

MEASUREMENT SCEPTICISM, CONSTRUCT VALIDATION, AND METHODOLOGY OF WELL-BEING THEORISING*

VICTOR LANGE

*Section of Philosophy, University of Copenhagen & CoInAct Research Group,
University of Copenhagen*

THOR GRÜNBAUM

*Section of Philosophy, University of Copenhagen & Department of Psychology,
University of Copenhagen & CoInAct Research Group, University of Copenhagen*

Precise measurements of well-being would be of profound societal importance. Yet, the sceptical worry that we cannot use social science instruments and tests to measure well-being is widely discussed by philosophers and scientists. A recent and interesting philosophical argument has pointed to the psychometric procedures of construct validation to address this sceptical worry. The argument has proposed that these procedures could warrant confidence in our ability to measure well-being. The present paper evaluates whether this type of argument succeeds. The answer is that it depends on which methodological background assumptions are motivating the sceptical worry to begin with. We show this by doing two things. First, we clarify (a) the different types of well-being theories involved in the science of well-being, and (b) the general methodological dimensions of well-being theorising. Second, we apply these distinctions and argue that construct validation is an unsuccessful response to measurement scepticism if this scepticism is motivated by a form of methodological non-naturalism. In the light of this, the overall point of the paper is that philosophers and scientists, when discussing measurement of well-being, should explicate their deeper methodological commitments. We further suggest that making such explicit commitments might present philosophers with a dilemma.

* Shared first authorship

Contact: Victor Lange <victorlange.sohrab@gmail.com>
Thor Grünbaum <tgr@hum.ku.dk>

1. Introduction

In the last 40 years, we have witnessed a notable increase in scientific measurements of well-being and the philosophical discussion of these measurements (e.g., Kahneman, Diener, & Schwartz 1999; Sumner 1996; Bok 2010). Such measurements inform various decision processes. Policy-making may centre on estimations of the involved well-being of different choice outcomes (Frey & Stutzer 2002; Diener, Lucas, Schimmack, & Helliwell 2009; Alkire 2016). Likewise, levels of patient well-being are important when considering clinical procedures (Saunders & Burgoyne 2002; Kaasa, Mastekaasa, & Naess 1988). The evaluation of economic policies, social interventions, and clinical procedures often assumes that well-being is a graded phenomenon. A person can have more or less of it. Ideally, it would be preferable if such policies, interventions, and procedures could be evaluated in terms of whether they give us more or less well-being. Only few philosophers would be prepared to argue against these basic assumptions. The planning and the evaluation of policies and social interventions in terms of well-being make the measurement of well-being a central task. Without some objective form of estimation of the actual well-being of the members of some population, evaluating social interventions in terms of changes to their well-being becomes a problematic enterprise.

Here, however, fundamental controversies appear. Many philosophers are sceptical of the idea that scales and tests used in the social sciences can measure well-being (we call this type of sceptical stance *well-being measurement scepticism*). Well-being measurement scepticism is intuitive and has been a prominent theme in the philosophical literature. It often builds on the basic intuition that well-being is very different from other psychological attributes and traits (like personality traits and intelligence). This paper aims to make two contributions in relation to such scepticism.

First, the paper introduces two distinctions: one concerning different types of well-being theories involved in the science of well-being and another concerning the general methodological positions on well-being theorising. Second, utilising these distinctions, we discuss a recent type of argument against well-being measurement scepticism presented by Alexandrova (2017: 106). According to Alexandrova (for related arguments, see Bishop 2015 and Hersch 2022), the psychometric procedures of construct validation should warrant confidence in the ability of well-being scales to measure well-being. We argue that Alexandrova's argument assumes that well-being theories must be open to empirically driven theory revision. Consequently, the argument from construct validation, as well as other related forms of well-being coherentism, cannot be used to answer measurement scepticism, if such scepticism is motivated by a type of methodological non-naturalism common in moral philosophy and related theories of well-being.

We draw a twofold lesson from these arguments. First, the success of a response to measurement scepticism like Alexandrova's depends on which methodological background assumptions motivated the sceptical stance in the first place. Second, more generally, when answering measurement scepticism, one should explicate one's methodological commitments. If we neglect making explicit the methodological assumptions central to a theory of well-being, we risk hindering progress in our understanding of how well-being as a graded phenomenon should be measured and fed into policy-making, health planning, and various forms of social interventions.

We proceed as follows. Section 2 presents some clarificatory distinctions concerning measurements of well-being and connected sceptical worries—moreover, the section introduces construct validation as a promising response to such worries. In Section 3, we present the central argument of the paper according to which construct validation cannot be used to respond to measurement scepticism if this scepticism is motivated on methodological non-naturalist grounds. Finally, in Section 4, we discuss some further implications of accepting methodological non-naturalism for how to solve the present issues.

2. Well-Being Measurement Scepticism and Construct Validation

2.1. Types of Well-Being Theories

In order to evaluate how best to respond to measurement scepticism, we need to be clear about the theoretical context. It is common in the philosophical literature to stress the difference between philosophical theories of well-being (for instance, Tiberius 2018 and Hurka 1993) and the theories of well-being used by psychologists and economists (Alkire 2016; Clark 2016). In line with this, we draw the following distinction between two types of well-being theories.

First, an important kind of theory is the *characterising* well-being theory. Such theories characterise what well-being consists of, either for human beings in general or for a specific group of human beings. The term 'well-being' is often thought to denote what is non-instrumentally good for a person's life or what makes a person's life go best (Crisp 2017; Parfit 1984: app. I). Characterising well-being theories would be theories about these matters. An example of a characterising well-being theory for human beings in general could be Kraut's (2007) theory of flourishing. An example of characterising well-being theory for a specific group of individuals could be Alexandrova's theory of child well-being (Alexandrova 2017: 54). Note that if a characterising well-being theory is a theory about the nature of well-being, it does not follow that characterising well-being theory is a

value theory about well-being as a strictly normative phenomenon. That would just be one type of characterising theory. Other philosophers might characterise well-being more descriptively as a causal phenomenon (Bishop 2015).

Second, a separate kind of theory is the *associative* well-being theory. Associative theories do not claim to characterise what well-being is, neither for human beings in general nor for some specific group. Instead, they describe the relations between socio-economic and psychological factors that are associated with the well-being of human beings in general or specific groups of human beings. Taylor's notion of 'markers of well-being' is useful here (Taylor 2015: 75). A marker of well-being is understood as any phenomenon that is either a constituent, a producer, or a reliable indicator of well-being as conceived by mainstream well-being theories (Taylor 2015: 77). When measuring a marker of well-being (e.g., measuring positive affect in some individuals), it is underspecified what relationship to well-being one takes this marker to have (for instance, is positive affect a constituent, a producer, or an indicator of well-being?). Hence, we might say that associative well-being theories are theories of how markers of well-being relate to each other (e.g., how positive affect relates to health) and to other phenomena that are not markers of well-being (e.g., how positive affect relates to certain demographic facts). Associative theories are hereby not theories of well-being strictly speaking. To link the content of associative well-being theories directly to claims about well-being, one must draw on characterising well-being theories. This means that associative theories specify networks for markers of well-being without taking any stance on the nature of these markers.¹

The distinction between characterising and associative well-being theories can help us understand the type of well-being theories involved in empirical research. Consider the well-known tripartite theory, developed by Diener, Oishi, and Lucas (1997). This theory claims that subjective well-being consists of three elements: (i) positive affect, (ii) relative lack of negative affect, and (iii) cognitive evaluation of one's own life. Does the tripartite theory characterise what well-being is? Or does it only describe the psychological associations involved in the scores of people's self-report of their lives, while remaining underspecified about exactly how to relate these scores to well-being in a characterising way?

We remain agnostic with respect to how to best understand any particular psychological theory. We stress only that it is important to clarify whether one understands a theory as a characterising or an associative theory, if one is to evaluate its success. If asked directly, many scientists would perhaps understand their well-being theories as associative theories about the markers of well-being, whereas most philosophers are interested in well-being theories as characteris-

1. This separates associative well-being theories from the theorising done by Bishop (2015). Bishop (2015) presents a causal network, which is proposed to be a theory of the nature of well-being and its causal dimensions. This is a characterising theory.

ing theories. As the following sections argue, philosophers disagree on whether science can determine or substantially inform us on the question of which characterising well-being theory is the correct one.

2.2. Well-Being Measurement Scepticism

Several scales and instruments have been developed to measure well-being. Examples include the subjective measurement procedures of Satisfaction With Life Scale (SWLS, Diener, Emmons, Larsen, & Griffin 1985), the 10-item mood scale of the Positive and Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen 1988), the Subjective Happiness Scale (Lyubomirsky & Lepper 1999), and the objective measurement procedures such as the Human Development Index (Anand & Sen 1994). Despite this broad supply of instruments, one might be sceptical about our ability to measure well-being. Such a sceptical attitude about the measurability of well-being has some intuitive appeal and is frequently encountered in public debates.² More importantly, in the philosophical literature on well-being, problems connected to measuring well-being are a prominent theme (Angner 2010; 2013b; Fumagalli 2022; Hausman 2011: 7; Haybron 2020; Ingelström & van der Deijl 2021; Taylor 2015; van der Deijl 2017a).

