et al., 2003).

The story of validation of the SWLS is typical. All the questionnaire-based measures of health-related quality of life, flourishing, and emotional state go through a similar process.¹⁴ I recount this story in detail (p.126) to emphasise that measures of well-being are not selected haphazardly, not normally anyway, and it is thus not fair to criticise them by appeal to any particular apparent problem. Instead we need to criticise the whole package, and this is another part of my reply.

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 goods, even when faithfully detected by well-being questionnaires, have different values for different people? 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, and it is enough to establish that the sweeping sceptical verdict Hausman reaches is not warranted on the basis of reasons he provides. In actual fact construct validation does not live up to its great ambition to check questionnaires against *all* the relevant knowledge. This is my argument in the final chapter. For now I take stock.

5.5. Measurable After All

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 said little to undermine Hausman's main contention that well-being in its most expansive sense is not measurable with the current (or possibly any) tools. Scholars of happiness may well conclude that I have yielded too much ground. Positive psychologists, whose whole enterprise is premised on marshalling scientific method to improve overall well-being for individuals and organisations, will (p.127) be particularly unimpressed by my concession. I intend this consequence. On my story the science of well-being need not strive to be the science of all-things-considered well-being to justify its existence. Instead it is on firmest ground when it concentrates on well-defined kinds and contexts. If I wanted to know how to do better in life,

well-being science would not be my first destination. It may be informative about specific components of well-being, such as Haybron-happiness, healthy relationships, mental health, and so on. I may also learn useful facts about major components of my well-being to the extent that I am a member of a kind whose well-being is well understood by science. But science is unlikely to speak more directly about my overall well-being as a unique person that I am. For that I would instead go to someone who truly knows me, who can judge my own personal well-being stew in Hausman's sense—an old and thoughtful friend, a wise therapist, a mentor, a trustworthy religious leader if I were a believer. None of this justifies abandoning the pursuit of this science; it is just the realistic interpretation of the available knowledge.

Also disappointed will be those for whom this general sense of individual well-being is the most central and significant to human life. To them my invocation of a different, contextual sense of well-being will 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. Again this is intentional. Redefinitions happen. 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 (2008, p. 116): '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' (author's italics). 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 (p.128) 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 bureaucracies, which live by numbers—and this train appears to be unstoppable now—well-being may have to be *made* measurable even if it was not initially.

Notes:

(1.) 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. Though distinct, these arguments perhaps echo Hausman's concerns.

(2.) Angner (2013) defends measurability of happiness as a psychological state. He grants that data from self-reports of happiness are easily misinterpreted as data about well-being but does not discuss whether well-being itself is measurable.

(3.) See Dolan (2000) on the relevant science and Carel (2016) on its significance.

(4.) For a brief history of the first psychometrics laboratory see <http://www.psychometrics.cam.ac.uk/about-us/our-history/first-psychometric-laboratory>.

(5.) For example, DeVet et al. (2011), Simms (2008).

(6.) I concentrate on construct validity at the expense of other validities because among measurement theorists the consensus seems to be that construct validity encompasses all other types of validity, such as criterion, predictive, discriminant, and content validity (Strauss & Smith, 2009). See Chapter 6 for a discussion of content validity.

(7.) There is a difference between *exploratory* and *confirmatory* factor analysis (see De Vet et al., 2011, pp. 169–172, among others). The former is used to reduce the number of items in a questionnaire by identifying the one(s) that best predict the overall ratings. The latter, on the other hand, tests that the factors that best summarise the data also conform with a theory of the underlying phenomenon if there is one. This distinction is not important for the present argument.

(8.) See Diener et al. (2008, pp. 74–93) for summary and references.

(9.) Angner (2015) argued more recently that the social indicators approach, motivated by an objective list theory of well-being, has yet a third measurement story to offer.

(10.) On general theory of measurement see Suppes (1998), Tal (2016), Cartwright and Bradburn (2011). The representational account is developed in Krantz et al. (1971).

(11.) I have in mind Wilson (2013), Angner (2009), Borsboom (2005, p. 86), and Michell (1999).

(12.) Though see Kahneman et al. (1997) for an example of a representational approach to happiness.

(13.) Chang (2004), van Fraassen (2008), Tal (2013).

(14.) See Lyubomirsky and Lepper (1999)'s Subjective Happiness Scale, Diener, Wirtz, et al. (2010)'s SPANE and Flourishing Scale. All are examples of validation of well-being relevant measures using roughly the same methods.



Access brought to you by: McGill University