Universiteit Antwerpen
CSB | Centrum voor Sociaal Beleid Herman Deleeck

Matthew D. Adler and Koen Decancq

Well-Being Measurement

WORKING PAPER

No. 21/05

May 2021

Universiteit Antwerpen

University of Antwerp
Herman Deleeck Centre for Social Policy
https://www.uantwerpen.be/en/research-groups/csb/

Well-Being Measurement[1]

Matthew D. Adler[2] and Koen Decancq[3]

May 2021

Abstract: The social welfare function (SWF) framework includes a well-being measure $w(\cdot)$, for converting outcomes into vectors (lists) of well-being numbers. These well-being numbers are interpersonally comparable. This chapter discusses the construction of the well-being measure. It supposes that $w(\cdot)$ operates on individual "histories," a history being a combination of an attribute bundle a and a preference R. That is $w(\cdot) = w(a, R)$. This setup is quite general. It encompasses preference-based well-being measures (namely those that assign well-being numbers to histories containing different bundles but the same preference in deference to that preference), as well as non-preference based measures. The chapter covers both, although mainly focusing on the former. Here, two approaches are discussed: the "equivalence approach," whereby an individual's well-being hinges on her attributes and her ordinal preference; and the "vNM approach," which uses lottery preferences rather than ordinal preferences.

Key words: well-being measurement, preference, interpersonal comparisons, prioritarianism, equivalence approach, equivalent income, money-metric well-being, von Neumann-Morgenstern (vNM) utility, capabilities, subjective well-being (SWB).

[1] This text is forthcoming as Chapter 3 in "Prioritarianism in Practice", edited by Matthew Adler and Ole Norheim (Cambridge University Press). We thank Erwin Ooghe, Paolo Piacquadio and participants in conferences at the University of Bergen and Duke University for comments and discussion. Decancq gratefully acknowledges support of the Research Foundation - Flanders.

[2] Duke University, Law School.

[3] Herman Deleeck Centre for Social Policy (University of Antwerp); Department of Economics (KU Leuven); Centre for Philosophy of Natural and Social Science (London School of Economics) and CORE (UC Louvain).

1 Introduction

This chapter discusses the construction of a well-being measure for purposes of the social welfare function (SWF) framework. Together with chapter 2, it provides a theoretical foundation for the *Prioritarianism in Practice* volume.

As discussed in chapter 2, the SWF framework includes three major components: a well-being measure $w(\cdot)$, for converting outcomes into vectors (lists) of well-being numbers; the SWF itself, which is a rule for ranking well-being vectors; and an uncertainty module, for ranking policies understood as probability distributions over outcomes. A given outcome x corresponds to the well-being vector $(w_1(x), \ldots, w_N(x))$, with $w_i(x)$ denoting the well-being of individual i in outcome x.

Chapter 2 also discussed the question of well-being comparability. It explored four different types of well-being comparisons: intra- and interpersonal comparisons of well-being levels and differences. The chapter defended all four types of comparisons, both for reasons having to do with the informational requirements of utilitarian and prioritarian SWFs, and as a matter of common sense. The well-being measure should be such as to embody information about all four types of comparisons.[4]

But how indeed are we to construct a well-being measure that embodies these various kinds of well-being information? That is the topic of this chapter.

We start, in section 3.2, by introducing the notion of a preference-based well-being measure. Generically, an individual's well-being in an outcome depends upon her "attributes" (income, health, longevity, happiness, etc.) and may also depend upon her preferences. A preference-based well-being measure builds in such dependence on preferences, in a systematic way. The succeeding two sections focus on the two principal approaches to constructing an interpersonally comparable preference-based well-being measure: the equivalence approach (Section 3.3) and the vNM approach (Section 3.4).[5] Section 3.5 addresses two non-preference-

---

[4] A well-being measure $w(\cdot)$ embodies information about these comparisons as follows.

Intra- and interpersonal comparisons of well-being levels. $w_i(x) \geq w_j(y)$ iff (if and only if) individual i in outcome x is at least as well off as individual j in outcome y, with x and y identical or distinct. If $i = j$, this is an intrapersonal level comparison. If $i \neq j$, this is an interpersonal level comparison.

Intra- and interpersonal comparisons of well-being differences. $w_i(x) - w_j(y) \geq w_k(z) - w_l(zz)$ iff the well-being difference between individual i in outcome x and individual j in outcome y is at least as large as the well-being difference between individual k in outcome z and individual l in outcome zz, with x, y, z, zz the same outcome or partly or wholly distinct. If $i = j = k = l$, this is an intrapersonal difference comparison. If not $i = j = k = l$, this is an interpersonal difference comparison. See Adler, chapter 2, sections 2.2, 2.5, this volume.

[5] A vast amount of work in economics is devoted to the theory and estimation of individual utility functions—a utility function being a representation of individual preferences. However, most writing about utility functions is either skeptical about interpersonal comparability or simply does not address the issue. What is distinctive about the scholarship that we label the "equivalence approach" and "vNM approach" is its ambition both to make interpersonal comparisons and to respect preferences. The literatures covered in Sections 3.3 and 3.4 are—as far as we're aware—the most systematic attempts to do both.

based approaches: Objective lists and "capabilities", on the one hand, and subjective well-being (SWB), on the other. Although the authors of this chapter favor a preference account of well-being, the SWF framework is not committed to such an account.

A terminological note: The reader will observe that the term "iff" is used with some frequency in the text and notes. This is shorthand for "if and only if."

2 Attributes, Preferences, and Well-Being Measures

The SWF framework is meant to be a tractable decision procedure for evaluating governmental policies. In order for this procedure to be tractable, a given outcome x cannot be a complete "possible world"; thinking in detail about whole possible worlds is beyond the cognitive capacity of human decisionmakers. Rather, an outcome is a *model* of a possible world. (See Adler, chapter 2, section 2.2, this volume).

In particular, the *welfarist* decisionmaker—who believes that ethical evaluation depends upon individual well-being—conceptualizes an outcome as a series of attribute bundles for each member of the population of ethical concern. A given outcome x assigns to each individual i an attribute bundle. Let $a_i(x)$ be the attribute bundle of a particular individual i in outcome x. To avoid notational clutter we drop the explicit reference to the outcome or the individual when that is convenient and use a as our generic symbol for an attribute bundle. The set of all attribute bundles under consideration is denoted **A**.[6]

The types of individual attributes described by an attribute bundle are *some* (not necessarily all) of the characteristics of an individual that determine her well-being, such as her income or consumption; her leisure; her health; and/or the level of various public goods that she and others experience (e.g., environmental quality). This list is not exhaustive and other characteristics can be added to the list. Nor is it mandatory to include all these characteristics. Indeed, in most applied work only some of these standard attributes are included. For example, the optimal tax tradition in SWF scholarship often assumes that an individual's attribute bundle consists of her consumption and her leisure and does not describe her health, the quality of her environment, or other characteristics. (See Tuomala and Weinzierl, chapter 4, this volume.) How fully to describe attribute bundles is a matter of balancing the accuracy gains of more complicated bundles against the tractability costs. The literature, to date, lacks a rigorous account of how to undertake this balancing, and we will not attempt to tackle the problem here. Rather, we direct

[6] **A** should include, at least, every bundle which is possessed by some individual i in some outcome x in the set **O** of outcomes that is being used in the analysis (Adler, chapter 2, section 2.2, this volume.) It may include additional bundles. Our analysis in this chapter doesn't depend on which bundles are included in **A**.

our attention to an orthogonal problem, namely how to construct a well-being measure for any characterization of individual attributes.