Well-being measurement sceptics usually point to fundamental problems with the measurement scales or instruments. Here is a list of some of the problems (the list is not supposed to be exhaustive and the items might be partly overlapping). First, a number of researchers have pointed out that the scores of measurement instruments often diverge (Fumagalli 2022). That is, measures of well-being can vary significantly depending on the type of measurement instrument or scale. For instance, there are divergencies between psychological and economic measures (Benjamin, Cooper, Heffetz, & Kimball 2020), physiological measures and hedonic measures (Stone & Mackie 2013), and momentary and retrospective measures of subjective well-being (Alexandrova 2005). Some of these divergencies might be explained by the fact that the instruments target different phenomena (Margolis et al. 2021). But we also find measurement divergencies with respect to different measures of the same phenomenon (Fumagalli 2019; Haybron 2008: ch. 5; Martela & Sheldon 2019). Second, measurement scales are often not interval scales. Within individuals and across individuals, the difference on a scale between 2 and 3 and between 6 and 7 might not have the same value. One important reason for this is that scores on subjective self-report scales are a combination of two unknowns, a subjec-

2. See for example the following:

<https://www.prospectmagazine.co.uk/magazine/why-its-impossible-to-measure-happiness>

<https://www.bbc.com/news/magazine-11765401>

tive feeling and a cognitive evaluation or reporting function (Ingelström & van der Deijl 2021; Bond & Lang 2019; for a more positive evaluation, Plant 2020). Economists in particular are often sceptical about the use of ordinal data from happiness questionnaires to aggregate and compare group averages (Bond & Lang 2019; Schröder & Yitzhaki 2017). Third, points on a given measurement scale might not be consistent over time (Fabian 2022; Kaiser 2022). Cases of hedonic adaptation (for discussion, see van der Deijl 2017b) and the dependence of reports on evaluations (for discussion, see Haybron 2007) have been prominent in the discussion of this issue. Collectively, these measurement problems ground a sceptical attitude about the claim that we are able to measure well-being.

Instances of well-being measurement scepticism might occur on a spectrum ranging from a local scope (in which the scepticism is about one or more specific scales) to a more global scope (expanding the scepticism to all current scales and potentially future ones). Hausman (2015) might be an example of a local well-being measurement scepticism. Hausman (2015: 129) points to problem with the widely used SWLS. When using SWLS, subjects indicate whether they agree with statements such as ‘I am satisfied with my life’, on a seven-point Likert scale (Diener et al. 1985). Some studies have shown that in general participants evaluate their lives as better if a disabled person is present in the room, when compared to a control condition (Strack, Schwarz, Chassein, Kern, & Wagner 1990). Further, participants rate their satisfaction with life to be greater if they find a dime before answering the survey questions (Schwarz & Strack 1991). In the end, this makes the life satisfaction reports unstable and significantly context-dependent.³ In the more global end of the spectrum, White (2013) extends many of the sceptical worries of Hausman (2010; 2015) to cover all attempts to measure well-being. Note that the argument of this paper concerns both local and global forms of well-being measurement scepticism.

Well-being measurement scepticism could target two different psychometric aspects of well-being scales. First, one might take the worries of measurement scepticism to be about the fact that the scores of well-being scales are unreliable. This scepticism could be local (concerning the reliability of a limited set of scales) or more global (concerning most or all well-being scales). A score of a particular well-being scale would be reliable to the degree that the scale consistently produced the same score when repeated under similar conditions (Markus & Borsboom 2013: 58–59; McDonald 2013). If the sceptical worry concerns the

3. Importantly, other places, Hausman (2012) is moderately optimistic about measuring well-being by preferences. Note that some replication studies indicate that most context-effects are small to non-existent (Lucas, Oishi, & Diener 2016).

reliability of the scores of well-being scales, the empirical well-being literature contains many relevant considerations, since researchers already investigate the reliability of different well-being scales (e.g., Diener et al. 1985; Diener & Tay 2014). If we take measurement scepticism to be only a worry about reliability of scores, then these considerations would provide a plausible reply. This paper does not concern the issues of reliability.

Second, measurement scepticism could be interpreted as a worry about whether the measurements are best explained by the target phenomenon (i.e., well-being). While the worry about reliability concerns the consistency of results produced by a scale, the worry about the explanation of the results concerns our confidence in whether the results of a scale are produced by actually measuring the intended attribute or phenomenon (in our case, well-being) (Shadish, Cook, & Campbell 2002). Whereas the first kind of worry concerns the degree of measurement precision associated with a test score, this second kind of worry concerns its so-called validity. Even if a scale reliably produces the same results, it could still be an open question whether the results are best explained as measures of one common attribute or as systematic measurement artefacts. Again, scepticism about validity could be both more local (concerning the ability of a limited set of scales to actually measure well-being) and more global (concerning the ability of most or all well-being scales to actually measure well-being).

For the rest of the paper, when we talk about measurement scepticism, we mean the kind of scepticism that targets the validity of well-being scales. Accordingly, the measurement problems listed above are taken to justify a sceptical attitude to the claim that common measures of well-being are able to measure well-being (rather than some other phenomenon or nothing at all).⁴ This still leaves it open how one's fundamental background assumptions could motivate one's well-being measurement scepticism about measurement validity. One could imagine a measurement sceptic saying that results of well-being scales could not be measurements of well-being since well-being is not the kind of thing that can be measured—for instance, it is not a uniform phenomenon or proper scientific object (see Bechtoldt 1959 and Brodbeck 1957 for this type of view). Or, one could imagine a measurement sceptic saying that we could not be sure that results of well-being scales were in fact measurements of well-being, since there is currently no consensus about which well-being theory is correct, and that therefore we are 'in the dark' when interpreting and measuring well-

4. Alexandrova (2017: 106) seems to understand Hausman (2015: 129) to express a worry about measurement validity. Discussions of measurement scepticism concerning the validity of well-being scales is a prominent topic in the philosophical literature (see, for example, van Der Deijl 2017a; Taylor 2015; Alexandrova 2017; Hausman 2011: 7).

being (see Rodogno 2015 and Wren-Lewis 2014 for relevant philosophical discussions of such motivations).⁵

2.3. Construct Validation as a Response

Social scientists of well-being often validate their well-being scales (Lui & Fernando 2018; Koushede et al. 2019; Joshanloo, Capone, Petrillo, & Caso 2017; van Dierendonck 2005; Lundgren-Nilson, Jonsdottir, Ahlborg, & Tennant 2013; Pavot & Diener 2009). These validation studies typically examine correlational patterns in large sets of survey data to see what hypothesis about well-being has the best fit to the data. One example would be Arthaud-Day, Rode, Mooney, and Near (2005) who examined the tripartite theory (Diener et al. 1997). By observing both discriminant and convergent correlational patterns between results of different measurement scales assumed to measure subjective well-being, Arthaud-Day et al. (2005: 470) concluded that the correlational patterns supported the tripartite theory. According to Alexandrova (2017: 120), such validation studies have consistently shown that well-being scales like SWLS exhibit construct validity, meaning that we have reasons to be confident in the scale's ability to measure (in this case subjective) well-being, or at least very central aspects of it (Diener et al. 1985; Schneider & Schimmack 2009).⁶

When the scores of independent measurement instruments are converging, we have *prima facie* reason to increase our confidence in both the truth of the results and the ability of the instruments to measure what they are intended to measure (Cartwright 1991; Wimsatt 1981; Woodward 2006). Think of eyewitnesses: if they independently converge in the events they report, this seems to give us some reason for increasing our confidence. Given this basic logic, the psychometric procedures of construct validation *prima facie* appear to provide an answer to the second type of sceptical worry concerning the validity of scales.

5. One could also imagine a measurement sceptic that argues that well-being is a purely qualitative phenomenon. Since measurement presupposes that the measured phenomenon is quantitative in nature, well-being cannot be measured (for this type of argument, see Michell 2003; 2008). Economists sometimes express a related scepticism about the way in which psychologists turn ordinal scale data into quantitative and statistical information about group averages (Bond & Lang 2019). In this paper, we set aside this type of scepticism. For the sake of the argument, we assume a broad notion of measurement according to which measurement "is defined as the assignment of numerals to objects or events according to rule" (Stevens 1946: 677). Our list of possible motivations for measurement scepticism is not exhaustive.

6. We are not suggesting that measures of *subjective* well-being should be understood as measures of well-being *per se*. The term 'subjective well-being' is used to refer to the aspect of well-being concerned with feelings (or emotions) and evaluative judgement of one's own life. It is controversial whether these subjective aspects are exhaustive of human well-being (for discussion, see Sumner 1996).

