



Can welfare be measured with a preference-satisfaction index?

Willem van der Deijl

To cite this article: Willem van der Deijl (2017): Can welfare be measured with a preference-satisfaction index?, Journal of Economic Methodology, DOI: [10.1080/1350178X.2017.1413586](https://doi.org/10.1080/1350178X.2017.1413586)

To link to this article: <https://doi.org/10.1080/1350178X.2017.1413586>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 15 Dec 2017.



Submit your article to this journal [↗](#)



Article views: 138



View related articles [↗](#)



View Crossmark data [↗](#)

Can welfare be measured with a preference-satisfaction index?

Willem van der Deijl 

Faculty of Philosophy, Erasmus University Rotterdam, Rotterdam, The Netherlands

ABSTRACT

Welfare in economics is generally conceived of in terms of the satisfaction of preferences, but a general, comparable index measure of welfare is generally not taken to be possible. In recent years, in response to the usage of measures of subjective well-being as indices of welfare in economics, a number of economists have started to develop measures of welfare based on preference-satisfaction. In order to evaluate the success of such measures, I formulate criteria of policy-relevance and theoretical success in the context of preference-satisfaction measures of welfare. I present a detailed case study of the methodological choices put forward in a prominent generalized proposal for measuring welfare through preferences recently published in the *American Economic Review*. I contrast this with an alternative welfare measure which also uses preferences to weight aspects of welfare: the ICECAP-A measure. I assess the methodology of both approaches in detail and argue that the two goals of a preference measure of welfare can only be satisfied at the expense of making a measure prohibitively costly.

ARTICLE HISTORY

Received 30 March 2017
Accepted 2 December 2017

KEYWORDS

Welfare; utility;
measurement; happiness;
well-being

1. Introduction

Since the second half of the twentieth-century welfare economics has been built upon the foundational assumption that welfare should be conceived of in terms of the satisfaction of preferences and that it is highly doubtful that preferences can result in a cardinal measures of welfare, or in measures that are interpersonally comparable (Binmore, 2009b; Colander, 2007). Lionel Robbins' plea for expelling measures of psychological feelings of satisfaction out of the science of economics – for the reason that it failed to be scientifically measurable – is said to be pivotal in the formation of this view:

There is no means of testing the magnitude of A's satisfaction as compared with B's. If we tested the state of their blood-streams, that would be a test of blood, not satisfaction. Introspection does not enable A to discover what is going on in B's mind, nor B to discover what is going on in A's. (Robbins, 1932, p. 140 [original emphasis])

CONTACT Willem van der Deijl  wjavanderdeijl@gmail.com
Centre de Recherche en Ethique, Université de Montréal, Montreal, Canada

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

In response, it became an accepted view that rather than identifying welfare with pleasure, welfare should be conceptualized in terms of whether people get what they want. However, this concept of satisfaction of preferences is difficult to measure in a general fashion – capturing all that people care about, or at least all preferences that are affected by policy. While partial measures of welfare are sometimes used in cost-benefit analyses – such as QALY's in the context of health-related utility – no overall, or generic, measures are generally taken to allow for meaningful comparisons. Consequently, the consensus appears to be that while preferences may be used for partial comparisons of welfare, there can be no general preference-satisfaction measure of welfare that allows for interesting comparisons (see the Frey and Stutzer quote below).

However, in recent years, the attitude of economists towards the measurement of welfare has been shaken. A number of economists have incorporated insights from psychology and have taken psychological feelings of happiness and life-satisfaction to be measurable after all through self-reports – under the heading of subjective well-being (SWB henceforth; Clark, Frijters, & Shields, 2008; Di Tella & MacCulloch, 2006; Dolan, Peasgood, & White, 2008; Kahneman & Krueger, 2006 and MacKerron, 2012 for an alternative approach to SWB measurement of welfare). As Bruno Frey and Alois Stutzer write:

Only a few years ago, most economists took it for granted that utility cannot possibly be measured, and not even reasonably be approximated. Things have dramatically changed: happiness research has made great progress, with economists playing a leading role. (2016, p. 21)

The view that these 'happiness economists' endorse either takes welfare to be constituted by happiness or life-satisfaction, or takes such measures to be good approximations of preference-satisfaction. Happiness economics has been growing expansively, but has also encountered much criticism (e.g. Barrotta, 2008; Fleurbaey & Blanchet, 2013; Hausman, 2010; Stewart, 2014; Sugden, 2008).¹

This narrative so far is based on pragmatic reasons: Happiness made place for preference-satisfactionism – or preferentism – as a central conception of welfare in economics because happiness was not considered measurable, while preference-satisfaction could be operationalized to some extent; only to shift back when happiness became considered measurable again. There is, however, also a deep philosophical disagreement on the nature of welfare between these two approaches. Welfare – or well-being (I will use these terms interchangeably) – is a term describing that which makes life good for the person who is living it (Sumner, 1996; Tiberius, 2006). While both happiness-based theories of well-being (such as hedonism) and preference-satisfaction theories have used 'utility' as a term to describe the good, they are quite different. Happiness-based theories locate well-being in psychological states, be it in terms of good feelings – in case of hedonism – or life-satisfaction – for life-satisfaction theories. Preference-satisfaction theories are, different from hedonism, not mental-state accounts of welfare. While preferences are mental-states, the satisfaction of preference is not (Griffin, 1986). Welfare, on a preference-satisfaction account represents the extent to which the world corresponds with how one wants it to be. On this account, regardless of how strongly it influences a person's sense of happiness or satisfaction, whenever a preference is satisfied, welfare is, *ceteris paribus*, increased. Much of the disagreement between the happiness-based and preference-satisfaction conceptions of well-being has focused on this contention (cf. Heathwood, 2006).