In principle, the well-being numbers inputted into an SWF are *lifetime numbers*. (See Adler, chapter 2, section 2.2, this volume.) Outcome x corresponds to the vector ($w_1(x)$, …., $w_N(x)$), with $w_i(x)$ indicating the lifetime well-being of individual i in outcome x. Note that if well-being numbers were meant instead to measure an individual's "sublifetime well-being"—her well-being during a portion of her life, e.g., a year—then the vector corresponding to each outcome would consist in a series of numbers for each individual i, rather than a single number $w_i(x)$.

Some policy-analysis scholarship employs multi-period models; an individual is described as living for one or more periods, and as having characteristics that can vary from period to period. An explicitly multi-period representation of individual lives is often adopted, for example, in public health scholarship. (See Cookson, Norheim and Skarda, chapter 6, this volume. See also Hammitt and Treich, chapter 7, this volume.) Individuals are seen as endowed with a lifespan and with a health state for each year alive. Much of the economics literature, however, employs simple, one-period models. Each individual has some single cluster of characteristics during the one period. In effect, the one-period framework ignores both interpersonal heterogeneity in lifespan and variation over time in a given person's attributes. This one-period setup is quite unrealistic, to be sure—individuals in reality *do* die at different ages and *do* experience intertemporal change in their characteristics—but may be substantially more tractable than the multi-period approach.

What follows is agnostic as between single-period and multi-period models. If the one-period representation is used, then each bundle a describes a combination of attributes during the single period. If the modeler instead follows the multi-period route, then each bundle a is a series of period bundles; it describes the individual's lifespan (the number of periods she exists) and, for each period, her attributes at that time.

On a preference view of well-being, an individual's well-being in outcome x depends on her attribute bundle, plus her preferences. We denote the preference of individual i in outcome x as $R_i(x)$.[7] Again, we drop the explicit reference to the outcome or individual when that is convenient. Moreover, it will be useful for the rest of this chapter to introduce a clear symbolism that allows us to make a distinction between *ordinal* preferences and *lottery*

[7] Note that we are assuming that an individual's preferences remain the same over her lifetime. In a multi-period model, this assumption is genuinely restrictive. See Fleurbaey and Tadenuma (2014) and Adler (2020) discussing how, respectively, the equivalence approach and vNM approach might be generalized to allow for variation in preferences over time. Space limitations prevent our discussing this here.

preferences. Ordinal preferences play a central role in the "equivalence approach" (Section 3.3), while lottery preferences are central to the "vNM approach" (Section 3.4).

We use R with a single bar on top to denote an ordinal preference: $\bar{R}$. A given ordinal preference $\bar{R}$ is a ranking of the set **A** of attribute bundles. When we write $a\ \bar{R}\ a^{\star}$, this should be read as: attribute bundle a is weakly preferred to $a^{\star}$ according to the ordinal preference $\bar{R}$. The set of all ordinal preferences under consideration is denoted $\bar{\mathbf{R}}$.[8] We use R with a double bar on top to denote a lottery preference: $\bar{\bar{R}}$. A given lottery preference $\bar{\bar{R}}$ is a ranking of the lottery set $\mathbf{L}^{\mathbf{A}}$, i.e., the set of all (finite) lotteries over **A**. The set of all lottery preferences under consideration is denoted $\bar{\bar{\mathbf{R}}}$. $a\ \bar{\bar{R}}\ a*$ means that the lottery preference $\bar{\bar{R}}$ weakly prefers bundle a to bundle $a*$.[9]

The generic symbol R (without bar) will be used as an umbrella term to refer both to an ordinal preference $\bar{R}$ and a lottery preference $\bar{\bar{R}}$. Having the generic symbol R at hand will make it easier to express what the equivalence and vNM approaches have in common.

The history $h_i(x)$ of individual i in outcome x is a combination of her attribute bundle $a_i(x)$ and her preference $R_i(x)$, so that $h_i(x) = (a_i(x), R_i(x))$. Also here we drop the explicit reference to the outcome or individual in the symbol when that is convenient. We write $h = (a, R)$ to indicate that a history consists of an attribute bundle and a preference. The project of well-being measurement can be seen as assigning well-being numbers to histories: $w(\cdot) = w(a, R)$, with R either an ordinal preference $\bar{R}$ (in the equivalence approach) or a lottery preference $\bar{\bar{R}}$ (in the vNM approach).[10]

---

[8] "Under consideration," here, has an analogous meaning as for the set **A** of attributes bundles (see above, note 6). $\bar{\mathbf{R}}$ should include every preference $\bar{R}$ possessed by some individual i in some outcome x in the set **O** of outcomes, and may include additional preferences. "Under consideration" with respect to the set $\bar{\bar{\mathbf{R}}}$ of lottery preferences, as stated in the text immediately below, is defined similarly.

[9] More precisely: $\bar{R}$ is a transitive, reflexive, and complete binary relation (an ordering) on **A**. From the weak preference $\bar{R}$ we can derive strict preference $\bar{P}$ and indifference $\bar{I}$ in the standard manner. $a\ \bar{P}\ a*$ iff $a\ \bar{R}\ a*$ and not $a*\ \bar{R}\ a$. $a\ \bar{I}\ a*$ iff $a\ \bar{R}\ a*$ and $a*\ \bar{R}\ a$.

A finite lottery l in $\mathbf{L}^{\mathbf{A}}$ assigns a probability number $\pi_l(a)$ to each attribute bundle in **A** such that: $0 \le \pi_l(a) \le 1$, $\pi_l(a) > 0$ for only a finite number of a, and $\sum_{a \in \mathbf{A}} \pi_l(a) = 1$. A lottery preference $\bar{\bar{R}}$ is an ordering of $\mathbf{L}^{\mathbf{A}}$.

Note that each attribute bundle a corresponds to a "degenerate" lottery: the lottery that assigns probability 1 to that bundle, and 0 to every other bundle. Thus, a lottery preference $\bar{\bar{R}}$, in its ranking of the lottery set $\mathbf{L}^{\mathbf{A}}$, includes a ranking of the attribute set **A** itself. We require a lottery preference only to rank finite rather than infinite lotteries because the apparatus of vNM utility theory, and thus the vNM approach to well-being measurement, can be formulated without the additional complexity of infinite lotteries. See Adler (2016, p. 482, n. 11).

[10] Let $(a_i(x), \bar{R}_i(x))$ denote the history (according to the equivalence approach) of individual i in outcome x. The equivalence approach then sets $w_i(x)$ equal to the well-being number of $(a_i(x), \bar{R}_i(x))$. Abusing symbolism slightly: $w_i(x) = w(a_i(x), \bar{R}_i(x))$. Strictly speaking $w(\cdot)$ which assigns well-being numbers to individuals in outcomes is a different function from one that assigns well-being numbers to histories. We should use a different

At this juncture, it's important to concede that not all preference relations are equally appealing for a preference-based well-being analysis. Consider, for instance, Ann, who has whimsical preferences that are internally inconsistent, or Bob, whose preferences are formed based on fake news and other forms of misinformation. Preferences can therefore be "laundered" first, so that the resulting laundered preference relations tell us how individuals would rank attribute bundles were they sufficiently well-informed and satisfied stipulated rationality criteria. The laundering may even go further and correct some preferences that are considered ethically unappealing, such as sadistic preferences. To be sure, the precise nature of this laundering is itself a topic for normative discussion (see, e.g., Goodin, 1986). Laundered preferences may differ from the preferences which are revealed by actual behavior (see, Gul and Pesendorfer, 2008 for a discussion).