Since the validation procedures aim to be both an examination of whether a given scale really measures the target-attribute and an optimisation of the scale's capacity to do so, construct validation is relevant for answering the sceptical worry that the measurement results are best explained as measurement of disparate phenomena or as measurement artefacts.⁷

To better understand the relation between construct validation and measurement scepticism about well-being, let us outline what we take to be the core procedures of construct validation. Construct validation of a scale or measurement instrument is the attempt to certify whether the scale or instrument is really measuring what it purports to be measuring.⁸ Many sources of evidence can be relevant for the assessment of validity, relating not only to how scores of the test under consideration correlate with scores of other tests, but for instance also to the experience of taking the test, the real-life performance of test takers, the cognitive models of the cognitive processes involved in taking the test, experimental results, and the fit with well-confirmed models and theories in the field (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 1999; Embretson 1983; Slaney 2017). For the purpose of this paper, two central features of construct validation are of particular interest.

The first central feature is that construct validation as one central source of evidence uses correlational patterns concerning, on the one hand, scores of different scales measuring the same and other attributes and, on the other hand, scores of similar scales measuring the same and other attributes (Campbell & Fiske 1959; Strauss & Smith 2009). Suppose that we develop a scale, *S*, to measure well-being in a given population (e.g., in university students). Construct validation draws on the idea that if *S* really measures well-being in this population, then two types of correlational patterns should be observed between *S* and results from other scales measuring the same or other attributes.

One type of correlation would be *convergent* correlations: significant correlations, either positive or negative, between scores of *S* and scores of other relevant

7. Note that Alexandrova (Alexandrova & Haybron 2016) has been critical about the *actual practice* of construct validation in the sciences of well-being. Having this in mind, when we say that Alexandrova has proposed construct validation as a tool for answering measurement scepticism, we refer to her ideal of how construct validation should proceed (see Alexandrova 2017: 150). Even though Alexandrova (2017) might be the only author to have explicitly formulated a reply to well-being measurement scepticism based on construct validation, many naturalistically minded philosophers of well-being have presented similar ideas (e.g., Angner 2013a; Hersch 2022; Rodogno 2015; Bishop 2015).

8. We remain neutral with respect to the foundational debate about whether validity is a property of scales or inferences and interpretation of scores obtained by using the scales (see Borsboom, Cramer, Kievit, Scholten, & Franic 2009; Hood 2009; Sireci 2009; Slaney 2017). We proceed to talk as if validity is a property of scales, but our argument does not depend on it. It could be reformulated along the lines of the second conception.

scales that measure the same target attribute or other attributes that we would view as related. For example, we would expect individuals with high scores in S to score high also in other scales measuring well-being or other phenomena we would expect to be positively related to well-being (e.g., physical health, social life, income). Moreover, we would expect such individuals to score low in scales measuring, for example, psychopathology.

The other type of correlational pattern would be *discriminant* correlations: non-significant correlations between scores in S and scores in other scales measuring attributes we would view as unrelated. For example, we would expect that individuals with high scores in S would not correlate in any significant way with measurements of general intelligence or perceptual acuity, or other attributes or psychological phenomena that we would find unrelated to well-being.⁹

It is important to notice that even though the use of correlational patterns might appear data-driven, it rests upon the use of theory (Cronbach & Meehl 1955; Slaney 2017). With regards to measurements of well-being, characterising well-being theories come into play in several ways. To begin with, we need a theory to tell us something about the nature of the attribute under investigation. In our case, to design a well-being scale, we need a theory to tell us what well-being consists of for a given population (e.g., does it consist in emotional life or a sense of meaning?). Moreover, we need a theory to tell us what correlational patterns to expect, if any, and why we should expect positive or negative correlations, strong or weak correlations. The design and evaluation of results of the scale is only made possible by a theory of the target phenomenon—that is, a characterising well-being theory.

The second central feature of validation is that construct validation is holistically testing both a scale and the theory that underlies it (Alexandrova & Haybron 2016; Slaney 2017). Just as observing the relevant correlational patterns might lead us to revise our scale (e.g., delete, add, or reformulate items on our scale), construct validation requires that we are *open* to revising our theory of the target phenomenon (in this example, our characterising theory of what well-being consists in). Whether to revise the scale or the theory is a judgement call for the researcher, but she should be *open* to revising the theory over time. This means that if the researcher consistently obtains results suggesting the theory is inadequate, she might be rationally forced to revise it or at least to provide very good reasons for not doing so. Since the claim that construct validation requires openness to empirically based theory-revision is a central point of contention, let us elaborate on what the claim entails.

9. What discriminant correlations to expect when measuring well-being is not a trivial matter. For example, opposite to the example above, intelligence (or certain kinds of it) might very well be related to well-being (Dimitrijević, Marjanovic, & Dimitrijevic 2018).

A number of important ideas underlie this openness to theory-revision. One general idea of construct validity theory is that as a scientific psychologist or social scientist, you have no privileged, theory-neutral access to the target phenomenon you are trying to measure. If two scales, embodying two different and conflicting theories supposed to measure the same phenomenon, provide us with conflicting measurement results, we have no theory-neutral perspective from which we can settle which scale is correct.

This is a perfectly general point about establishment and calibration of measurement scales, also for physical measures like measures of temperature or pressure (Chang 2004; Tal 2017a, 2017b). Given only a vague theoretical understanding and a few basic assumptions, such as the assumptions that the target phenomenon is unified, well-behaved, and measurable, we can start measuring the target phenomenon by using crude ordinal measures and a vague theory. By repeated measuring with the same and different types of instruments, we slowly establish reliable scales. This might enable us to enrich our theory by rejecting false assumptions, increasing confidence in others, and adding new ones. This might enable us to improve the scales, and so on.

Historically, theories about construct validation have been committed to this coherentist feature of validation (Cronbach & Meehl 1955). According to Cronbach and Meehl, when engaged in psychological testing and measuring, the psychologist is trying to measure something which is not directly observable. You need to start from vague theory and crude scales. The only possible way forward seems to be adjustments back and forth between theory, scales, and data obtained with the scales. This is why construct validation requires that we are open to theory-revision on the basis of correlational data. That is, if we keep getting negative correlations when we expect them to be positive (or vice versa) or insignificant correlations when we expect them to be significant (or vice versa), then at some point it would be irrational not to revise one's theory. However, that we are required to remain *open* to theory-revision should not be conflated with the claim that construct validation requires us to *actually* revise our theory: as stressed, it is a judgement call whether to revise the scale, the theory, or to reject the data as faulty.

Openness to theory-revision is not only a general feature of validation; it is required if construct validation is supposed to answer well-being measurement scepticism. That is, the reason why construct validation can raise our confidence in our ability to measure well-being is because the procedures examine and potentially revise *both* the scale and the underlying theory of the phenomenon. Well-being measurement scepticism is the worry that well-being might not be the kind of thing that we can measure (by some current or all possible scales). To answer this sceptical worry, we need validation procedures that can make us confident that our scales are measuring the attribute as conceptualised by our

theory, but also that our conceptualisation and theory are appropriate in relation to the attribute. Construct validation addresses both kinds of confidence but only because it demands that both the involved scale and theories are open to revision (in our case, the relevant characterising well-being theories). This means that even if validation does not actually lead us to revise our characterising well-being theory (because we make other reasonable revisionary prioritisations), the only reason why construct validation actually raises our confidence in measuring well-being is because our theory was open to revision but nevertheless stood the test of correlational examination.

Summing up this line of reasoning, a basic assumption of construct validation theory is that the theorising and the measuring of an attribute proceed in parallel and are mutually adjusted since we have no theory-neutral access to the target attribute. This means that no theory enjoys an a priori privileged position in conceptualising the attribute. Different theories make different predictions about correlational patterns obtained with a variety of scales designed to measure the same, related, and unrelated attributes. The theory underlying the scales with the most attractive correlational patterns does, everything else being equal, appear to be the best supported conceptualisation of the attribute. In sum, construct validation intuitively appears to be a promising answer to well-being measurement scepticism (in both local and global forms).

3. Methodological Controversies about Openness to Theory-Revision

3.1. Construct Validation and Rational Theory Choice

How successful is a response to well-being scepticism that relies on construct validation? As we shall see in the following, for a response based on construct validation to work, it must make crucial assumptions about the methodology of well-being theorising.

Let us first revisit construct validation and its central assumption of openness to theory-revision by looking at an example. Suppose that we develop a scale, S^1 , to measure well-being among university students, and we build this on an underlying characterising well-being theory, T^1 . S^1 measures well-being by asking participants to assess their degree of self-discovery (see Waterman et al. 2010). T^1 states that the well-being of university students is a matter of students developing and realising their academic and general higher psychological capacities such as reflective skills, meaningful conversations, and nourishing of their talents (T^1 could draw inspiration from a version of perfectionism as known from moral philosophy, for instance, Wall 2017 and Hurka 1993).