This paper is concerned with one particular reply to the trend to measure welfare in terms of happiness, which has come from economists who endorse the view that welfare should be conceived of in terms of preferences rather than in terms of happiness or life-satisfaction,

and do not believe that the latter are necessarily good proxies of the former. Namely, the response that rejects happiness measures because they do not (necessarily) cohere with people's preference-satisfaction, but that, despite skepticism on theoretical grounds (Hausman, 1995; List, 2003; see Binmore, 2009a), welfare can nevertheless be measured in practice by means of eliciting people's preferences, and measuring the extent of their satisfaction. The most prominent example of such criticism comes from Benjamin, Heffetz, Kimball, and Szembrot (2014) who argue that: 'widely-used SWB measures may not capture all factors that enter into preferences' (Benjamin et al., 2014, p. 2700). In reaction, they are revising the age-old skepticism about the possibility to measure preference-satisfaction meaningfully on an individual, general level. Benjamin et al. formulate a general formal framework and related measurement methodology for measuring a preference-based well-being index, novel in its aim to measure welfare comprehensively.

The purpose of this article is to investigate the success of this response; not, however, in terms of the plausibility of the underlying theory of well-being, but in terms of the success of the claim that a preference-satisfactionist account of welfare can feasibly be developed into an individual index measure of well-being. There are a range of reasons for why preference-satisfaction may or may not be a plausible candidate for a measure of well-being – particularly in the context of policy. One could hold against Benjamin et al. that preferences may be other-regarding, immoral (Hausman, 2012, ch. 7; Heathwood, 2010),² and the satisfaction of preferences need not feel good, or may go completely unnoticed (Parfit, 1984), and as such, are implausible to serve as a basis of an account of welfare, particularly in the realm of policy. On the other hand, it may be defended by suggesting that even if preference-satisfactionism is not a plausible theory of well-being in light of these problems, preference-satisfaction may still be seen as good evidence for welfare (Hausman, 2012, ch. 8). Moreover, there is space to bite the bullet on these concerns (Lukas, 2009), and the subjectivity of the measures may be seen as an advantage in the policy context as it limits the concern that the measure is based on a paternalistic notion of welfare (Haybron & Tiberius, 2015). However, putting aside conceptual problems with preferentism, and regardless of some deeper theoretical concerns about the possibility of developing a preference-based utility index, the main contribution of this article is to argue that an individual well-being index based on preferences would be so data-demanding given its basic theoretical commitments, that it would be practically infeasible as a useful tool for policy. These theoretical commitments involve individualism about preferences – the preferences used to assess my well-being should be *my* preferences – and unrestrictedness with respect to preferences – the set of preferences that count towards my well-being should not exclude preferences that I may plausibly have, and that would count towards my well-being.

To make this argument, I assess the empirical strategies of Benjamin et al.'s approach in order to see how they deal with measurement challenges. I contrast their approach with an alternative way to use preferences in a general welfare measure. The ICEpop CAPability measure for Adults – or ICECAP-A (Al-Janabi, Flynn, & Coast, 2012; Al-Janabi et al., 2013; Flynn et al., 2015) – is a 'generic measure of well-being that can be used across health and other areas of public provision of goods and services' (Flynn et al., 2015, p. 267). Different from Benjamin et al.'s methodology, the ICECAP-A uses the preference-approach to welfare pragmatically – as a proxy for people's valuation – and is neither committed to individualism nor unrestrictedness. However, it nevertheless faces methodological challenges that shed light on the data-demandingness of certain methodological choices in measuring preferences.

The discussion of the used empirical strategies and choices in both approaches illustrates the methodological challenges that a preference-measure is faced with. Not only does this discussion clarify the issues at hand in a concrete context, but also takes seriously the context of the development of concrete measures, in which theoretical problems may be tackled by pragmatic choices. Importantly, however, the main aim of the article is not to criticize the pragmatic choices that researchers make to operationalize preference-satisfactionism. Rather, it is to illustrate the trade-offs involved in staying faithful to the theoretical commitments of preferentialism in the concrete context of developing welfare measures.

Section 2 discusses the very idea of preference-measures of welfare, and the meaning of policy-relevance and theoretical commitments in this context, Section 3 discusses the two case studies in detail, and Section 4 draws more general conclusions from this discussion with respect to the feasibility of measuring welfare through preferences. Section 5 concludes.

2. Measuring preference-based welfare, the very idea

The idea that welfare is constituted by preference-satisfaction comes with a number of commitments. Preference-satisfaction theory is sometimes called a formal theory of welfare (Tiberius, 2004), as it does not substantively take position on any specific good, but rather leaves individuals to be the author of what is good for them. An important underlying motivation for the formal nature of preference-satisfaction accounts is a strong anti-paternalistic intuition that the individual should be the ultimate judge on what makes her life good. A well-known formulation of this idea by Peter Railton is that: 'it would be an intolerably alienated conception of someone's good to imagine that it may fail in any way to engage him' (Railton, 1986; see also Fletcher, 2013; Yelle, 2014). By making well-being solely dependent on what we want, it is impossible that we judge someone to be well on the basis of things that they do not care about. This liberal conception of well-being is the core theoretical commitment of the approach.

On the concrete level of measurement, this implies two things. Firstly, the measure should be individualistic. That is, a preference-satisfaction measure of an individual's welfare should be based on her preferences, and not on the preferences or values of the group she is a part of, arm-chair philosophers, or policy-makers. Secondly, the space of things people may have preferences over, should not be restricted. If it is the case that people have strong preferences over the success of their football team, such preferences should have a place in a general measure of welfare. More precisely, while some preference-based conceptions of welfare may exclude preferences on formal grounds (such as unstable desires, Chekola, 2007), it could not do so on substantive grounds (such as seeming silly, like Rawls, 1971's famous grass-counter's desires).³ These two concrete commitments that follow from the liberal commitments of preference-satisfactionism we can call *individualism* and *unrestrictedness*, respectively.

What does it mean to measure welfare through preferences? The kind of things people have preferences over can be represented as a finite set of dimensions of welfare. To make this precise, consider the following general formalization of welfare:

$$W_i \equiv f_i(\mathbf{w}_i) \quad (1)$$

where W_i is an individual welfare function for individual i , \mathbf{w}_i a vector of elements that constitute welfare for i , and f_i a function that describes how these combine into a welfare value.