In any event, the preferences that we will be discussing in this chapter have been laundered according to whichever criteria of rationality, good information, and ethical acceptability are endorsed as singling out *well-being relevant* preferences.[11] We don't take a position here about the specifics of laundering, beyond assuming that preferences satisfy certain basic formal rationality criteria. [12]

We can now bring to the fore two major similarities between the equivalence and vNM approaches. First, as already stated in the introduction to this chapter, both accept interpersonal comparability. Specifically, both endorse some version of the Same History Principle. [13]

> **SAME HISTORY PRINCIPLE**: For any two individuals i and j, and any two outcomes x and y, if the history of i in x is (a, R) and the history of j in y is (a, R), then i in x is equally well off as j in y.

---

symbol for the latter function, e.g., θ(·), and say that: $w_i(x) = \theta(a_i(x), \overline{R}_i(x))$. See Adler, chapter 2, section 2.A.3, this volume. However, so as to avoid multiplying symbols, we'll use $w(\cdot)$ to mean both types of function. For most of this chapter, $w(\cdot)$ is used in the sense $w(a, R)$, so there isn't much risk of confusion.

Let $(a_i(x), \overline{\overline{R}}_i(x))$ denote the history (according to the vNM approach) of individual i in outcome x. The vNM approach then sets $w_i(x)$ equal to the well-being number of $(a_i(x), \overline{\overline{R}}_i(x))$. Again abusing symbolism slightly: $w_i(x) = w(a_i(x), \overline{\overline{R}}_i(x))$.

[11] $\overline{R}$ is a *laundered* ordinal preference, and $\overline{\mathbf{R}}$ the set of all laundered ordinal preferences under consideration. $\overline{\overline{R}}$ is a *laundered* lottery preference, and $\overline{\overline{\mathbf{R}}}$ the set of all laundered lottery preferences under consideration.

[12] See note 9. See, however, Fleurbaey and Schokkaert (2013) and Decancq and Nys (2021) on the equivalence approach when individuals have incomplete preferences.

[13] The equivalence approach endorses the ordinal-preference version of the Same History Principle, namely: if two individuals have the same attributes and the same ordinal preferences, then they are equally well off. The vNM approach rejects this version of the Same History principle, for reasons to be explored in Section 3.4.2, but endorses the lottery-preference version: if two individuals have the same attributes and the same lottery preferences, then they are equally well off.

If two individuals have the same history—the same attributes and the same preferences—then they are equally well off. The Same History Principle seems compelling and makes rigorous the common-sense case for interpersonal comparability. (See Adler, chapter 2, this volume, section 2.5.) If two individuals, Isaac and Jules, are identical with respect to everything that determines well-being—they have the same attribute bundle *and* the same preferences—then surely their well-being levels are identical.

The direct upshot of the Same History Principle is to allow for interpersonal comparisons of well-being *levels*. But it also implies that differentiating between intra- and interpersonal comparisons of well-being *differences* is arbitrary. *If* intrapersonal comparisons of differences are meaningful then so, too, must be interpersonal difference comparisons.[14]

Second, both the equivalence and vNM approaches yield *preference-based* well-being measures, in a precise sense. A well-being measure $w(\cdot)$ is *preference-based* if it conforms to the following Deference Principle (namely, deference to preferences):

> **DEFERENCE PRINCIPLE**: For all attribute bundles a and $a^\star$ and all preferences R, $w(a,R) \geq w(a^\star,R)$ iff $a\ R\ a^\star$.

The Deference Principle states that well-being is higher in a preferred attribute bundle. The equivalence and vNM approaches both satisfy the Deference Principle. Whether to adopt the Deference Principle is a normative question. While the underlying idea of sovereignty is a central tenet of economics, the linkage between preferences and well-being is a contested topic in the philosophical literature on well-being (see Section 3.5). In our view, the case for the principle rests upon individual autonomy.

Consider, first, the *intrapersonal* case in which we are determining what the well-being of a given individual (Philippe) would be with two histories. That is to say: we are comparing Philippe having attribute bundle a to Philippe having attribute bundle $a^\star$, on the supposition that Philippe's preferences are R. The attribute bundle $a^\star$ may be obtained from a as the result of a policy intervention, for instance. If Philippe weakly prefers the initial attribute bundle, $a\ R\ a^\star$, but the social planner assigns a higher well-being score to the post-intervention history, $w(a,R) < w(a^\star,R)$, the social planner is substituting her own judgment about what makes Philippe's life goes well for his.

[14] Assume that we endorse *intrapersonal* difference comparisons. We are willing to say, concerning one individual (Stacey), that the difference in Stacey's well-being between outcomes *x* and *y* is greater than the difference in Stacey's well-being between outcomes *z* and *zz*. Now imagine that a second individual, Thom, has the same history in outcome *z** as Stacey does in *z*, and the same history in *zz** as Stacey does in *zz*. Then, by the Same History principle, we should accept that the difference in Stacey's well-being between *x* and *y* is greater than the difference in Thom's well-being between *z** and *zz**—an *interpersonal* difference comparison.

Turn now to the *interpersonal* case where (a, R) is being considered as a possible history for one person, and $(a^{\star}, R)$ as a possible history for a second, distinct, person. We are comparing, for example, the well-being of Philippe with (a, R) to the well-being of Stacey with $(a^{\star}, R)$. Note that, while the attribute bundles of Philippe and Stacey differ, they have the same preferences.[15] If $a\ R\ a^{\star}$, but $w(a, R) < w(a^{\star}, R)$, the social planner goes against the shared view of Philippe and Stacey that Philippe with attribute bundle a is at least as well-off as Stacey with attribute bundle $a^{\star}$. We believe that in both examples the social planner would violate the autonomy of the concerned individuals.

3 The Equivalence Approach

In this section, we discuss the equivalence approach to constructing a well-being measure in line with the Deference Principle.[16] Recall that, for the equivalence approach, an individual's well-being depends upon her attributes and her ordinal preference. A history $h = (a, \bar{R})$ is a combination of an attribute bundle and an ordinal preference $\bar{R}$. A well-being measure $w(\cdot)$ is a function that assigns a numerical value to each history, which can be used as input to an SWF.

An ordinal preference relation is an *indifference map*. The equivalence approach is built on the idea that this indifference map can be used to identify histories that are equally good from a well-being perspective. More precisely, attribute bundles situated on indifference curves that coincide (even if the curves belong to distinct preference orderings) are considered as equally good according to the equivalence approach. The history $(a, \bar{R})$ is equally good as $(a^{*}, \bar{R}^{*})$ if $\bar{R}$ and $\bar{R}^{*}$ share an identical indifference curve and a and $a^{\star}$ are on it. Consequently, one can say that well-being measures of the equivalence approach assign a numerical value to indifference curves.[17] The various well-being measures that belong to the equivalence approach disagree, however, on how to assign a numerical value to histories which are situated on non-coinciding indifference curves.

Two important families of well-being measures can be discerned within the equivalence approach, each with their roots in the theory of fair allocation (see Fleurbaey and Maniquet, 2011). In the first family, a collection of *reference sets* is used to assign a numerical value to each indifference curve. A prominent member of this family is the money metric well-being

---

[15] Decancq et al. (2015a) refer to the deference principle in interpersonal comparisons as the Same Preference Principle.

[16] An excellent introduction to the equivalence approach can be found in Fleurbaey and Blanchet (2013, chapter 4). A more technical discussion can be found in Fleurbaey and Maniquet (2011) and older, but still very relevant, sources are by Deaton (1980) and Deaton and Muellbauer (1980).