There are (at least) three types of scenarios in which there could be a pressure on us to revise our theory of well-being that underlies our well-being scale. In relation to S^1 and T^1 specifically, consider how these scenarios might manifest themselves. First, *opposing correlations*, suppose we observe convergent correlation in the survey data such that students that score high in S^1 (degree of self-discovery) are among the students who use less time at peer socialising. However, T^1 is predicting the opposite: it predicts that individuals with high scores in S^1 would be among the students spending most time at peer socialising. Second, *silence about correlations*, suppose we observe a significant positive correlation such that students with high scores in S^1 also score high in scales that measure the level of ambition of one's own life (e.g., ambition on behalf of one's own career). Suppose that prior to making the measurements, T^1 did not predict anything about this relationship. Third, *null results*, suppose we observe discriminant correlation such that students with high scores in S^1 are not generally among the students with high emotional health and they are not less prone to psychopathologies such as affective disorders (e.g., depression and anxiety). However, suppose that T^1 , in accordance with the majority of other well-being theories, predicted that people with high degrees of well-being generally have higher degrees of emotional health and are significantly less prone to psychopathologies.¹⁰

Now, suppose that another scale, S^2 , which has the underlying characterising well-being theory T^2 , has been developed. T^2 is a hedonic well-being theory: it takes well-being to consist in the ratio of positive to negative mental states. Assume that S^2 (and other scales designed on the basis of T^2) exhibit an attractive degree of correlation: their measurement results correlate more attractively with reports on life satisfaction, income, social life, and other relevant dimensions than results of S^1 . Furthermore, assume that T^2 is better at predicting the convergent and discriminant correlational patterns with respect to related and unrelated attributes. Given that T^1 and T^2 are theories about the same attribute (namely, well-being of university students), all other things being equal, T^2 seems to be a better theory of what well-being of university students consists in (following the logic of construct validation).

If construct validation is supposed to work in the same way when investigating well-being as in other domains of psychology, the above three scenarios could (at least, in principle) put pressure on us to revise our *characterising* theory, T^1 . Each scenario could reasonably be taken to indicate that T^1 theorised about well-being in a mistaken way. Given the sketched relationships between T^1 and the observed correlations in the survey data, one could reasonably argue that

10. Such expectation might not be universal. See Westerhof and Keyes (2010) and Ryff and Keyes (1995) for an interesting view that suggests a more nuanced relation between well-being and mental illness.

the correlational patterns are evidence that the well-being of university students is *not* flourishing, at least not in the sense specified by T¹, and that it is more plausible that their well-being is a matter of hedonic mental states. Following the standard logic behind construct validation, the three possible scenarios could be relevant for justifying revision of our theory or choosing between different competing theories of the same attribute.

3.2. Methodology of Well-Being Theorising

Despite being common in psychology, this kind of correlation-based revision of characterising well-being theories is a controversial idea. This becomes clear when we consider two philosophical positions on the methodology of well-being theorising.

First, *methodological naturalism* about characterising well-being theorising is the position that well-being theorising should proceed in a fashion similar to theorising about other natural phenomena—that is, in direct relation to and informed by scientific measurement attempts. Methodological naturalism would in this sense be the view that the big questions about the nature of well-being should be answered by drawing on scientific, empirical work. Moreover, such naturalists would take operationalisation of well-being in terms of relevant measures to be a real constraint on characterising well-being theories. Many of the methodological naturalists would aim for a reflective equilibrium or coherentism where empirical data from social science are given equal weight to intuitions and conceptual analysis in the sense that strong empirical evidence can lead one to reject intuitions and adjust conceptual considerations.¹¹

Second, *methodological non-naturalism* about characterising well-being theorising is the position that the question of what essentially and non-derivatively makes a person's life go best, or what is non-instrumentally good for her, is a question with a strong evaluative dimension that should be decided by the methods found in (moral) philosophy.¹² That is, just as the appropriate way to theorise about what characterises the morally right action is to engage in systematic and careful deliberation using thought-experiments and counter-examples, the question of which characterising well-being theory is correct should be approached

11. Many contemporary authors seem to adopt the position of methodological naturalism, or a position closely related to it, for example, Rodogno (2015), Angner (2013a), Alexandrova (2017), Hersch (2022), Tiberius and Plakias (2010), Tiberius (2007; 2018), and Bishop (2015).

12. To be clear, we are not suggesting that moral philosophy is or should be committed methodological non-naturalism. We are merely assuming that methodological non-naturalism is a prevalent stance in moral philosophy (compared to philosophy of language or philosophy of mind).

by careful conceptual deliberation. This makes theorising about well-being an *a priori* activity. Methodological non-naturalists sometimes use possible states of affairs to argue against wellbeing theories—for example, imagine someone who is happy while also deprived of basic needs. However, whether these states actually obtain is irrelevant to them (e.g., does poverty correlate with happiness?). By contrast, the methodological naturalists think that whether these states obtain is relevant.¹³ A methodological non-naturalist does not consider well-being an empirical matter in the sense that empirical material (such as survey data or other measurements) could never decide the classic debates between competing characterising theories of well-being. Consequently, for a methodological non-naturalist, it is not necessarily a requirement for a characterising theory of well-being that it can be operationalised in terms of psychological and economic measures of well-being. To sum up, the methodological non-naturalist does not aim for a reflective equilibrium or coherentism in the above-sketched sense.¹⁴

The fundamental differences between methodological naturalism and non-naturalism are related to views about the relevant epistemic virtues of characterising well-being theories. Since methodological naturalists would take well-being theorising to be similar to or continuous with any other theorising in science, they might stress abductive-inductive reasoning and classical epistemic values such as accuracy, consistency, scope, simplicity, and fruitfulness (Kuhn 1977).

13. We thank a reviewer for suggesting this way of phrasing the difference between the naturalist and non-naturalist.

14. Feldman (2010) is a clear example of an author endorsing methodological non-naturalism. In general, Feldman (2010: chs. 12, 13, 14) objects to the contemporary tendency among philosophers to take scientific work to be relevant to characterising well-being theorising. After discussing arguments in favour of the relevance of scientific inquiry to a philosophy of well-being (or happiness), Feldman writes (2010: 269):

Furthermore, I have not said that [Richard J.] Davidson's research is of no philosophical interest. For all I know, it might be relevant to some philosophical question [on the nature of well-being or happiness]. Again, at present, I cannot think of any philosophical question to which that research would be relevant. . . . Furthermore, I have not said that the empirical research of others in the "positive psychology" field is either pointless or irrelevant to philosophical questions about happiness. Maybe there is some other researcher who has discovered something that bears on some philosophical question. In spite of the fact that I have looked, I have not found any such research. I doubt that it will be found.

Moreover, considering the method by which many contemporary moral philosophers present and defend their well-being theories, it seems plausible that methodological non-naturalism is a common position. Take for example the well-known well-being theories by Crisp (2006b), Kraut (2007), and Hurka (1993): they all proceed by traditional conceptual reflection to figure out what characterises human well-being. Hersch (2022: part 3) and Alexandrova (2017) share the impression that a significant number of (moral) philosophers endorse some form of methodological non-naturalism with regards to characterising well-being theories. See also Fumagalli's discussion of what he calls 'theory-based approaches' (Fumagalli 2022).

By contrast, since methodological non-naturalists would take well-being theorising to be essentially conceptual in nature, they would often take the goal of well-being theories to present necessary and sufficient conditions for something to count as well-being. According to the methodological non-naturalist, a central epistemic virtue of a well-being theory would be resistance to counter-examples.

The distinction between methodological naturalism and non-naturalism in characterising well-being theory is further supported by the way in which it relates to Hersch's (2022) distinction between well-being coherentism and foundationalism. According to well-being coherentism, beliefs about well-being are justified when they cohere with other mutually supportive beliefs about well-being. By contrast, well-being foundationalism is the view that beliefs about well-being are justified only with reference to some basic truths about well-being (Hersch 2022: part 3). Hersch argues that well-being coherentism should be developed along the lines of Chang's notion of 'epistemic iteration' (Chang 2004; see also our Section 3.4). Thus, coherentism embodies the epistemological outlook of construct validation and the openness to correlation-based theory-revision.

This suggests that methodological non-naturalism would often pair up with foundationalism (which is clearly Hersch's impression). According to methodological non-naturalism, there is an asymmetric dependency: Characterising theories constrain measurements, but measures cannot constrain and decide basic theoretical questions. According to methodological non-naturalism, characterising well-being theories are immune to empirical revision. By contrast, from the perspective of methodological naturalism, characterising well-being theories should cohere with our measurements and be open to empirical revision.

Summing up, where the methodological naturalist would ask "given insights from all these different ways to study well-being, what is well-being?", the methodological non-naturalist would say "given that well-being is *x*, all these other supposed ways to study well-being are actually studying something different".

3.3. Methodological Non-Naturalism and Openness to Theory-Revision

Methodological non-naturalism cannot accept that *characterising* well-being theories are open to revision based on correlations in survey data (although it can allow that *associative* theories are open to such revision). Consequently, methodological non-naturalism is incompatible with construct validation procedures as a way to respond to well-being measurement scepticism.