For example, hedonism can be characterized in this way, where $f_i = f$ is a linear function, while $w_i = w$ is a vector of one element, namely pleasure,⁴ while in case of objective list theories $w_i = w$, contains a number of elements that contribute to people's well-being in some way, $f_i = f$.⁵ Preference-satisfaction can be represented in two equivalent ways. In a limited form, it can be expressed as a version of (1) in which w_i has one element, s_i ; the extent to which one's preferences are satisfied, where $f_i = f$ is a linear unindexed function. Equivalently, we can also say that $f_i(w_i)$ is a completely individualized version in which both the contents of w_i as well as the way they are combined are fully determined by the structure of a person's preferences:

$$W_{pi} \equiv fs_i = f_i(w_i) \tag{2}$$

A measure of welfare is thus one that represents W_{pi} somehow. Such measures can differ in various ways, allowing for different types of comparisons. Most significantly, a measure of W_{pi} may be used to make comparisons between individuals, or within individuals over time. Because empirical work on the measurement of welfare aims to guide policy, it is useful to look at possible aims of well-being policy. Assuming that a policy-maker is interested in the well-being of affected citizen when comparing two alternative policy actions to the status quo, a policy-maker would be interested in total welfare effects, but also in where in the distribution the changes in welfare lie.

In the ideal case, measures of preference-satisfaction – henceforth, simply utility – are cardinally interpretable and interpersonally interpretable (see Table 1). If this is the case, a policy-maker can evaluate the degree and equality of welfare in the current state, and the effect of policies on both. In a less ideal case, utility can be compared across individuals, but only provides ordinal information. In this case, the worst-off in society can be identified, but the magnitudes of changes for different individuals cannot be compared. However, it may also be the case that a measure is cardinal, but not interpersonally comparable. In these cases, its policy relevance depends on whether the *differences* between utility scores – utility units – can be meaningfully compared. If units of utility can be made compared, but levels cannot, aggregated changes in welfare can be estimated, even though the distribution of welfare cannot be identified. This thus allows for utilitarian considerations in policy-making. If units of utility cannot be made comparable between individuals, it may still be the case that for every individual the units represent the same changes in utility – i.e. each individual has her own cardinal scale. For policy purposes, this is the worst situation, together with the case that the measures are neither cardinal, nor interpersonally comparable. In both these cases, welfare indices would be able to indicate whether the welfare of individuals has increased or decreased, but not how welfare of different individuals compares, nor how the aggregate magnitude of such welfare changes compares to other welfare changes.

Table 1. Level and comparability of preference-satisfaction measures.

	Cardinal	Ordinal
Interpersonal comparability	<i>Ideal:</i> both the magnitude of welfare changes and their impact on inequality can be assessed.	Relative positions, such as the worst-off, can be identified, but nothing can be said about the magnitude of interpersonal or intrapersonal utility differences.
No interpersonal comparability	<i>If differences comparable:</i> utilitarian considerations <i>If differences not comparable:</i> Individual intrapersonal judgments	Ordinal individual intrapersonal judgments

The relationship between the comparability levels of measures and policy evaluation is best illustrated with an example. Consider a policy-maker who has the option between keeping the status quo or doing policy A that affects Erik and Sophie. There is a measure of utility that is 6 for Erik and 8 for Sophie. We know that doing A would increase Erik's utility by 1, while it would decrease Sophie's utility by .5. In case the measure can be interpreted cardinally and is interpersonally comparable, we can see that the situation brings Erik and Sophie closer together, while improving aggregate welfare. In case they are ordinal and interpersonally comparable, we can see that Erik is worst-off, but we have no idea how Sophie's decrease compares in magnitude to Erik's. In case the utility measure is cardinal but levels of welfare are incomparable (and changes are comparable), we can see that A improves overall welfare, though we do not know if it equalizes the distribution. In the worst case – an ordinal measure that is not interpersonally comparable (or a cardinal measure of which neither the differences nor levels are comparable) – we know A improves Erik's welfare and decreases Sophie's, but we cannot say how this changes the distribution of welfare, nor whether Erik's increase is larger or smaller than Sophie's decrease. The *cardinality* and *comparability* thus jointly make up the policy-relevance of utility measures of welfare.

A final consideration is *feasibility* of measures. Any measure of welfare, but particularly ones that are to be used to inform policy, should not be overly demanding on either respondents or government offices conducting them. In particular, in the discussion that follows, the term feasibility will be used to distinguish between welfare measures that use data that can be obtained through reasonably long questionnaires (or other informational bases) and measures that require so much information from individuals that measurement would be prohibitively costly; in particular for the individual being evaluated.

3. In practice

Benjamin et al. (2014) motivate their article by describing 'the principle of revealed preference' as the cornerstone of economics: 'the ultimate criterion for judging what makes a person better off is what she chooses' (2014, p. 2698). However, while in some economic instances this principle may be informative, in particular in the policy context, people do not always make choices about options that may matter for their well-being. The aim of their paper is to provide an index of preferences to serve as an indicator of well-being in particularly these contexts. More precisely, they want to develop: 'an individual-level index that combines together different aspects of well-being that may be measured by survey questions' (2014, p. 2699)

They acknowledge and appreciate SWB measures as a candidate for this purpose. However, while such measures may be multi-dimensional, the weights attached to these dimensions are generally assigned by researchers themselves, and are thus 'ad hoc' (2700). Consequently, a person can score high on such indices without this reflecting this person's preference.

The article provides a theoretical framework as well as an empirical illustration of this framework, which results from a number of pragmatic choices. Theoretically, the proposal is based on the consumption theoretic framework in which changes in utility are assumed to be proportional to changes in consumed goods:

$$\Delta u \propto \sum_{m=1}^M p_m \Delta c_m \quad (3)$$

where M is the set of goods, m , consumed at price p_m , and quantity c_m . While this framework may be sufficient for assessing the impact of market transactions on welfare, the same does not apply to non-market goods, with which policy typically is engaged with. Thus, Benjamin et al. propose to broaden the framework accordingly. Rather than just market goods, they propose a welfare function exists out of a set \mathbf{w} of welfare components, w_j over which people have preferences. Formally, this is captured by the following:

$$\Delta u = \sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} \Delta w_j \quad (4)$$

Intuitively, this captures the idea that over a given time period, the changes in welfare are given by all the changes in welfare components multiplied by their impact on welfare, which corresponds to how strongly they are preferred. So, if only health improves, while all the other welfare components remain the same, welfare changes just as much as how strongly the health improvement was preferred over alternatives.