[17] Hence, all preference information that is not contained in the shape of the indifference curves is ignored, such as risk aversion (see Adler, 2016; section 3.4, this chapter) or other information about the intensity of preferences (see Sen 1979, p. 11).

measure. In the second family, each indifference curve is labelled by its intersection with a *monotone path*. A prominent member of the second family is the ray utility well-being measure.[18] First, we discuss these families and then turn to the use of the so-called *equivalent income* measure of well-being. Interestingly, the equivalent income measure can be interpreted as a member of both families. We conclude with a clarification regarding difference comparisons. Except where explicitly stated, we simplify our presentation of the equivalence approach by assuming that preferences are monotonic.[19]

3.1 Measuring Well-being with a Collection of Reference Sets

In the first family of well-being measures, a collection of nested reference sets $B = (B_\lambda)_{\lambda \in \mathbb{R}^+}$ is used to assign a numerical value to each indifference curve that can be used as input in an SWF. Each member of the collection of reference sets is indexed by a real number λ in such a way that larger reference sets get larger numbers, so that $\lambda \leq \lambda'$ iff $B_\lambda \subseteq B_{\lambda'}$. Figure 1 provides an example where two reference sets are depicted in a two-attribute space. The reference sets are nested and indexed so that $\lambda < \lambda'$.

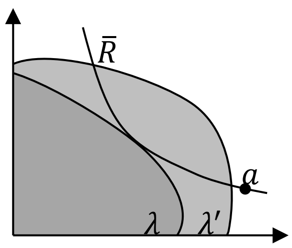

Figure 1. Two nested reference sets and a history

Well-being measures of this family assign the index of the reference set to the history to which it is *equivalent*. A reference set B_λ is said to be equivalent to a history $(a, \bar{R})$ when the individual at hand is indifferent (with her preferences $\bar{R}$) between her bundle of attributes a and being able to choose freely from the reference set B_λ. In the example of Figure 1, the individual is indifferent between her attribute bundle a and being able to choose freely from the reference set indexed by λ. Once each history is assigned a reference set from the collection of reference sets, they can be compared using the indices of the equivalent reference sets.[20] This captures the intuition that choosing freely from a larger reference set cannot be worse than

[18] In a series of recent papers, Fleurbaey and Maniquet (2017, 2018, 2019) have compared the axiomatic properties of the money metric and ray utility well-being measures.

[19] That is: each ordinal preference $\bar{R}$ in $\bar{\mathbf{R}}$ is such that, if attribute bundle *a* has a higher level than *a** of at least one type of attribute, and no lower level of any, then $a \; \bar{P} \; a*$.

[20] We assume that the collection of reference sets is such that: for each indifference curve there is one and only one equivalent reference set.

being able to choose from a smaller reference set. Note furthermore that larger well-being numbers are attached to histories which are situated on higher indifference curves. Hence, we have that $w(a', \bar{R}) \geq w(a, \bar{R})$ iff $a' \ \bar{R}\, a$. This observation illustrates that the Deference Principle is indeed satisfied.[21]

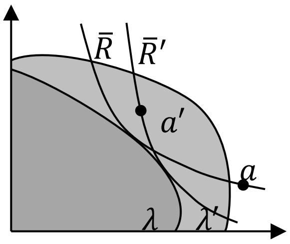

Figure 2. Two histories that are equivalent to the same reference set

The same well-being level is given to all histories that are equivalent to the same reference set, as illustrated in Figure 2. Both individuals are indifferent between their own attribute bundle and choosing freely from the same set B_λ. Even though they have very different attribute bundles and different preferences, the well-being level of both histories is said to be the same because both histories are considered equally good as being able to choose freely from the same reference set. The question which histories are equally good from a well-being perspective is particularly relevant when the well-being measures are used as arguments of a prioritarian SWF. As discussed in Chapter 2, prioritarian SWFs are averse to well-being inequality, and the choice of the well-being measure determines what the ideal of equality precisely entails. When using a well-being measure based on a reference set, the goal is to offer all individuals the same equivalent reference set to choose from. In this sense, the equivalence approach connects the measurement of well-being to an evaluation of opportunity sets. This idea was evoked by Deaton and Muellbauer (1980, Chapter 1) and received renewed attention in the capability approach (Sen, 1985).

The choice of the collection of reference sets, and the way in which they are indexed, matters for the precise well-being levels that are attached to the histories. This choice is a normative one, which should ideally be guided by an ethical theory.[22] In the general case, however, it may seem hard to come up with a natural and intuitive way of indexing the reference sets. Specific cases are arguably easier.

[21] The analytic argument for why the Deference Principle is satisfied runs as follows. We assume (see note 20) that for any two attribute bundles a and a^*, there is a single reference set B_λ to which a is equivalent according to $\bar{R}$, and a single reference set B_{λ^*} to which a^* is equivalent according to $\bar{R}$. If $a \ \bar{I} \ a^*$, then it must be the case that $B_\lambda = B_{\lambda^*}$, hence $\lambda = \lambda^*$. If $a \ \bar{P} \ a^*$, then it can't be the case that $B_\lambda = B_{\lambda^*}$; nor can it be the case that B_λ is a proper subset of B_{λ^*}; hence B_{λ^*} is a proper subset of B_λ and thus $\lambda > \lambda^*$.

[22] See Fleurbaey and Blanchet (2013; p. 141) for some suggestions how such an ethical theory may be developed.

One specific case is the so-called *money metric* well-being measure.[23] Consider a collection of reference sets that are defined as follows $B_\lambda = \{a | \sum a_j \times p_j^* \leq \lambda\}$, where $p^* = (p_1^*, \dots, p_j^*, \dots, p_k^*)$ is a positive reference price vector. These reference sets are reminiscent of standard budget constraints, which contain all (attribute) bundles that can be afforded with an income λ, at prices p^*. In a two-dimensional setting, the reference sets are right-angled triangles with their right angle in the origin, see Figure 3. For a given reference price vector, the reference sets are indeed nested, as more attribute bundles can be afforded with a higher income level.

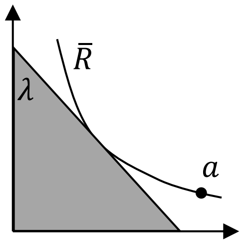

Figure 3. The money metric well-being measure

When using a money metric measure of well-being, well-being is evaluated by looking at the minimal income that is necessary to afford an attribute bundle that is at least as good (according to preferences $\bar{R}$) as the attribute bundle a, given reference prices p^*. Multiplying all reference prices with a constant does not change the shape of the reference sets, but multiplies the well-being measure with the constant. Using a scale invariant SWF to aggregate the well-being measure (e.g., an Atkinson SWF; see Adler, chapter 2, section 2.9, this volume) introduces a convenient independence of the social evaluation with respect to the constant.

The use of money metrics has been criticized for its dependence on the choice of the reference prices p^* (see, for instance, Roberts 1980). But the choice of reference prices need not be arbitrary and should ideally be part of a fully-fledged ethical theory of social welfare. When the attributes are all market goods and the reference prices are chosen equal to the market prices, the money metric well-being measure coincides with ordinary income and, hence, social states will be evaluated with an SWF defined over ordinary incomes. When a lower price is chosen than the market price for a specific attribute, the well-being level of individuals who have a lot of this attribute in their bundle is lowered. However, when using money metrics in a context where the attributes are not available on the market (such as health, political freedom, or social interactions), market prices do not exist to provide a benchmark.

A second critique against the use of money metric well-being measures has been formulated by Blackorby and Donaldson (1988). They observe that a standard *quasi-concave* SWF (e.g., a prioritarian SWF) combined with money metric well-being measures may fail to approve

[23] Money metrics were introduced by McKenzie (1957) and Samuelson (1974) as a tool to number indifference curves.

transfers from richer to poorer individuals. Only an SWF that gives absolute priority to the worst-off approves all richer-to-poorer transfers.