Maybe the conflict between these validation procedures and methodological non-naturalism is only apparent. One could object that since data always underdetermines one's choice of theory, empirical data never rationally forces

one to adjust one's theory. Correlation-based openness to theory-revision is so inconsequential that methodological non-naturalism can take it on board easily. Following this line of thought, the methodological non-naturalist might stress that construct validation never unambiguously demands theory-revision, and that methodological non-naturalism is in fact compatible with this validation procedure. In other words, since construct validation proceeds holistically by adjusting measurement scales and underlying theory in parallel, it is always a judgement call whether the involved scale or the theory should be the target of revision: the mismatch between data and theory might in principle be explained away by faulty measurement scales and measurement noise (Duhem 1914/1954). With this in mind, one could adopt holism about testing (Quine 1951) and claim that no observation can ever 'demand' a revision of one's well-being theory.

This objection does not work. As stressed above, even though it is true that validation procedures do not necessarily demand theory-revision in any given situation, validation demands that theories are *open* to a revision based on correlational data—meaning, at least, that researchers must provide reasons for not revising their theories in the light of inconsistent correlational patterns. In other words, as conflicting evidence piles up by means of well-tested measurement scales, at some point it might be most rational to revise one's theory.¹⁵ Dogmatism about one's theory would be untrue to the spirit of construct validation and rob the procedures of their relevance as an answer to well-being measurement scepticism.¹⁶ Given the absence of a theory-neutral direct access to the target attribute, we need to remain open to revision of our characterising well-being theories: that is how we can gain confidence that our theories are actually conceptualising the attribute properly and that our scales are actually measuring a real phenomenon.

However, this is exactly the kind of openness to a correlation-based revision of characterising well-being theories that a methodological non-naturalist would find mistaken. Recall how methodological non-naturalism considers well-being to be profoundly connected to value theoretical and moral questions and the associated philosophical methodology. Unless a convincing argument is provided prior to the presentation of the procedures of construct validation, many moral philosophers would most likely be reluctant to accept that characterising well-being theories should be open to an empirically justified theory-revision.

15. There are important examples of theory-revision due to observation of correlational patterns—for instance, in intelligence research (Jensen 1987), in emotion research (Barrett 2006), in research on implicit attitudes (Feest 2020), and in cognitive motor neuroscience (Grünbaum & Christensen 2020).

16. One might back up this claim with arguments to the effect that radical holism about testing leads to irrationality (Laudan 1990) or arguments showing that that auxiliary claims and measurement scales often can or have undergone independent testing (Chang 2004; Sober 1999).

Many value theorists and moral philosophers would probably view the empirical openness to revision in construct validation as an inappropriate way to blend descriptive (patterns of correlational results) and normative matters (ideas of what has non-instrumental value in life).

Methodological non-naturalism is incompatible with the kind of openness of characterising well-being theories that is involved in construct validation and which is required for validation procedures to be an answer to measurement scepticism. The important implication of this is that construct validation does not work as a reply to well-being measurement scepticism if such scepticism is motivated on grounds of methodological non-naturalism. In such a case, the relevant sceptic would not accept that construct validation could increase our confidence in measuring well-being since this increase presupposes that our characterising well-being theories are open to revision based on correlational data.

One reasonable question now becomes whether well-being measurement scepticism as motivated by methodological non-naturalism is a plausible position and whether any authors occupy it. The answer is ‘yes’—in short, because there appears to be a straightforward connection between methodological non-naturalism and measurement scepticism. Take the following two considerations.

First, contemporary authors explicitly articulating a form of well-being measurement scepticism often endorse methodological non-naturalism with respect to characterising well-being theories. For example, Hausman (2011: 141) explicitly endorses a characterising well-being theory similar to that of Kraut’s (2007), which takes well-being to consist in flourishing. Hausman’s way of determining that this is the correct characterising well-being theory relies on a non-naturalist methodology: his identification of well-being with flourishing is motivated by philosophical reflection, not by examination of correlational patterns or any other common scientific procedure (see for example the reasoning in Hausman 2011: 64, 130).

Second, methodological non-naturalism would in many cases make well-being measurement scepticism a natural position. If one reasons in accordance with the schema “given that well-being is x , all these other supposed ways to study well-being are actually studying something different”, measures that are not measuring x are not measuring well-being at all. Take the example of Feldman (2010) who accepts a form of methodological non-naturalism and explicitly expresses a form of measurement scepticism with considerable scope (Feldman 2010: 235):

While I recognize that all of these are genuine sources of worry concerning these Satisfaction with Life Domain test instruments, I think the problem runs far deeper. I think the problem is that, even if all the wrinkles could be ironed out, any such instrument would still be measuring the

wrong thing. That is, the test instrument would not be measuring happiness [or well-being].

White's (2013) measurement scepticism is also motivated by his endorsement of methodological non-naturalism.

In general, methodological non-naturalism often aligns with some form of well-being measurement scepticism with regards to the current supply of well-being scales. If one is an invariantist (to use Alexandrova's terminology) in the sense that one and only one characterising well-being theory is true (say, some version of hedonism), then one will tend to be sceptic towards the ability of scales inspired by any other characterising well-being theory (say, objective list theory) to measure well-being (at least of their ability to measure well-being in any direct way).

Importantly, this is not to say that an invariantist methodological non-naturalist necessarily will be sceptical towards any scientific measurement of well-being. She might indeed be very positive towards measurements of well-being inspired by the characterising well-being theory that she supports. Neither is it to say that such an invariantist will find well-being scales inspired by other theories totally useless—such an invariantist could think that these scales measured well-being in an indirect way. All we suggest is that there are likely links between methodological non-naturalism and well-being measurement scepticism, not that there is any strict implication. Considering the fact that many moral philosophers appear to be methodological non-naturalists (e.g., Crisp 2006a; 2006b; Kraut 2007; Hurka 1993), it therefore seems reasonable to expect that moral philosophy contains several well-being measurement sceptics who endorse scepticism on methodological non-naturalist grounds.

With the above points in mind, well-being measurement scepticism motivated by a form of methodological non-naturalism is an important and plausible position. Given the assumption that methodological non-naturalism cannot accept empirical openness to theory revision, it is clear that construct validation is not a neutral common ground on which to refute measurement scepticism. Construct validation can serve as a successful answer to measurement scepticism only if both the sceptic and the non-sceptic about well-being measurement accept methodological naturalism.

3.4. The Need for a Further Argument

Suppose that one accepts well-being measurement scepticism on methodological non-naturalistic grounds. Given this assumption, an argument against measurement scepticism would need to provide a further argument for why char-

acterising well-being theories should be open to a correlation-based revision as required by construct validation. That is, we would need to show that methodological non-naturalists should accept the right form of openness to correlation-based theory-revision. In the following, we discuss two possible ways to provide such an argument and conclude that both are begging the question against methodological non-naturalism.

(a) Alexandrova (2017: xl) has introduced a distinction between *high-level* and *mid-level* theories of well-being. High-level well-being theories are theories as known from traditional moral philosophy—such as hedonism, preference-satisfaction theory, and objective list theory. Such theories specify what non-derivatively or ultimately makes a person's life go well, they apply to humans in general, and consider the whole of an individual's life (Parfit 1984: app. I). By contrast, *mid-level* theories of well-being are theories that state what well-being consists in for a specific kind of individual (e.g., one mid-level theory might specify what well-being is for children, while another specifies it for cancer patients). Such theories state what is good for individuals in virtue of these individuals being members of a specific group. Mid-level theories might draw upon elements from multiple different high-level theories. A criterion specific to mid-level theories is that they must be translatable into measurement procedures. According to Alexandrova (2017), philosophical theories are typically high-level theories, while the characterising well-being theories used in psychology and economics are mid-level theories.

The distinction between high-level and mid-level theories would enable the reply that only mid-level theories should be open to correlation-based revision, while high-level theories should be immune to this kind of revision (Alexandrova seems to hold such a view, 2017: part 5.3). Hence, the suggestion would be that correlational patterns should only inform us on how to theorise about what well-being consists in for a specific group of individuals, not for human beings in general. By making this distinction between levels of well-being theories, it seems possible to hold a position that preserves both the intuitions of immunity and the openness to a correlation-based revision of characterising well-being theories.

Should the methodological non-naturalist accept that characterising well-being theories at the middle-level are open to empirically based theory-revision? Probably not. Arguing for the appropriateness of only mid-level theory-revision requires a further argument in addition to simply highlighting the two levels of well-being theorising, and such an argument appears to beg the question against the methodological non-naturalist. Consider the following two stances on high-level well-being theories.

First, assume that a person is a *high-level invariantist* with respect to well-being theories, if she thinks that one, and only one, of the high-level theories is

true (Crisp 2006a and 2006b is an invariantist with regards to hedonism, whereas Kraut 2007 is with regards to flourishing). Assume moreover that this invariantist is a methodological non-naturalist and further that she thinks that any justification of a well-being theory should follow a foundationalist procedure according to which the fundamental elements of the specific high-level theory that is true should be imported into the mid-level theory (this seems reasonable, since this specific high-level theory is viewed as the *only* true view of the fundamental nature of well-being).