Interestingly, the theoretical framework does not stop here, but in fact, the welfare index that is proposed is defined as follows⁶:

$$W_{bhks,i} \sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} w_j \quad (5)$$

Compared to (4), the Δ is removed from before w_j . In practice then, an empirical strategy to get at the index is to measure the $\partial u(\mathbf{w})$ from a stated preference survey, while the w_j 's should be obtained independently from, among other sources, SWB surveys. In the stated preference survey, respondents are asked to report what their preferences are between two options in which 1–3 aspects are altered positively, and 1–3 aspects are altered negatively, and they are asked whether they prefer either one of the options slightly, somewhat, or much.

Benjamin et al. (2014) acknowledge that 'since the marginal utilities are defined only up to an arbitrary constant, so is the index' (2014, pp. 2704–2705). As the index weighs the *level* of welfare components with their *marginal* relative utility, it can only sensibly compare changes in welfare within an individual over time. Even in the theoretical ideal that a full preference map can be made for a particular individual that maps the desirability of all the possible values of \mathbf{w} , only 'ordinal welfare comparisons could then be made between any of the individual's SWB-survey occasions' (2014, p. 2705). This indicates that even in the theoretical ideal, the index falls within the lower right corner of Table 1, with a highly limited policy-applicability.

However, in their proposed (and executed) empirical strategy, a number of pragmatic choices need to be made that complicate the validity of even such minimal intrapersonal ordinal comparisons.

3.1. Selection of aspects

The methodology Benjamin et al. propose to elicit preferences is a stated preference method. As a central part of the method, respondents are asked to make hypothetical choices between aspects in \mathbf{w} . This requires that that the elements of \mathbf{w} are specified. This creates a challenge for Benjamin et al. After all, in order for the preference measure to be valid, it requires the

set of aspects of welfare to be ‘*exhaustive*’, as well as ‘*non-overlapping*’ (2014, p. 2707). Nevertheless, because they would not want their ‘*ex-ante beliefs*’ to be an influence, and because they would not want to miss out on any important aspects, they decide to create ‘*as comprehensive a list of candidate fundamental aspects as we practically can*’ (2014, p. 2707). The constructed list is made up of philosophical lists (such as Nussbaum, 2000) and aspects of well-being from both the empirical and philosophical literature, such as SWB constructs, as well as some aspects the authors themselves contributed. Benjamin et al. acknowledge that this may lead to overlapping aspects on the list.

The resulting list contains 129 aspects, including, for example, ‘The extent to which humanity does things worthy of pride’, ‘The amount of pleasure in your life’, ‘Equality of income in your nation’, ‘You not feeling anxious’, and ‘People getting the rewards and punishments they deserve’ (2715–2718). They acknowledge that many of these may overlap. The amount of pleasure in one’s life and ‘you not feeling anxious’ is one such example. In order to abridge this list into non-overlapping aspects of welfare, Benjamin et al. propose a data-driven strategy (2014 appendix). They argue that when a combination of two aspects is considered less desirable to an individual than the sum of the two separate aspects, it implies an overlap, and one of the two aspects may have to be deleted from the list.

This methodological choice has an important theoretical attraction: it minimizes paternalism with respect to the contents of w_i , and thereby maximizes unrestrictedness.⁷ However, it comes with a number of problems.⁸ The overlap between many of the concepts involved may very well be detected by the method proposed, but not clearly solved. Consider for example, the overlap between health and pleasure. Feeling pleasure, up to an extent, must surely be seen as part of being healthy, while at the same time being in bad health will affect how much pleasure one feels. The proposed method will thus quite likely find overlap between pleasure and health, but this certainly does not mean that these are fully separate aspects and that one of them should be dropped. Quite likely many items on the list will overlap in a similar fashion. A potential solution to avoid overlap is to rely on theoretically motivated lists, such as objective lists in philosophy (e.g. Griffin, 1986; see also Alkire, 2002). However, Benjamin et al. rightly argue that relying on such *ex ante* lists would be hostile to the unrestrictedness commitment of preference-satisfactionism.

3.2. Data-demandingness and pooling of preferences

Another problem arises from the large number of elements in w , namely, it now becomes effortful to map a person’s preferences. This would not only require an individual to rank 129 aspects in life, but because no particular functional structure can be assumed a priori, the large possible set of all possible values of the elements in w_i may take have to be compared *vis-à-vis* each other as whole. This means that the number of required comparisons to construct a full preference map increases exponentially. The intuition behind this is that when one aspect, such as health, suddenly deteriorates, this may not only affect the relative importance of health improvements in comparison to other aspects, but it may also affect the preference someone has between other aspects, such as between *lack of anxiety* and *your sense of achievement*. Such a non-linear relationship between preferences over these aspects vastly increases the possible preference-comparisons required to construct a full preference-map. This leads Benjamin et al. to a number of pragmatically motivated choices.

Firstly, preferences are elicited only at the present level of values of w . This allows Benjamin et al. to limit themselves to asking respondents to make hypothetical trade-offs at the present level of w . This means that they need not ask respondents how they would make the trade-off between anxiety and sense of achievement *given that their health level is X*, where X may vary from very close to very far from their actual level of health. Secondly, in order to overcome the data-demandingness of the proposed theoretical framework, Benjamin et al. assume that preferences are locally linear, thus assuming away the non-linear effect just described. These assumptions, however, are highly restrictive. It excludes preferences that people plausibly may have. Nevertheless, even with these assumptions, deriving individual marginal utilities requires individuals to make large sets of comparisons in their proposed stated preference surveys.

The way Benjamin et al. overcome this difficulty is by pooling preferences across respondents. This means that the responses from the hypothetical trade-offs are pooled together as if they all came from a single person. This heavily reduces the data-demandingness of the method, but such a method is equivalent to assuming a representative agent. As Benjamin et al. acknowledge, '[d]oing so is difficult to justify theoretically' (2014, p. 2731). After all, in light of the individualistic commitment of the preference-satisfactionist approach, it is peculiar to assume that all individuals have (roughly) the same preferences. Their proposed method to counter this concern is to identify different groups, or 'types' according to their demographic characteristics.