It should be stressed that the prioritarian SWF conforms to the Pigou-Dalton axiom in terms of well-being. (See Adler, chapter 2, this volume, section 2.3.) A pure, gap-diminishing transfer of well-being from someone better off, to someone worse off, will be approved by the prioritarian SWF. If $w(\cdot)$ is constructed using the money-metric approach, then the prioritarian SWF will approve pure, gap-diminishing transfers of money-metric well-being from those at higher levels of money-metric well-being, to those at lower levels. The Blackorby and Donaldson critique concerns the Pigou-Dalton axiom *in the space of attributes*. A quasi-concave (e.g., prioritarian) SWF defined over money-metric well-being will fail to approve some *attribute* transfers from richer to poorer individuals.

This observation has led to several different responses in the literature. First, Blackorby and Donaldson (1988, p. 129) themselves conclude that a social welfare analysis based on money metric well-being measures is flawed. Their conclusion struck a heavy blow against the use of money metric well-being measures in (applied) social welfare analysis. Fleurbaey and Maniquet (2011, p. 21), by contrast, conclude that this observation justifies the leximin (or maximin) SWF, rather than undermining the use of money metric measures as arguments in a social welfare analysis. Recently, Bosmans et al. (2018) have revisited the observation by Blackorby and Donaldson (1988) and question the premise that the approval of richer-to-poorer transfers is an essential requirement for an SWF in a multi-attribute setting. While such transfers clearly lead to a more equitable situation, their price in terms of efficiency may be large. Indeed, a richer-to-poorer transfer in a multi-attribute setting may transfer attributes from a rich person who values the attributes greatly to a poor person who does not value these attributes. Only SWFs that give an absolute priority to equity considerations over efficiency concerns (such as leximin or maximin) consider all such transfers desirable.

3.2 Measuring Well-being with a Monotone Path

We turn now to the second family of well-being measures that belong to the equivalence approach.[24] In this family, a so-called *monotone path* is used to assign a numerical value to each indifference curve. A monotone path is a set of attribute bundles such that, for any two bundles on the path, one of the two "vector-dominates" the other—meaning that one of the two bundles has at least as much of each type of attribute as the other bundle, and strictly more of

[24] Formally, this second family can be seen as a special case of the first (consider a collection of reference sets which each consist of the bundle on the monotone path plus all the attribute bundles which are vector dominated by that bundle). Yet, we believe that it is pedagogical to treat it as a separate family.

at least one type of attribute.[25] Let the attribute bundles on the monotone path be indexed by a real number κ in such a way that vector-dominating attribute bundles get a larger number; that is, $\kappa' > \kappa$ iff $a_{\kappa'}$vector dominates a_κ in the sense just stated; see Figure 4 for an illustration. Attribute bundles on the monotone path can be ranked by the index κ, irrespective of the preferences of the concerned individuals. That means that attribute bundles on the monotone path that have more in some attributes and no less in any are better from the perspective of well-being.[26]

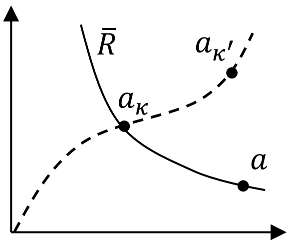

Figure 4. A monotone path and a history

Each history (a, $\bar{R}$) can be assigned an equivalent attribute bundle on the monotone path, which is situated where the indifference curve of preference $\bar{R}$ through a and the monotone path cross. The index of the equivalent attribute bundle can then be used to quantify the well-being of all histories that are situated on the indifference curve. Once each history is assigned an index, comparing them in terms of well-being is straightforward. The indices can also be used as arguments of an SWF. Furthermore, we see that larger indices are attached to histories which are situated on higher indifference curves, so that the Deference Principle is satisfied.[27]

Well-being measures of this family assign the same well-being level to all histories that are equivalent to the same attribute bundle on the monotone path, as illustrated in Figure 5. Both individuals in the figure are indifferent between their own attribute bundle and the attribute bundle a_κ that is situated on the monotone path. By comparing Figure 2 and 5, it is clear that the underlying logic to identifying the histories which are equally good from a well-being perspective differs between the two families of the equivalence approach. Hence, also the ideal of well-being equality differs. While for the first family the ideal of equality is to offer all individuals the same equivalent reference set to choose from, the ideal for the second family is that the actual attribute bundles are equivalent to the same bundle on the monotone path.

[25] We assume that each indifference curve intersects once and only once with the monotone path.

[26] Fleurbaey and Trannoy (2003) prove that ranking attribute bundles with vector dominance is in general incompatible with the idea of respecting preferences. Restricting vector dominance to a monotone path, on the other hand, is compatible with respect for preferences, as shown by Decancq et al. (2015b).

[27] The Deference Principle follows immediately from the definition of monotone path plus the assumption of monotonic preferences. If $\bar{R}$ is monotonic, then—as between any two bundles on the monotone path—it must strictly prefer the one that vector dominates the other and is assigned a larger index number.

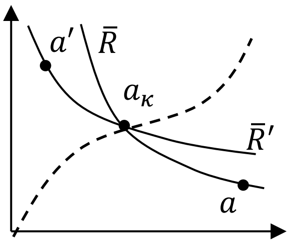

Figure 5. Two histories that are equivalent to the same attribute bundle on the monotone path.

Analogous to the first family, the choice of the monotone path and the way in which the attribute bundles on the monotone path are indexed is a normative one, with important consequences for the precise well-being numbers that are attached to the histories.

The so-called *ray utilities* (or quantity metrics according to Deaton and Muellbauer, 1980, p. 179–182) are prominent examples of well-being measures that use a monotone path to assign a numerical value to indifference curves.[28] In this case, the monotone path is a straight line or ray that connects the origin with a reference bundle r, and the index κ is the ratio of the distance from the origin to a_κ to the distance to r. In other words, the well-being of history $(a, \bar{R})$ is measured by the proportional share of the reference bundle r that is considered equally good (according to preferences $\bar{R}$) as the attribute bundle a. In the illustration of Figure 6, the history $(a, \bar{R})$ obtains a ray utility well-being number of 0.5 since the equivalent attribute bundle a_κ is situated halfway on the ray through the reference bundle r.

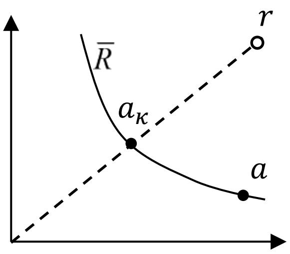

Figure 6. The ray utility well-being measure

How to choose the reference bundle is a normative question. Ray utilities can be criticized along the same lines as the money metrics and a similar challenge arises to develop a full-fledged ethical theory of the choice of the reference bundle. The specific context may determine the appeal of the choice of a specific reference bundle. The reference bundle can consist of the total amount of resources available in the society for each attribute, for instance (Pazner and Schmeidler, 1978). In the context of multidimensional poverty measurement, Decancq et al.

[28] Ray utilities were introduced by Samuelson (1977), Pazner and Schmeidler (1978), and Deaton (1979). See Fleurbaey and Tadenuma (2014) for a recent discussion.

(2019) propose to select the bundle that contains the poverty threshold in each attribute as reference bundle.

Once each history is assigned an equivalent attribute bundle on the ray, evaluating the social welfare of the social outcome becomes a one-dimensional problem (along the ray). Note that shifting the reference bundle r itself along the ray leads to a rescaling of all well-being numbers with a constant. When a scale invariant SWF is used to aggregate the well-being numbers in a second stage (e.g. an Atkinson SWF; see Adler, chapter 2, section 2.9, this volume), the precise position of the reference bundle along the ray becomes irrelevant, so that only the direction of the ray matters.