Given the foundationalist procedure of this high-level invariantist, the invariantist would argue that correlational patterns are of only very limited relevance to the evaluation of mid-level theories. That is, correlational patterns could be of use only if the relevant competing mid-level theories were all based on the specific high-level theory taken to be true by the invariantist (that is, such patterns would only be of relevance in relation to an 'internal competition' between different mid-level versions stemming from the same higher-level theory). Furthermore, the invariantist might argue, if the competing mid-level theories were based upon different high-level theories (a kind of 'external competition'), the selection between them should be made in virtue of the high-level theory that figured as the foundation of the mid-level theory. The mid-level theory with the correct high-level inspiration would come out as a winner no matter what correlational patterns supported it compared to other mid-level theories. Consequently, the high-level invariantist could state that correlational patterns of survey data could never decide the main elements of mid-level theories, though they might play a role in deciding how one should translate the correct high-level theory into a suitable mid-level theory. That is, from the perspective of a methodological non-naturalist invariantist about well-being, a characterising mid-level theory could only be open to empirically based theory-revision in the sense of finding the right translation from high-level to mid-level theory; it could never be open to empirically based theory-revision in the sense of playing a role in specifying what well-being is.

Second, assume that a person is a *high-level variantist* in the sense that she thinks that one needs to draw on more than one of the high-level theories to understand the nature of well-being (Alexandrova seems to endorse this kind of pluralism, 2017: 45). Yet, despite being a variantist, she might still be a methodological non-naturalist also in relation to mid-level theories. That is, she thinks that since mid-level theories are theories of what makes a specific group of individuals' lives go best with respect to them being members of this group, the method to arrive at plausible mid-level theories is the same as arriving at high-level theories (namely by the method of systematic reflection using basic intuitions, thought-experiments, and counter-examples). Put differently, even though the variantist is open to deciding between different mid-level theories

with different high-level inspirations, she might take the very nature of well-being theorising to be of such a kind that this decision is *not* to be justified by correlational patterns of survey data (as required by construct validation), but by conceptual analysis.

With the above in mind, it is clear that simply stressing the distinction between different levels of well-being theories should not convince the methodological non-naturalist to be open to empirically based theory-revision. Both a high-level invariantist and a high-level variantist could be methodological non-naturalists and insist that the correctness of mid-level theories should be decided by conceptual analysis and not by a coherentist procedure open to theory-revision. The use of correlational patterns as an evidence source for mid-level well-being theorising requires another argument and this argument must address the basic disagreement about the methodology and epistemology of characterising well-being theories.

(b) The second way to provide an independent argument to address the non-naturalist's rejection of the openness to theory-revision would more explicitly follow a coherentist line and argue that to ensure scientific progress in the science of well-being, characterising well-being theories must be open to revision justified by correlational patterns.

First, one might point to what Hersch (2022) calls 'the coordination problem'. This refers to the problem of coordinating and integrating the knowledge gained in the philosophy of well-being and in the science of well-being, implying that the theories and methods of each domain can benefit from each other. It seems reasonable to say that for the ambition of coordination to be satisfied, characterising well-being theories must be open to revision justified by correlational patterns.

Second, one might use the notion of epistemic iteration by Chang (2004). The basic idea is that scientific progress is ensured by an epistemic process in which one adopts a certain system of knowledge (in our case, a set of beliefs and measurement procedures about well-being). This system is then utilised in accordance with two primary commitments. First, one uses it under a principle of respect (the idea that the system is an achievement and should not be discarded, unless a better system is in fact available). Second, one uses it under the imperative of progress (the idea that the system is not correct and needs improvement). In this sense, construct validation is a process of epistemic iteration. That is, (a) one adopts a system of knowledge (a theory and a related scale), (b) one applies it and corrects it by revising the scale and theory on the basis of correlational patterns (as well as other sources of data and theoretical development), (c) the steps are iterated and the result is a progressively optimised system of knowledge of what well-being is and how to measure it. Hence, one could argue that characterising well-being should be open to revision justified by correlational patterns,

since this is crucial to ensure that the theorising about well-being follow the general and plausible view of epistemic progress as characterised by epistemic iteration.

Again, this kind of reasoning begs the question by assuming a specific kind of methodological and epistemological outlook on well-being theorising.¹⁷ According to the methodological non-naturalist, the coordination problem is not to be solved by philosophical theories being revisable on the ground of empirical results. This would not count as any progress in understanding what well-being is. Instead, progress would be made if scales were better designed in relation to the most plausible characterising well-being theories—theories developed using the standard methods of (moral) philosophy. Moreover, such a non-naturalist could remain unimpressed by the reference to epistemic iteration. She might stress that such epistemic iteration might have worked for measurements of physical phenomena, for example temperature or mass, but since well-being is a phenomenon with a decisive evaluative dimension, the iteration approach is inappropriate here.

It is not a trivial task to provide an independent argument for the openness to empirically based theory-revision of characterising well-being theories that does not beg the question against methodological non-naturalism. However, if construct validation is to work as an answer to measurement scepticism motivated on methodological non-naturalist grounds, it must provide exactly such an argument.

4. Methodological Non-Naturalism and Further Implications

Endorsing methodological non-naturalism about well-being is not without problems. Let us briefly examine two interesting implications of accepting the position.

First, accepting methodological non-naturalism would have consequences for our understanding of the possible roles of construct validation in the science of well-being. Since non-naturalism implies that openness to a correlation-based revision of characterising well-being theories is inappropriate, construct validation could *not* help us in any way with figuring out what the fundamental nature of well-being is. Construct validation could only inform us on our ability to measure phenomena associated to well-being. For example, consider the construct validation as carried out by Arthaud-Day et al. (2005) who, as already mentioned, examined the tripartite theory (Diener et al. 1997). Since, given meth-

17. For instance, Hassoun (2019) and Fumagalli (2022) would most likely raise concerns about this coherentist spirit of theory-revision.

odological non-naturalism, confirmation of the tripartite theory by correlational data could only be a confirmation of the theory as an associative theory, all the procedures of construct validation could tell us is something about how self-reports of affect and cognitive evaluation of one's life are related to each other. Since the informational and justificatory traffic only goes in one direction (from philosophical specification of what well-being is to empirical measurements), construct validation could only play a role in helping us understand how best to measure what we already know to be well-being. It could not play a real role in the specification of the nature of well-being

Second, these considerations point to a more general implication of methodological non-naturalism: namely, that such a position divorces philosophical theorising about the nature of well-being from the science of well-being. By embracing methodological non-naturalism and providing characterising well-being theories with immunity to an empirically based revision, we separate philosophy from the empirical investigation of well-being and hinder any use of scientific procedures to answer measurement scepticism about well-being. Given widespread disagreement among methodological non-naturalists about the correct characterising theory of well-being, methodological non-naturalism seems like an unattractive theoretical point of departure for any attempt to measure well-being directly. Given the importance of measurements of well-being for the design and evaluation of social policies and interventions, this would appear to be unfortunate.

It begins to look like the nature of well-being as a graded phenomenon and the imperative to empirically measure it in order to evaluate policies force a kind of dilemma upon us. If we adopt methodological naturalism about well-being theorising, we have a way to answer well-being measurement scepticism. Methodological naturalism implies that developing characterising well-being theories is a task also for science. This implies an openness to data driven revision of characterising well-being theories, which is unacceptable for many value theorists and moral philosophers. It would fundamentally neglect the normative dimension of well-being. If we adopt methodological non-naturalism, we can preserve the fundamental normative character of well-being. Methodological non-naturalism implies that developing characterising well-being theories is solely a task for philosophy. This choice implies a separation of philosophy and science. Consequently, measurement scepticism cannot be answered by reference to traditional scientific procedures, like construct validation.

Ideally, the design, choice, and evaluation of social policies and interventions should be informed by whether they provide members a given population with more or less well-being. Such policy-making and policy-evaluation would be greatly facilitated by objective measurements of well-being. Well-being mea-

surement scepticism presents us with an important challenge to this ideal. Meeting this challenge requires us to explicate basic methodological assumptions about the grounds on which the scepticism is articulated and countered. The basic challenge to the methodological naturalist's reply to well-being measurement scepticism is to demonstrate that well-being is a measurable psychological attribute just like personality traits and intelligence. However, many philosophers would object that well-being has substantial and irreducible evaluative dimensions. The basic challenge to the methodological non-naturalist is that the position has no convincing answers to well-being measurement scepticism. The consequence would seem to be that measures of well-being could not justifiably feed into our choice of social policies and interventions.

To be sure, the person who is whole-heartedly committed to either methodological naturalism or non-naturalism would see no dilemma here. The choice would be straightforward for her. Yet, we suspect that many philosophers would be hesitant to take position on this issue, especially in the light of the above-mentioned dilemma like situation.