The general problem is also related to the large selection of welfare aspects. Because the number of welfare aspects Benjamin et al. consider is so large, constructing preference maps is very data-demanding, which in turn requires simplifying assumptions that are difficult to justify in light of the individualist commitment of the approach. The weights used to aggregate the different welfare aspects in w may very well be substantially different from the weights that an individual would want. To see this, consider someone whose preferences deviate from typical responses from her type. For example, while everyone of her type cares a lot about their health and the well-being of their families, one person cares much more about a sense of achievement. Her aggregate welfare, W_i , will then be weighed by weights that do not at all resonate with her own preferences. Notably, even without attempting to develop a cardinal or interpersonally comparable measure of welfare, Benjamin et al. run into steep trade-offs between feasibility and theoretical commitments.

3.3. An alternative: ICECAP-A

An alternative preference-based specification of a welfare measure is provided by the ICECAP-A. Preference-based indicators have been widely used to evaluate outcomes in cost-benefit analyses, in particular in the field of health care. In health care health-related quality of life is generally evaluated with a well-known preference-based measure, the QALY. A QALY captures the utility of being in a particular health state.⁹ Health states can be defined in a number of dimensions. For example, a prominent measure (EQ-5D) uses: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each of these is scaled from 1 to 5, such that 13551 means that a health state comprises little mobility and much depression, medium self-care, but no pain and full ability to do usual activities. Nevertheless, the distinction between health-related quality of life and quality of life itself is not always clear, as the impact of decreased health states may affect a broad spectrum of aspects of our lives

(Al-Janabi et al., 2012). In order to overcome such problems, a number of health economists have proposed measures of general welfare. One such measure is the ICECAP-A. The ICECAP-A is based on a capability framework, and as such rejects the use of individualistic preferences as a direct measure of welfare.¹⁰ Nevertheless, in order to operationalize the approach, Flynn et al. use what they call the “Cookson’s compromise” in assuming that population values obtained from choice-based tasks can be used as evidence for valuation with the capability approach (Cookson, 2005)’ (Flynn et al., 2015, pp. 259–260). Effectively, like Benjamin et al., it uses preferences to assign weights to different welfare components.¹¹ Like Benjamin et al.’s empirical proposal, it pools the noted preferences (Flynn et al., 2015). Both assume the framework in (2), and use stated hypothetical preferences to estimate $f(\cdot)$. However, there are a number of significant differences in the way the difficulties described above are addressed.

First, rather than using a list that is as comprehensive as possible, the capability approach that the developers of ICECAP-A endorse allows them to limit their welfare goods to a carefully constructed concise list of five capabilities of an approximately equally high level of abstraction: stability, attachment, autonomy, achievement, and enjoyment. This list is not (only) informed by the academic literature, but extracted from structured interviews and focus groups (Al-Janabi et al., 2012). Secondly, Flynn et al. (2015) assume that their welfare measure can be calibrated between 0 (having no capabilities whatsoever) and 1 (having full capability). These two assumptions are arguably quite strong. It is restrictive to limit the set of w over which agents have preferences relevant for well-being to five capabilities. Nevertheless, if well-being were five-dimensional, it is not implausible to say that perfect well-being is reached if all these five dimensions are at their maximal levels. Al-Janabi et al. use a discrete choice-based valualational experiment, using best-worst scaling, in order to elicit preferences.¹² This method, in combination with the two assumptions above, allow for a cardinal measure of welfare (or, the value of welfare).¹³ The measure ultimately presents a population level measure scaled from 0 to 1, in which each of the five capabilities affects well-being in an additive, but non-linear way. In other words, the difference between levels 1 and 2 enjoyment, and levels 3 and 4 enjoyment is weighted differently, but these weights do not depend on the level of the other four capabilities. Would they have done so, then, this would greatly increase the data-demand. As Flynn et al. (2015) explain: ‘a DCE [discrete choice experiment] capable of estimating even only two-way interactions would have been prohibitively costly, in terms of sample size’.

By filling out a questionnaire asking a respondent her five levels of capability, researchers can assign a welfare-value to the respondent on the basis of the weights from a population-wide discrete-choice experiment. The cardinal and interpersonally comparable nature of the measure of welfare fits in the top left concern of Table 1, and as such has much policy-relevance. Not only does it allow identifying the worst-off in society in terms of this measure, but it also helps to identify the impact of different interventions, and it can help assess how welfare has developed over time. Moreover, if the estimated weights can be extrapolated to other contexts,¹⁴ it does not require much data to estimate. However, while the measure only uses preferences as an estimate of individual valuation, as an individualistic measure of welfare, the policy-relevance comes at a high cost from a preference-satisfactionist perspective. The methodology takes the population seriously in the determination of the welfare function, both in terms of its contents – the formulation of the five capabilities that make up w – as well as its weights. However, because population values are used, it fails

to meet the individualistic commitment inherent to preference-satisfactionism as an account of welfare. If someone has a different view than the rest of the population in terms of the weights and contents of the well-being measure, the welfare function may be alien to her, and it may be possible that a person is evaluated on a standard, without her endorsing the values on which this is based.

4. Steep trade-offs

We have identified a number of different dimensions on which preference-based welfare measure may differ:

- (Q1) Does it restrict the list over which one can have preferences?
- (Q2) Does it restrict the functional form of the preference function?
- (Q3) Does it use individual preferences or does it use preferences that are pooled across individuals to weigh welfare aspects?
- (Q4) Does it result in an ordinal or cardinal measure?
- (Q5) Does it result in an interpersonally comparable or incomparable measure?
- (Q6) Does it require a lot of data from individuals who are being assessed?

In the discussion above, we saw that Benjamin et al.'s measure and the ICECAP-A make the following choices (Table 2):

While Benjamin et al. only minimally restrict preferences and aim at a measure that is ordinal and incomparable between individuals, and is relatively data-demanding, the ICECAP-A measure restricts the objects over which individuals have preferences to five capabilities, is cardinal and interpersonally comparable, and is not very data-demanding. However, both use group-level preferences, even though both submit that their methods can be used for substrata of the population, rather than the population as a whole, to get closer to the individual level.

The first three considerations on the list relate to the theoretical commitments of the approach. While neither approach truly provides a welfare measure that respects individual preference, Benjamin et al. do go at great lengths to leave the aspects over which individuals can have preferences as open as possible, while the ICECAP-A only captures preferences over five abstract capabilities. While the ICECAP-A allows for a non-linear relationship between the evaluated welfare goods and utility, neither approach is able to accommodate non-linear interactions between welfare goods.