3.3 Measuring Well-being with Equivalent Incomes

In this subsection we discuss the equivalent income well-being measure, which has received some recent attention in theoretical and empirical work.[29] To introduce the equivalent income well-being measure, let us write the attribute bundle $a = (y, b)$ in which y denotes income and b a vector of non-income dimensions. For the non-income attributes we select a reference value (denoted $\tilde{b}$), for instance the situation in which the individual suffers no health problems, obtains the desired level of social interactions, etc. The equivalent income y^* is the income level that, combined with the reference value on the non-income attributes, would place the individual in a situation that she finds equally good as her actual situation. More precisely, the equivalent income y^* is defined by

$$(y, b)\ \bar{I}\ (y^*, \tilde{b})$$

where $\bar{I}$ denotes indifference (according to ordinal preferences $\bar{R}$).

Figure 7 illustrates the case in which there is one non-income dimension, specifically health. The individual is indifferent between the current attribute bundle $a = (y, b)$ and a hypothetical bundle $(y^*, \tilde{b})$. She considers having the equivalent income level and being in the reference health situation equally good as her current situation with more income and a worse health. We call the difference between the current income level y and the equivalent income level y^*, the *willingness-to-pay* to be in the reference situation.

[29] See Fleurbaey and Schokkaert (2013), Decancq et al. (2015a,b), Decancq and Schokkaert (2016), or Fleurbaey (2016) for recent discussions of the equivalent income well-being measure.

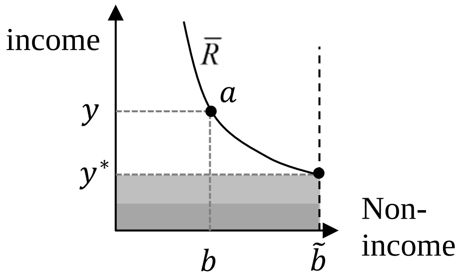

Figure 7. The equivalent income well-being measure

The equivalent income well-being measure can be interpreted as a member of both families that belong to the equivalence approach. Equivalent incomes result from a specific choice of reference sets as well as a specific choice of the monotone path. Figure 7 illustrates two rectangular reference sets $B_\lambda = \{(y,b) | y \leq \lambda, b \leq \tilde{b}\}$, for which the index λ coincides with the equivalent income measure y^*. Alternatively, the dashed line formed by all attribute bundles with non-income dimension equal to $\tilde{b}$ is a monotone path. By indexing the attribute bundles on this monotone path with their income level, the resulting well-being measure coincides with the equivalent income measure.

How to choose the reference value $\tilde{b}$? Similar to other examples of the equivalence approach, selecting the reference value is a normative choice. Observe that two histories on the monotone path can be ranked based on their income level, *irrespective of the preferences*. In the case when the attributes are income and health, Fleurbaey and Schokkaert (2013) argue that "perfect health" (i.e. a situation without health problems) constitutes an appealing reference value. The argument is that in the hypothetical situation where nobody suffers from health problems, well-being comparisons can be made based on income, without taking preferences or the willingness-to-pay for health improvements into account. Yet, as soon as some people suffer from health problems, their willingness-to-pay for health improvements matters to evaluate the severity of their health problems. Preferences about the importance of health matter when comparing situations with poor health, but not when health is perfect.

Reference values need not be the same for all individuals. Fleurbaey and Blanchet (2013 p. 149) and Decancq et al. (2015a) propose to use individualized reference values. These reference values can be chosen as the best value of the non-income attribute for each individual, for instance. They may prefer a certain amount of leisure time, for instance. Deviations from this best value are considered as worse by the individual; see Figure 8 for an illustration.[30]

[30] In Figure 8, the preferences are not monotonic. We assume merely that each indifference curve intersects once with the monotone path (the vertical line at $\tilde{b}$) and that the more income is always preferred to less. If so, the Deference Principle is satisfied.

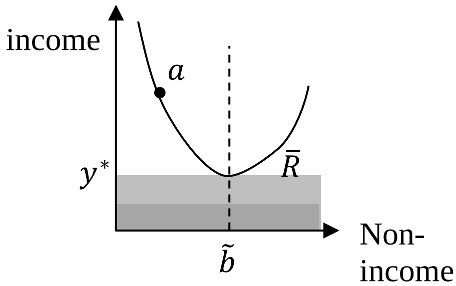

Figure 8. Equivalent incomes when preferences are not monotone

Conveniently, equivalent income is measured in monetary units (in the same units as the income dimension). This has the advantage that the values can be easily interpreted. Yet, the monetary measurement unit of equivalent income should not be interpreted as a reductionist attempt to neglect the importance of the non-income attributes. On the contrary, when using equivalent incomes, each individual values the difference between her actual achievements in the non-income attributes and the reference situation according to her own preferences, by means of her willingness-to-pay.

3.4 The Equivalence Approach and Difference Comparisons

Finally, we turn to the question how the equivalence approach can provide information about well-being differences. Comparisons of well-being levels are insufficient for a prioritarian SWF or the utilitarian SWF (see Adler, chapter 2, section 2.5, this volume).[31] Both require *more* than well-being level information. At a minimum, these SWFs require well-being *difference* information.[32] How can the equivalence approach provide such information?

The answer, in a nutshell, is that the equivalence approach assigns a unique number to each history $(a, \bar{R})$. In the reference-set version of the approach, this is accomplished by assigning an index number to each reference set; the unique number $w(a, \bar{R})$ is, then, the index number

---

[31] Imagine a well-being measure $w(\cdot)$ that assigns well-being numbers to histories, but is understood only to contain information about the well-being levels of the histories. If so, we can define a new measure $w^+(\cdot)$, by picking any strictly increasing function $f(\cdot)$, and defining $w^+(\cdot)$ as follows: $w^+(a, \bar{R}) = f(w(a, \bar{R}))$ for every history $(a, \bar{R})$. Note that $w^+(\cdot)$ and $w(\cdot)$ agree in their intra- and interpersonal comparisons of well-being levels. Moreover, the leximin SWF will be invariant to the substitution of $w^+(\cdot)$ for $w(\cdot)$. But this is not true of either the utilitarian SWF or a prioritarian SWF. See Adler, chapter 2, section 2.5, this volume.

[32] As discussed in chapter 2, section 2.9.1, prioritarian SWFs indeed require *more* than level and difference information. While utilitarian SWFs are invariant to a common cardinal rescaling of well-being numbers, prioritarian SWFs are not. Atkinson prioritarian SWFs *are* invariant to a common ratio rescaling of well-being numbers. How to go from a well-being measure that contains well-being level and difference information, to one that contains well-being level, difference, and ratio information, is addressed in chapter 2, section 2.9.1, 2.A.3.1.

Our focus in this chapter is simply how to construct a $w(\cdot)$ that is preference-based and contains both well-being level and well-being difference information.

of the reference set such that $\bar{R}$ is indifferent between a and that set. In the monotone-path version, this is accomplished by picking an index number for each location along the path. The unique number $w(a, \bar{R})$ is then the index number of the location along the path such that $\bar{R}$ is indifferent between a and that location.

We need to be cautious in how we use the unique well-being number assigned to each history $(a, \bar{R})$. The numbers may contain spurious information. They may "indicate" a well-being fact regarding histories that we don't believe to be meaningful.[33]

However, if we *do* believe that well-being difference comparisons are meaningful, then we *can* read off such information from $w(\cdot)$ as constructed with the equivalence approach. We can say that the well-being difference between $(a, \bar{R})$ and $(a^*, \bar{R}^*)$ is at least as large as the well-being difference between $(a^+, \bar{R}^+)$ and $(a^{++}, \bar{R}^{++})$ iff $w(a, \bar{R}) - w(a^*, \bar{R}^*) \geq w(a^+, \bar{R}^+) - w(a^{++}, \bar{R}^{++})$.