5. Conclusion

We can conclude that the psychometric procedures of construct validation can be used to answer well-being measurement scepticism only if this scepticism is motivated on methodological naturalist grounds. This is clear since construct validation both as a set of general psychometric procedures and as a reply to measurement scepticism requires that characterising well-being theories are open to a correlation-based revision—an openness the methodological non-naturalist cannot accept. We have moreover argued that it seems difficult to provide an independent argument for the appropriateness of this openness that does not beg the question against methodological non-naturalism. Finally, we suggested that both methodological naturalism and non-naturalism comes with implications that could be perceived as unattractive. Methodological naturalism about well-being theorising implies an openness to revision of characterising well-being theories based on survey data, whereas methodological non-naturalism implies a methodological separation of the science of well-being and the philosophy of well-being.

Acknowledgements

We are very grateful for the comments and discussions concerning central ideas of the paper provided by the participants at the workshop 'Issues in measure-

ments of well-being' at University of Copenhagen, December 2019. In particular, we are grateful to Anna Alexandrova, Eric Angner, Valerie Tiberius, and Willem van der Deijl. Moreover, we thank you Alexander Heap and Marion Goodman for their comments on earlier versions of the manuscript.

References

- Alexandrova, Anna (2005). Subjective Well-Being and Kahneman's 'Objective Happiness'. *Journal of Happiness Studies*, 6(3), 301–24.
- Alexandrova, Anna (2017). *A Philosophy for the Science of Well-Being*. Oxford University Press.
- Alexandrova, Anna and Daniel M. Haybron (2016). Is Construct Validation Valid? *Philosophy of Science*, 83(5), 1098–109.
- Alkire, Sabina (2016). The Capability Approach and Well-Being Measurement for Public Policy. In M. D. Adler and M. Fleurbaey (Ed.), *The Oxford Handbook of Well-Being Public Policy* (615–45). Oxford University Press.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Anand, Sudhir and Amartya Sen (1994). Human Development Index: Methodology and Measurement (No. HDOCPA-1994-02). *New York: Human Development Report Office, United Nations Development Programme*.
- Angner, Eric (2010). Are Subjective Measures of Well-Being 'Direct'? *Australasian Journal of Philosophy*, 89(1), 115–30.
- Angner, Eric (2013a). Is Empirical Research Relevant to Philosophical Questions? *Res Philosophica*, 90(3), 343–63.
- Angner, Eric (2013b). Is It Possible to Measure Happiness? *European Journal for Philosophy of Science*, 3(2), 221–40.
- Arthaud-Day, Marne L., Joseph C. Rode, Christine H. Mooney, and Janet P. Near (2005). The Subjective Well-Being Construct: A Test of Its Convergent, Discriminant, and Factorial Validity. *Social Indicators Research*, 74(3), 445–76.
- Barrett, Lisa Feldman (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science* 1(1), 28–58.
- Bechtoldt, Harold P. (1959). Construct Validity: A Critique. *American Psychologist*, 14, 619–29.
- Benjamin, Dan, Kristen Cooper, Ori Heffetz, and Miles Kimball (2020). Self-Reported Wellbeing Indicators Are a Valuable Complement to Traditional Economic Indicators but Are Not Yet Ready to Compete with Them. *Behavioural Public Policy*, 4(2), 198–209.
- Bishop, Michael (2015). *The Good Life: Unifying the Philosophy and Psychology of Well-Being*. Oxford University Press.
- Bok, Sissela (2010). *Exploring Happiness: From Aristotle to Brain Science*. Yale University Press.
- Bond, Timothy N. and Kevin Lang (2019). The Sad Truth about Happiness Scales. *Journal of Political Economy*, 127(4), 1629–40.

- Borsboom, Denny, Angélique O. J. Cramer, Rogier A. Kievit, Annemarie Z. Scholten, and Sanja Franic (2009). The End of Construct Validity. In R. Lissitz (Ed.), *The Concept of Validity* (135–70). Information Age Publishers.
- Brodbeck, May (1957). The Philosophy of Science and Educational Research. *Review of Educational Research*, 27, 427–40.
- Campbell, Donald T. and Donald W. Fiske (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56(2), 81–105.
- Cartwright, Nancy (1991). Replicability, Reproducibility, and Robustness – Comments on Harry Collins. *History of Political Economy*, 23(1), 143–55.
- Chang, Hasok (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.
- Clark, Andrew E. (2016). SWB as a Measure of Individual Well-Being. In M. D. Adler and M. Fleuerbaey (Ed.), *The Oxford Handbook of Well-Being Public Policy* (518–53). Oxford University Press.
- Crisp, Roger (2006a). Hedonism Reconsidered. *Philosophy and Phenomenological Research* 73(3), 619–45.
- Crisp, Roger (2006b). *Reasons and the Good*. Clarendon Press.
- Crisp, Roger (2017). Well-Being. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). <https://plato.stanford.edu/archives/fall2017/entries/well-being/>
- Cronbach, Lee J. and Paul E. Meehl (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281–302.
- Diener, Ed and Louis Tay (2014). Review of the Day Reconstruction Method (DRM). *Social Indicators Research*, 116(1), 255–67.
- Diener, Ed, Richard Lucas, Ulrich Schimmack, and John Helliwell (2009). *Well-Being for Public Policy*. Oxford University Press.
- Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49(1), 71–75.
- Diener, Ed, Shigehiro Oishi, Richard E. Lucas (1997). Recent Findings on Subjective Well Being. *Indian Journal of Clinical Psychology*, 24, 25–41.
- Dimitrijević, Ana A., Zorana J. Marjanovic, and Aleksandar Dimitrijevic (2018). Which-ever Intelligence Makes You Happy: The Role of Academic, Emotional, and Practical Abilities in Predicting Psychological Well-Being. *Personality and Individual Differences*, 132, 6–13.
- Duhem, Pierre (1954). *The Aim and Structure of Physical Theory* (P. W. Wiener, Trans.). Princeton University Press. (Original work published 1914)
- Embretson, Susan (1983). Construct Validity: Construct Representation versus Nomothetic Span. *Psychological Bulletin*, 93, 179–97.
- Fabian, Mark (2022). Scale Norming Undermines the Use of Life Satisfaction Scale Data for Welfare Analysis. *Journal of Happiness Studies*, 23(4), 1509–41.
- Feest, Uljana (2020). Construct Validity in Psychological Tests – The Case of Implicit Social Cognition. *European Journal for Philosophy of Science*, 10(1), 1–24.
- Feldman, Fred (2010). *What Is This Thing Called Happiness?* Oxford University Press.
- Frey, Bruno S. and Alois Stutzer (2002). *Happiness and Economics*. Princeton University Press.
- Fumagalli, Roberto (2019). (F)utility Exposed. *Philosophy of Science*, 86(5), 955–66.
- Fumagalli, Roberto (2022). A Reformed Division of Labor for the Science of Well-Being. *Philosophy*, 97(4), 1–35.

- Grünbaum, Thor and Mark Schram Christensen (2020). Measures of Agency. *Neuroscience of Consciousness*, 6(1), niaa019.
- Hassoun, Nicole (2019). Thoughts on Philosophy and the Science of Well-Being. *Res Philosophica*, 96(4), 521–28.
- Hausman, Daniel M. (2010). Hedonism and Welfare Economics. *Economics and Philosophy*, 26, 321–44.
- Hausman, Daniel M. (2011). *Why Satisfy Preferences?* (No. 1124). Papers on Economics and Evolution. Max-Planck-Institute für Ökonomik.
- Hausman, Daniel M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press.
- Hausman, Daniel M. (2015). *Valuing Health: Well-Being, Freedom, and Suffering*. Oxford University Press.
- Haybron, Daniel M. (2007). Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall. *Nous*, 41(3), 394–428.
- Haybron, Daniel M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford University Press.
- Haybron, Daniel M. (2020). Happiness. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). <https://plato.stanford.edu/archives/sum2020/entries/happiness/>
- Hersch, Gil (2022). Well-Being Coherentism. *British Journal for the Philosophy of Science*, 73(4), 1045–65.
- Hood, S. Brian (2009). Validity in Psychological Testing and Scientific Realism. *Theory & Psychology*, 19(4), 451–73.
- Hurka, Thomas (1993). *Perfectionism*. Oxford University Press.
- Ingelström, Mats and Willem van der Deijl (2021). Can Happiness Measures Be Calibrated? *Synthese*, 199(3), 5719–46.
- Jensen, Arthur R. (1987). The “g” beyond Factor Analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, and J. C. Witt (Eds.), *Buros-Nebraska Symposium on Measurement & Testing, Vol. 3: The Influence of Cognitive Psychology on Testing* (87–142). Lawrence Erlbaum Associates.
- Joshanloo, Mohsen, Vincenza Capone, Giovanna Petrillo, and Daniela Caso (2017). Discriminant Validity of Hedonic, Social, and Psychological Well-Being in Two Italian Samples. *Personality and Individual Differences*, 109, 23–27.
- Kaasa, Stein, Arne Mastekaasa, and S. Naess (1988). Quality of Life of Lung Cancer Patients in a Randomized Clinical Trial Evaluated by a Psychosocial Well-Being Questionnaire. *Acta Oncologica*, 27(4), 335–42.
- Kahneman, Daniel, Ed Diener, and Edward Schwarz (Eds.) (1999). *Well-Being: The Foundations of Hedonic Psychology*. Russell Sage Foundation Press.
- Kaiser, Casper (2022). Using Memories to Assess the Intrapersonal Comparability of Wellbeing Reports. *Journal of Economic Behavior & Organization*, 193, 410–42.
- Koushede, Vibeke, Mathias Lasgaard, Carsten Hinrichsen, Charlotte Meilstrup, Line Nielsen, Signe Boe Rrayce, Manuel Torres-Sahli, Dora Gudrun Gudmundsdottir, Sarah Stewarts-Brown, and Ziggi I. Santini (2019). Measuring Mental Well-Being in Denmark: Validation of the Original and Short Version of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS and SWEMWBS) and Cross-Cultural Comparison across Four European Settings. *Psychiatry Research*, 271, 502–9.
- Kraut, Richard (2007). *What Is Good and Why: The Ethics of Well-Being*. Harvard University Press.