Questions 4 and 5 determine how useful the measure ultimately is for policy-makers. While the ICECAP-A is (plausibly) taken by its developers to represent a cardinal and

Table 2. Methodological aspects of the two preference-satisfaction measures.

	Benjamin et al. (2014)	ICEPCAP-A
Unrestricted preferences?	Yes (or minimally restricted)	No
Restricted function form?	Linearity restriction and no interactions	No interaction, but no linearity restriction
Individual preferences or preferences pooled across individuals?	Pooled	Pooled
Ordinal or cardinal?	Ordinal	Cardinal
Interpersonally Comparable?	Incomparable	Comparable
Data-demanding?	Relatively data-demanding	Not very data-demanding

interpersonally comparable measure of welfare, Benjamin et al.'s measure only captures ordinal changes in welfare that are not interpersonally comparable. While Benjamin et al. explicitly have a policy-context in mind when they consider the application of their measure, it is unclear how their measure can be helpful in guiding welfare-driven policies. Because of the limited comparability, it is neither able to make utilitarian judgments about the changes in aggregate welfare, nor is it able to identify relevant differences in welfare between individuals.

The final question is about feasibility. As Benjamin et al. note, even under restrictive assumptions, fully comparing the different possible combinations of 129 welfare aspects leads to such a large number of comparisons that this would be completely infeasible to do for each individual, and, despite the much lower number of welfare aspects, the same applies the ICECAP-A. But, in light of the pragmatic choice to group preferences, the ICECAP-A strongly limits the data-demandingness; and so does the Benjamin et al. measure to a more limited extent.

The problem is thus as follows.¹⁵ A preference-satisfaction measure of welfare that remains faithful to its theoretical commitments does not limit the preference space in such a way that significant kinds of preferences that people may hold are not represented (Q1). Moreover, it evaluates a person's welfare by her own preferences, and not those of others (Q3). In order to allow for the preferences people may have, this would also involve allowing for non-linear relationships between welfare goods and welfare, as well as non-linear interactions between welfare goods (Q2). In order for a measure to allow for making interesting comparisons, it should be interpersonally comparable (Q5) or cardinal (Q4), and ideally both. However, achieving each of these things increases the data-demandingness of measures. In particular, combining (Q1) with (Q3) is highly data-demanding, as there are simply a lot of dimensions on which lives may be different that can be relevant for welfare, and this means that there are a lot of different possible combinations of such dimensions that make up welfare. For every dimension added to the set of w_i , the number of possible welfare states increases exponentially. The ICECAP-A is based on five dimensions that may vary over four different levels. This leads to 1024 (4^5) possible values of w_i that should be ranked in order to get at a full preference-map. In case of Benjamin et al., if we assume that each of the 129 dimensions can vary over four different levels, we get at a total of $4.63 * 10^{77}$ possible states of the world to compare. However, even if the individualistic commitment is dropped (as is done by both Benjamin et al. and the ICECAP-A), there are still significant data constraints that prohibit fulfilling all the other features. As we saw, adding the possibility of non-linear relationships and information needed to make measures comparable and cardinal only increases the informational load needed to come to an individual level measure of welfare.

It is highly contentious whether a cardinal interpersonally comparable welfare measure could exist in practice (e.g. Binmore, 2009b; Hausman, 1995; List, 2003), but if so, a calibrated scale would be needed to standardize the preference ranking (as the ICECAP-A does). However, this requires that we do not only evaluate changes of welfare at the present level of w , but that we construct a full preference-map that will help us indicate how far away we are from calibrated points (e.g. 0 and 1). This again, raises the number of hypothetical choices a particular individual would have to make even further. Staying faithful to theoretical commitments and providing a useful policy measure are thus heavily data-demanding features of a welfare measure based on preference-satisfactionism.

This problem is explicitly acknowledged by both groups of authors when it comes to the question whether a measure should use individual or population-level preferences. Benjamin et al. motivate their choice to use group-preferences rather than individual level preferences explicitly by referring to avoiding data-overdemandingness, and Flynn et al. (2015) also note that even with five welfare dimensions 'it was not feasible to provide respondents with all 1024 possible scenarios'. In both cases the measure can be individualized, but only at the cost of being over-demanding to individuals who are being assessed.¹⁶

A measure that requires individuals to make thousands of hypothetical comparisons in order to get at an individual measure of welfare is infeasible; in order to reduce this number, theoretical commitments of preference-satisfactionism, policy-relevance, or some aspects of both need to be sacrificed. Hence, as feasibility is the hardest constraint for researchers wanting to provide a helpful measure for policy purposes, a trade-off needs to be made between the different aspects of theoretical faithfulness and policy-relevance. Without suggesting a correct way to make such trade-offs, it should be clear that such trade-offs are steep. Neither of the two measures ultimately stays faithful to the individualism of preference-satisfactionism. Furthermore, Benjamin et al.'s measure falls short of providing a policy-relevant measure, while the ICECAP-A heavily restrict the space of valuation.

5. Conclusion

For theoretical reasons, the possibility that welfare could be measured by means of a comparable index of general preference-satisfaction has for a long time been seen with a skeptical attitude. Such attitude may have been overly skeptical. The theoretical frameworks of both reviewed approaches show that at a concrete level plausible pragmatic choices can be made that result in indices of utility. However, in light of methodological considerations, a preference measure of welfare falls short either in terms of normative and theoretical commitments, in terms of policy-relevance, or faces the charge that it is too demanding to be put into practice.

So, to what extent can welfare be measured with a preference-based utility index? While a preference-based method of indexing welfare connects closely to economic theory, the trade-offs identified in this article pose a stark challenge to this possibility. However, the trade-offs also offer an opportunity to bring different approach to welfare measurement closer together. For example, by restricting the space of possible preferences in a welfare-measure, the developers of the ICECAP have incorporated insights from the more objective capability approach. While this may make their approach less attractive to a pure preference-satisfactionist, it should make the approach more appealing to capability scholars. If indeed, a preference-measure needs to let go of its unrestrictedness, the usage of insights from alternative approaches – such as the capability approach and objective list theories – may be a way to build a synthesis in the ongoing debate on how to measure welfare.