Conversely, in choosing an indexing scheme for reference sets or monotone paths, we should be sensitive to what these numbers imply as regards both well-being level and well-being difference comparisons. This is critical if we will be using the numbers as inputs into a utilitarian or prioritarian SWF. Indeed, we have stressed throughout this section that the choice of reference sets or a monotone path involves normative judgment. *If* we will be interpreting $w(\cdot)$ numbers to indicate not only well-being levels but also well-being differences, then the normative judgment in indexing reference sets or a monotone path is—precisely—a normative judgment regarding both level and difference comparisons between histories.[34]

4 The vNM Approach

The vNM approach conceptualizes a history as a combination of an attribute bundle and a lottery preference: $h = (a, \bar{\bar{R}})$. It assumes, further, that the lottery preference satisfies the axioms of vNM utility theory, and that the ranking of history lotteries satisfies a principle of Lottery Deference. Using these assumptions plus several others, the vNM approach derives its "Fundamental Formula," which states that the well-being number of a given history is the *utility* number of the attribute bundle in that history—as per an appropriately scaled vNM utility function representing the lottery preference.

---

[33] For example, we may not believe that it is meaningful to assign a single magnitude to the well-being difference between histories. See chapter 2, section 2.9.2, 2.A.3.2. If so, we should not interpret a well-being measure $w(\cdot)$ as constructed with the equivalence approach as containing such information. For example, $w(\cdot)$ may be such that $w(a, \bar{R})$ is 10, while $w(a^*, \bar{R}^*)$ is 7. The numbers "indicate" that the magnitude of the well-being difference between the two histories is 3. But if we believe that such a magnitude is spurious information, we shouldn't read $w(\cdot)$ as indicating this.

[34] While the Deference Principle itself determines our well-being level comparisons of two histories with the same preference, it *doesn't* determine our well-being level comparisons of histories containing different preferences, nor does it determine our difference comparisons between histories.

The vNM approach originates with the pioneering scholarship of John Harsanyi. Adler, building upon Harsanyi's foundational work, has further developed the methodology; and the version of the vNM approach set forth here will be as presented in Adler (2016; 2019 ch. 2, appendix). vNM utilities (shorthand for "von Neumann-Morgenstern") are, of course, a pervasive tool in modern economics. What is distinctive about the line of thought originating with Harsanyi is the use of vNM utility functions as the linchpin of a theory of interpersonal comparisons. We use the term "vNM approach" to mean, specifically, this theoretical apparatus—as opposed to vNM utility theory more generally.

To date, little empirical work has been undertaken that explicitly references the vNM approach. (By contrast, the equivalence approach has been implemented in a sizeable empirical literature.) See Decancq and Neumann (2016); Decancq and Schokkaert, chapter 5, this volume, for empirical applications of the vNM approach. However, some scholarship on SWFs does employ vNM utility functions as the input to an SWF.[35] The vNM approach (or something like it) is implicit in such scholarship.

In what follows, we first present the approach; then contrast it with the equivalence approach; and finally discuss the special case of homogeneous preferences.

4.1 The vNM Approach: Elements

vNM utility functions are ubiquitous in modern economics, as mentioned. vNM utility theory shows the following. (Gilboa 2009; Kreps 2013.) If we require that a lottery preference $\bar{\bar{R}}$ (an ordering of the set of lotteries)[36] conform to a few basic rationality axioms, then $\bar{\bar{R}}$ can be *expectationally represented* by a vNM utility function, $u^{\bar{\bar{R}}}(\cdot)$. For any two lotteries, the first is preferred to the second by $\bar{\bar{R}}$ iff the expected utility of the first lottery is greater than the expected utility of the second, and $\bar{\bar{R}}$ is indifferent between the two lotteries iff the two lotteries have the same expected utilities—these expected utilities calculated using $u^{\bar{\bar{R}}}(\cdot)$.[37] The symbolism $u^{\bar{\bar{R}}}(\cdot)$ indicates that this is a utility function corresponding to $\bar{\bar{R}}$: the numbers it assigns to lotteries, by taking expected values, are such as to mirror the ranking of lotteries by $\bar{\bar{R}}$.

[35] For example, the SWF-based literature on climate change often uses a utilitarian SWF with individual utility, as a function of consumption, calculated with a vNM utility function of the constant-relative-risk-aversion from. See Botzen and van den Bergh (2014, p. 3).

[36] See note 9.

[37] That is: $l \; \bar{\bar{R}} \; l^*$ iff $\sum_{a \in \mathbf{A}} \pi_l(a) u^{\bar{\bar{R}}}(a) \geq \sum_{a \in \mathbf{A}} \pi_{l^*}(a) u^{\bar{\bar{R}}}(a)$.

vNM utility theory also shows that a utility function representing a lottery preference is *unique up to a positive affine transformation*. A utility function $u^{\bar{\bar{R}}}(\cdot)$ that expectationally represents lottery preference $\bar{\bar{R}}$ is not unique. Rather, there is an entire family of utility functions that do so. If we pick one member of this family ($u^{\bar{\bar{R}}}(\cdot)$), then every other member can be derived from $u^{\bar{\bar{R}}}(\cdot)$ by multiplying by a positive constant and adding a constant.

The vNM approach constructs a well-being measure $w(\cdot)$ for histories as follows. The ethical deliberator is posited to have a ranking of histories and of differences between histories. If this ranking is well-behaved, there exists a $w(\cdot)$ that represents it. $w(h) \geq w(h^*)$ iff history h is judged by the deliberator to be at least as good for well-being as history h^*. $w(h) - w(h^*) \geq w(h^+) - w(h^{++})$ iff the well-being difference between h and h^* is judged by the deliberator to be at least as large as the well-being difference between h^+ and h^{++}.

The ethical deliberator is also posited to have a ranking of *history lotteries*. From the set **A** of attribute bundles and set $\bar{\bar{\mathbf{R}}}$ of lottery preferences, we arrive at a set **H** of histories (all pairings of an attribute bundle and a lottery preference). $\mathbf{L^H}$ is the set of all lotteries over **H**. Each lottery L in this set is such as to assign probability numbers, summing to unity, to various histories. The deliberator is posited to have a ranking of $\mathbf{L^H}$. Further, this ranking is formally well-behaved so that there exists a function $\zeta(\cdot)$ that expectationally represents it. One history lottery L is judged by the deliberator to be at least as good as a second history lottery L^* iff the expected $\zeta(\cdot)$ value of L is at least as large as the expected $\zeta(\cdot)$ value of L^*.[38]

So we have three different mathematical indicators: a vNM utility function $u^{\bar{\bar{R}}}(\cdot)$ for each preference $\bar{\bar{R}}$ regarding lotteries over attribute bundles; $w(\cdot)$ representing the deliberator's judgments regarding the well-being levels of histories and well-being differences between them; and $\zeta(\cdot)$ expectationally representing the deliberator's ranking of history lotteries. The vNM approach now posits two axioms that create a tight linkage between these indicators.

First, it posits a "Lottery Deference" principle. The vNM approach accepts the Deference Principle (above) but goes further. It says not only that the ranking of histories with a common preference defers to that preference, but that the ranking of *lotteries* over histories with a common preference defers to that preference.

[38] Let $\pi_L(h)$ be the probability assigned to history h by L. Then $\zeta(\cdot)$ is such that the deliberator judges L to be at least as good for well-being as L^* iff $\sum_{h \in \mathbf{H}} \pi_L(h)\zeta(h) \geq \sum_{h \in \mathbf{H}} \pi_{L^*}(h)\zeta(h)$.