- Kuhn, Thomas (1977). Objectivity, Value, and Theory Choice. In *The Essential Tension: Selected Studies in Scientific Tradition and Change* (320–39). The University of Chicago Press.
- Laudan, Larry (1990). Demystifying Underdetermination. In C. W. Savage (Ed.), *Scientific Theories* (267–97). Vol. 14 of Minnesota Studies in the Philosophy of Science. University of Minnesota Press.
- Lucas, Richard E., Shigehiro Oishi, and Ed Diener (2016). What We Know about Context Effects in Self-Report Surveys of Well-Being: Comment on Deaton and Stone. *Oxford Economic Papers*, 68(4), 871–76.
- Lui, P. Priscilla and Gaithri A. Fernando (2018). Development and Initial Validation of a Multidimensional Scale Assessing Subjective Well-Being: The Well-Being Scale (WeBS). *Psychological Reports*, 121(1), 135–60.
- Lundgren-Nilson, Åsa, Ingibjörg H. Jonsdottir, Gunnar Ahlborg Jr., and Alan Tennant (2013). Construct Validity of the Psychological General Well Being Index (PGWBI) in a Sample of Patients Undergoing Treatment for Stress-Related Exhaustion: A Rasch Analysis. *Health and Quality of Life Outcomes*, 11(2), 1–9.
- Lyubomirsky, Sonja and Heidi S. Lepper (1999). A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research*, 46(2), 137–55.
- Margolis, Seth, Eric Schwitzgebel, Daniel J. Ozer, and Sonja Lyubomirsky (2021). Empirical Relationships among Five Types of Well-Being. In M. T. Lee, L. D. Kubzansky, and T. J. VanderWeele (Eds.), *Measuring Well-Being: Interdisciplinary Perspectives from the Social Sciences and the Humanities* (460–92). Oxford University Press.
- Markus, Keith A. and Denny Borsboom (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. Routledge.
- Martela, Frank and Kennon M. Sheldon (2019). Clarifying the Concept of Well-Being: Psychological Need Satisfaction as the Common Core Connecting Eudaimonic and Subjective Well-Being. *Review of General Psychology*, 23(4), 458–74.
- McDonald, Roderick P. (2013). Modern Test Theory. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 1, 118–43). Oxford University Press.
- Michell, Joel (2003). The Quantitative Imperative: Positivism, Naïve Realism and the Place of Qualitative Methods in Psychology. *Theory & Psychology*, 13, 5–31.
- Michell, Joel (2008). Is Psychometrics Pathological Science? *Measurement*, 6, 7–24.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.
- Pavot, William and Ed Diener (2009). Review of the Satisfaction with Life Scale. In Ed Diener (Ed.), *Assessing Well-Being: The Collected Works of Ed Diener*. Social Indicators Research Series 39. Springer.
- Plant, Michael (2020). A Happy Possibility about Happiness (and Other Subjective) Scales: An Investigation and Tentative Defence of the Cardinality Thesis. *Happier Lives Institute Working Paper*.
- Quine, Willard V. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20–43.
- Rodogno, Raffaele (2015). *Well-Being, Science, and Philosophy*. In J. H. Søraker, J. W. van der Rijt, J. de Boer, P. H. Wong, and P. Brey (Eds.), *Well-Being in Contemporary Society* (39–57). Springer.
- Ryff, Carol D. and Corey Lee M. Keyes (1995). The Structure of Psychological Well Being Revisited. *Journal of Personality and Social Psychology*, 69(4), 719–27.

- Saunders, Douglas S. and Robert W. Burgoyne (2002). Evaluating Health-Related Well-being Outcomes among Outpatient Adults with Human Immunodeficiency Virus Infection in HAART Era. *International Journal of STD & AIDS*, 13(10), 683–90.
- Schneider, Leann and Ulrich Schimmack (2009). Self-Informant Agreement in Well-Being Ratings: A Meta-Analysis. *Social Indicators Research*, 94(3), 363–76.
- Schröder, Carsten and Shlomo Yitzhaki (2017). Revisiting the Evidence for Cardinal Treatment of Ordinal Variables. *European Economic Review*, 92, 337–58.
- Schwarz, Nobert and Fritz Strack (1991). Evaluating One's life: A Judgement Model of Subjective Well-Being. In F. Strack, M. Argyle, and N. Schwarz (Eds.), *Subjective Well-Being* (27–47). Pergamon Press.
- Shadish, William, Tomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company.
- Sireci, Stephen G. (2009). Packing and Unpacking Sources of Validity Evidence: History Repeats Itself Again. In R. Lissitz (Ed.), *The Concept of Validity* (19–38). Information Age Publishers.
- Slaney, Kathleen L. (2017). *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions*. Palgrave Macmillan.
- Sober, Elliot (1999). Testability. *Proceedings and addresses of the American Philosophical Association*, 73(2), 47–76.
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 667–80.
- Stone, Aarthur A. and Christopher Mackie (Eds.) (2013). *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience*. National Academies Press.
- Strack, Fritz, Norbert Schwarz, Birgitte Chassein, Dieter Kern, and Dirk Wagner (1990). The Salience of Comparison Standards and the Activation of Social Norms: Consequences of Judgements of Happiness and Their Communication. *British Journal of Social Psychology*, 29(4), 303–14.
- Strauss, Milton E. and Gregory T. Smith (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, 5, 1–25.
- Sumner, L. Wayne (1996). *Welfare, Happiness, and Ethics*. Oxford University Press.
- Tal, Eran (2017a). Calibration: Modelling the Measurement Process. *Studies in History and Philosophy of Science Part A*, 65, 33–45.
- Tal, Eran (2017b). Measurement in Science. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). <https://plato.stanford.edu/archives/fall2017/entries/measurement-science/>
- Taylor, Tim Edwin (2015). The Markers of Wellbeing: A Basis for a Theory-Neutral Approach. *International Journal of Wellbeing*, 5(2), 75–90.
- Tiberius, Valerie (2007). Substance and Procedure in Theories of Prudential Value. *Australasian Journal of Philosophy*, 85(3), 373–91.
- Tiberius, Valerie (2018). *Well-Being As Value Fulfillment*. Oxford University Press.
- Tiberius, Valerie and Alexandra Plakias (2010). Well-Being. In John M. Doris (Ed.), *The Moral Psychology Handbook*. Oxford University Press.
- van der Deijl, Willem (2017a). Are Measures of Well-Being Philosophically Adequate? *Philosophy of Social Sciences*, 47(3), 209–34.
- Van der Deijl, Willem (2017b). Which Problem of Adaptation? *Utilitas*, 29(4), 474–92.
- van Dierendonck, Dirk (2005). The Construct Validity of Ryff's Scales of Psychological Well-Being and Its Extension with Spiritual Well-Being. *Personality and Individual Differences*, 36(3), 629–43.

- Wall, Steven (2017). Perfectionism in Moral and Political Philosophy. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 ed.). <https://plato.stanford.edu/archives/sum2019/entries/perfectionism-moral/>
- Waterman, Alan S., Seth J. Schwartz, Byron L. Zamboanga, Russell D. Ravert, Michelle K. Williams, V. Bede Agocha, Su Y. Kim, and M. Brent Donnellan (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric Properties, Demographic Comparisons, and Evidence of Validity. *The Journal of Positive Psychology*, 5(1), 41–61.
- Watson, David, Lee Anna Clark, and Auke Tellegen (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063–70.
- Westerhof, Gerben J. and Corey L. M. Keyes (2010). Mental Illness and Mental Health: The Two Continua Model Across the Lifespan. *Journal of Adult Development*, 17(2), 110–19.
- White, Mark D. (2013). Can We—and Should We—Measure Well-Being? *Review of Social Economy*, 71(4), 526–33.
- Wimsatt, William C. (1981). Robustness, Reliability, and Overdetermination. In *Reengineering Philosophy for Limited Beings* (43–74). Harvard University Press.
- Woodward, James (2006). Some Varieties of Robustness. *Journal of Economic Methodology*, 13(2), 219–40.
- Wren-Lewis, Sam (2014). How Successfully Can We Measure Well-Being through Measuring Happiness? *South African Journal of Philosophy*, 33(4), 417–32.