Notes

1. It similarly has been fiercely defended against the charge that happiness measures are not cardinal (Ferrer-i-Carbonell & Frijters, 2004), and that it does not capture what it is intended to capture (see for example Alexandrova & Haybron, 2016; Diener, Lucas, Schimmack, & Helliwell, 2009; van der Deijl, 2017; and Veenhoven, 2012 for a critical appraisal).

2. Particularly in the context of policy, it would arguably be problematic if governments would aim to satisfy preferences that are racist (Heathwood, 2010; Hausman, 2012, ch. 7). I would like to thank an anonymous referee for pointing this out.
3. A difficult question is whether a preferentist account of welfare should include other-regarding preferences or not, and whether this counts as a formal criterion. Both positions have arguments in favor. The argument does not rely on this question.
4. In case of qualitative hedonism, e contains a variety of versions of pleasure, in which case the relationship, between w_i and w_j is also more complex.
5. The capability approach, merely limits w_i to functionings and capabilities, without specifying f_i or w_i in any substantive sense (Sen, 1985).
6. To cohere with the rest of the formulations in this article the name of the index, $W_{bhks,i}$ is my own input, the rest comes directly from Benjamin et al. (2014, p. 2704).
7. The extent to which it minimizes paternalism is arguably still more restricted than necessary. It still excludes many possible aspects people may have preferences about, such as the status of the great barrier reef, that our children are successful in life, or having aesthetic experiences, just to name a few.
8. A main problem is mentioned, but a further issue is that this methodology is blind to the distinction between conceptual overlap and overlap due to causal relationships. To a person who cares about income, education and income may overlap in terms of preferences (due to expected income increase education may provide), but need not think of the two as similar concepts. A second issue is that Benjamin et al. assume that people's interpretation of concepts is fixed. This is dubious and, in the context of developing conceptual overlap, problematic. A person who is asked to make a trade-off between 'The happiness of your family' and 'The amount of pleasure in your life' will probably understand happiness to be something different than someone who is asked to make a trade-off between 'The happiness of your family' and 'The overall well-being of you and your family'. Moreover, the approach is highly data-demanding, adding to the concern discussed below.
9. One way in which this can be done is asking respondents to make time-trade-offs between health states (others are using standard gambles or visual analogue scales). Under the assumption that utility components of time and health states are separable, one may assess when a particular individual is indifferent between living for 10 years in a bad health state compared to living X years in a perfect health state. If the perfect health state is assigned the value 1, we know that $x/10$ must be the value of QALY of living for a year in the bad health state.
10. Both Al-Janabi et al. (2012), as well as Flynn et al. (2015) stress that they aim to get at a 'general', 'generic', or 'overall' (2015, p. 267) welfare measure that is not limited to health itself, but quality of life in general. However, as one anonymous referee pointed out, Al-Janabi et al. (2012) do exclude one item that people care about – 'the ability to live in a good or "just" world' (74) – on grounds of lying outside of the scope for policy, while running the risk of overshadowing other capabilities. While this may be interpreted as a pragmatic argument, it can also be interpreted as showing the authors are ultimately interested in a measure that is *so general as to capture all that affects wellbeing within policy*. This is thus not the exact same thing as general well-being itself. However, as this concept is still highly general, and as Benjamin et al's also maintain this policy focus, the conclusion below will not hinge on this distinction.
11. The remainder mainly considers the ICECAP-A measure as a preference-measure of welfare and identifies weaknesses of it as such. However, this is not the only way to understand this measure, and does not appear to be the way the developers themselves see the ICECAP-A. The authors instrumentally use preferences as a proxy of valuation. As the ICECAP-A is a capability measure, and uses population preferences a proxy of what individuals have 'reason to value' (Sen, 1999), many of the criticism may not apply. The ICECAP-A measure may very well be among the most defensible measures of welfare, both in terms of normative and empirical adequacy, if it is not seen as a preference-measure of welfare. The same arguments apply to a similar measure, the Adult Social Care Outcomes Toolkit (ASCOT; Forder & Caiels, 2011). However, while this measure takes on a more explicit utility-based framework, it aims to measure 'social outcome' or 'care-related quality of life' rather than well-being more generally.

12. The methodology is based on asking respondents to compare well-being states – combinations of different levels of the 5 dimensions of the ICECAP-A measure. The estimates used to scale each dimension are based on the conditional probability a well-being state is considered best (or worst) given the attribute level of the dimension.
13. The authors are not very clear on the exact interpretation of their measure. They ultimately present their weights as a ‘tariff’ to be used in economic valuations. It clearly presents the value of wellbeing states to each individual. It is not clear however, whether the authors believe there to be a distinction between the value of a person’s wellbeing state and a wellbeing state itself. While the two concepts appear to be the same in this context, the language the authors use avoids putting the two at par.
14. The authors themselves see the estimated weights as having value nation-wide in the UK, but also believe that particular contexts may require different weights.
15. As one anonymous referee pointed out, it may be argued that the problem may be held against any measure of welfare. However, while there are a variety of problems one can list with respect to the appropriateness of SWB measures of welfare, data-demandingness is not one of them. Because SWB assumes people are good evaluators of their own happiness, asking a single question is sufficient to estimate their welfare. This, thus neither requires the pooling of individuals, nor restricting specific types of happiness. Purely objective pluralist measures do face a difficult weighting problem: how should the different goods on a list be valued in different lives. This problem is similar to the problem discussed here, and objectivists about welfare generally do not formulate a clear answer to this question. However, the issue is different (perhaps worse) than the problem discussed here. Preferentists do have a theoretical solution: we need to know what people (would) want, in order to solve the weighing problem. As the argument shows, however, this requires so much information, that it is not realistically feasible.
16. Similarly, developers of the ASCOT write: ‘There are practical limits on the number of indicators any preference-based measure can utilize, mainly due to the limitations of preference elicitation techniques. They include the difficulty respondents have in ranking over many different attributes and the tractability of statistical analysis of these data’ (Forder & Caiels, 2011, p. 1768).

Acknowledgments

I would like to thank Werner Brouwer for working intensively with me on this topic. Furthermore, I would like to thank Jack Vromen, Conrad Heilmann, an anonymous referee, and audiences at ENPOSS 2017, and the 3rd International Conference Economic Philosophy for helpful comments.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Willem van der Deijl is a postdoctoral researcher at the Centre de Recherche en Ethique in Montreal. He holds a PhD from the Erasmus University Rotterdam in Philosophy and Economics. His research focuses on epistemic, axiological, and political issues regarding the concept of wellbeing.

ORCID

Willem van der Deijl  <http://orcid.org/0000-0002-8190-2705>

References

- Alexandrova, A., & Haybron, D. M. (2016). Is construct validation valid? *Philosophy of Science*, 83(5), 1098–1109.
- Al-Janabi, H., Flynn, T. N., & Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. *Quality of Life Research*, 21(1), 167–176.
- Al-Janabi, H., Peters, T. J., Brazier, J., Bryan, S., Flynn, T. N., Clemens, S., ... Coast, J. (2013). An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research*, 22(7), 1831–1840.
- Alkire, S. (2002). *Valuing freedoms: Sen's capability approach and poverty reduction*. Oxford: Oxford University Press.
- Barrotta, P. (2008). Why economists should be unhappy with the economics of happiness. *Economics and Philosophy*, 24(2), 145–165.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Szembrot, N. (2014). Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference. *American Economic Review*, 104(9), 2698–2735.
- Binmore, K. (2009a). Interpersonal comparison of utility. In H. Kincaid & D. Ross (Eds.), *Oxford Handbook of Philosophy of Economics* (pp. 540–559). New York, NY: Oxford University Press.
- Binmore, K. (2009b, March). Interpersonal comparison of utility.
- Chekola, M. (2007). Happiness, rationality, autonomy and the good life. *Journal of Happiness Studies*, 8(1), 51–78.
- Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1), 95–144.
- Colander, D. (2007). Retrospectives: Edgeworth's hedonimeter and the quest to measure utility. *Journal of Economic Perspectives*, 215–225.
- Cookson, R. (2005). QALYs and the capability approach. *Health Economics*, 14(8), 817–829.
- Di Tella, R., & MacCulloch, R. (2006). Some uses of happiness data in economics. *Journal of Economic Perspectives*, 20(1), 25–46.
- Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. F. (2009). *Well-being for public policy*. New York, NY: Oxford University Press.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29(1), 94–122.
- Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness?. *The Economic Journal*, 114(497), 641–659.
- Fletcher, G. (2013). A fresh start for the objective-list theory of well-being. *Utilitas*, 25(2), 206–220.
- Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. New York, NY: Oxford University Press.
- Flynn, T. N., Huynh, E., Peters, T. J., Al-Janabi, H., Clemens, S., Moody, A., & Coast, J. (2015). Scoring the icecap-a capability instrument. Estimation of a UK general population tariff. *Health Economics*, 24(3), 258–269.
- Forder, J. E., & Caiels, J. (2011). Measuring the outcomes of long-term care. *Social Science & Medicine*, 73(12), 1766–1774.
- Frey, B. S., & Stutzer, A. (2016). Policy consequences of happiness research. In Stefano Bartolini, Ennio Bilancini, Luigino Bruni, & Pier Luigi Porta (Eds.), *Policies for Happiness* (pp. 21–35). Oxford: Oxford University Press.
- Griffin, J. (1986). *Well-being: Its meaning, measurement, and moral importance*. Oxford: Clarendon.
- Hausman, D. M. (1995). The impossibility of interpersonal utility comparisons. *Mind*, 104(415), 473–490.
- Hausman, D. M. (2010). Hedonism and welfare economics. *Economics and Philosophy*, 26(3), 321–344.
- Hausman, D. M. (2012). *Preference, value, choice, and welfare*. New York, NY: Cambridge University Press.
- Haybron, D. M., & Tiberius, V. (2015). Well-being policy: What standard of well-being? *Journal of the American Philosophical Association*, 1(4), 712–733.
- Heathwood, C. (2006). Desire satisfactionism and hedonism. *Philosophical Studies*, 128(3), 539–563.

- Heathwood, C. (2010). Welfare. In J. Skorupski (Ed.), *The Routledge companion to ethics* (pp. 645–655). London: Routledge.
- Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, 20(1), 3–24.
- List, C. (2003). Are interpersonal comparisons of utility indeterminate? *Erkenntnis*, 58(2), 229–260.
- Lukas, M. (2009). Desire satisfactionism and the problem of irrelevant desires. *The Journal of Ethics and Social Philosophy*, 4, 1–24.
- MacKerron, G. (2012). Happiness economics from 35 000 feet. *Journal of Economic Surveys*, 26(4), 705–735.
- Nussbaum, M. C. (2000). *Women and human development: The capabilities approach* (Vol. 3). New York, NY: Cambridge University Press.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
- Railton, P. (1986). Facts and values. *Philosophical Topics*, 5–31.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Robbins, L. (1932). *An essay on the nature and significance of economic science*. New York, NY: MacMillan.
- Sen, A. K. (1985). *Commodities and capabilities*. Amsterdam: Oxford University Press.
- Sen, A. K. (1999). *Development as freedom*. Oxford: Oxford University Press.
- Stewart, F. (2014). Against happiness: A critical appraisal of the use of measures of happiness for evaluating progress in development. *Journal of Human Development and Capabilities*, 15(4), 293–307.
- Sugden, R. (2008). Capability, happiness and opportunity. In L. Bruni, F. Comim, & M. Pugno (Eds.), *Capabilities and Happiness* (pp. 299–322). New York, NY: Oxford University Press.
- Sumner, L. W. (1996). *Welfare, happiness, and ethics*. Oxford: Clarendon Press.
- Tiberius, V. (2004). Cultural differences and philosophical accounts of well-being. *Journal of Happiness Studies*, 5(3), 293–314.
- Tiberius, V. (2006). Well-being: Psychological research for philosophers. *Philosophy Compass*, 1(5), 493–505.
- van der Deijl, W. (2017). Which problem of adaptation? *Utilitas*, 29, 474–492.
- Veenhoven, R. (2012). Cross-national differences in happiness: Cultural measurement bias or effect of culture? *International Journal of Wellbeing*, 2, 333–353.
- Yelle, B. (2014). Alienation, deprivation, and the well-being of persons. *Utilitas*, 26(4), 367–384.