LOTTERY DEFERENCE PRINCIPLE: If history lotteries L and L^* are such as to assign non-zero probabilities to histories all of which contain the very same preference $\bar{\bar{R}}$, then the deliberator defers to $\bar{\bar{R}}$ in ranking the history lotteries.[39]

Adler defends the Lottery Deference Principle by appealing to the very same considerations of individual autonomy that justify the Deference Principle. Consider, for simplicity, the intrapersonal case. Consider the subset of histories constructed by combining attribute bundles with the very same lottery preference $\bar{\bar{R}}^+$. L and L^* are lotteries over this subset. Let's imagine that $\bar{\bar{R}}^+$ is indeed the actual lottery preference of Philippe. In comparing L to L^*, then, we are comparing lotteries over different possible lives for Philippe—that is, lives consisting of one or another attribute bundle paired with Philippe's lottery preference $\bar{\bar{R}}^+$.

The Deference Principle says we should respect Philippe's preference in determining which would be a better life for him to lead for certain. If $\bar{\bar{R}}^+$ prefers a to a^*—that is Philippe, holding preference $\bar{\bar{R}}^+$, favors a over a^*—then the Deference Principle tells us to prefer the life in which Philippe has a, to the one in which he has a^*. Surely then (or so Adler argues) we should also respect Philippe's preference in comparing one lottery over lives he might lead to a second.

It turns out that Lottery Deference implies a tight linkage between $\zeta(\cdot)$ and each $u^{\bar{\bar{R}}}(\cdot)$. The $\zeta(\cdot)$ value assigned to any given history $(a, \bar{\bar{R}})$ must be either equal to the utility value assigned to bundle a by $u^{\bar{\bar{R}}}(\cdot)$ or a positive affine transformation of that utility value.

Second, the vNM approach posits a "Bernoulli" principle, which won't be discussed at length here, but says—in a nutshell—that the deliberator is risk neutral with respect to well-being in ranking history lotteries. As between a history that yields a well-being value w, and a lottery with a 50% chance of increasing well-being to $w + \Delta w$, and a 50% chance of decreasing well-being to $w - \Delta w$, the deliberator judges the two to be equally good. She does not give more weight to a chance of a well-being gain over an equal chance of a well-being loss, or vice versa. Bernoulli (which is intuitively quite plausible) has powerful implications. Simplifying slightly, it implies that $w(h) = \zeta(h)$ for every h.[40] Understand that these indicators are conceptually distinct: $w(\cdot)$ tracks the deliberator's judgments of levels and differences, $\zeta(\cdot)$ her judgments

---

[39] Here is a precise statement of Lottery Deference. Let L, L^*, and $\bar{\bar{R}}^+$ be such that: if $\pi_L(a, \bar{\bar{R}}) > 0$ then $\bar{\bar{R}} = \bar{\bar{R}}^+$ and if $\pi_{L^*}(a, \bar{\bar{R}}) > 0$ then $\bar{\bar{R}} = \bar{\bar{R}}^+$. Let l and l^* be the two attribute-bundle lotteries corresponding to L and L^*, respectively, namely: $\pi_l(a) = \pi_L(a, \bar{\bar{R}}^+)$ and $\pi_{l^*}(a) = \pi_{L^*}(a, \bar{\bar{R}}^+)$ for all a. Then the deliberator judges L to be at least as good for well-being as L^* iff $l\ \bar{\bar{R}}^+\ l^*$.

[40] Strictly, it implies that $w(\cdot)$ is a positive affine transformation of $\zeta(\cdot)$; but since all positive affine transformations of $w(\cdot)$ equally well represent comparisons of history levels and differences, and all positive affine transformations of $\zeta(\cdot)$ equally well expectationally represent comparisons of history lotteries, we can without loss of generality set $w(\cdot) = \zeta(\cdot)$.

regarding lotteries. Bernoulli means that the $w(\cdot)$ numbers assigned to histories are equal to the $\zeta(\cdot)$ numbers.

Putting together Lottery Deference and Bernoulli, we end up with the Fundamental Formula of the vNM approach:

$$\textit{Fundamental Formula}: \; w(a,\bar{\bar{R}}) = s(u^{\bar{\bar{R}}})u^{\bar{\bar{R}}}(a) + t(u^{\bar{\bar{R}}}) \;,\; s(u^{\bar{\bar{R}}}) > 0$$

The Fundamental Formula says that the well-being number assigned to a history (a, $\bar{\bar{R}}$) is equal to, or a positive affine transformation of, the utility number of the attribute bundle in that history (a) as per a vNM utility function $u^{\bar{\bar{R}}}(\cdot)$ expectationally representing the preference in that history($\bar{\bar{R}}$).

The Fundamental Formula means that the ethical deliberator's judgments regarding well-being levels and differences are tightly constrained by the vNM utility functions representing the various preferences. Each lottery preference in $\bar{\bar{\mathbf{R}}}$ is represented by a vNM utility function $u^{\bar{\bar{R}}}(\cdot)$. For each such function, the deliberator chooses two scaling factors, $s(u^{\bar{\bar{R}}}) > 0$ and $t(u^{\bar{\bar{R}}})$. (The scaling factors used for the vNM utility function representing one lottery preference need not be the same as the factors used for the vNM utility function representing a different lottery preference; hence the scaling factors are written as $s(u^{\bar{\bar{R}}})$ and $t(u^{\bar{\bar{R}}})$ rather than simply s and t.) Having chosen these utility-function scaling factors, the deliberator has fully specified $w(\cdot)$. See Table 1 for an illustration of the role of the utility-function scaling factors. The axioms of Lottery Deference and Bernoulli mean that the deliberator's *only* flexibility in constructing $w(\cdot)$ is in choosing the scaling factors.

Table 1. The role of scaling factors in the vNM approach

Attribute bundle (a)	$u^{\overline{\overline{R}}}(a)$	$u^{\overline{\overline{R}}^+}(a)$
a^*	1	1
a^{**}	3	5
a^{***}	5	4

Attribute bundle (a)	$s(u^{\overline{\overline{R}}}) = 50$ $t(u^{\overline{\overline{R}}}) = 0$ $w(a, \overline{\overline{R}})$	$s(u^{\overline{\overline{R}}^+}) = 1$ $t(u^{\overline{\overline{R}}^+}) = 0$ $w(a, \overline{\overline{R}}^+)$
a^*	50	1
a^{**}	150	5
a^{***}	250	4

Attribute bundle (a)	$s(u^{\overline{\overline{R}}}) = 1$ $t(u^{\overline{\overline{R}}}) = 0$ $w(a, \overline{\overline{R}})$	$s(u^{\overline{\overline{R}}^+}) = 1$ $t(u^{\overline{\overline{R}}^+}) = 10$ $w(a, \overline{\overline{R}}^+)$
a^*	1	11
a^{**}	3	15
a^{***}	5	14

Attribute bundle (a)	$s(u^{\overline{\overline{R}}}) = 1$ $t(u^{\overline{\overline{R}}}) = 3$ $w(a, \overline{\overline{R}})$	$s(u^{\overline{\overline{R}}^+}) = 100$ $t(u^{\overline{\overline{R}}^+}) = -99$ $w(a, \overline{\overline{R}}^+)$
a^*	4	1
a^{**}	6	401
a^{***}	8	301

Explanation: The top part of the table displays two vNM utility functions: $u^{\overline{\overline{R}}}(\cdot)$, a vNM utility function representing preference $\overline{\overline{R}}$; and $u^{\overline{\overline{R}}^+}(\cdot)$, a vNM utility function representing preference $\overline{\overline{R}}^+$. This part of the table shows the values that each of these utility functions assigns to three attribute bundles: a^*, a^{**}, a^{***}. The bottom part of the table shows the well-being numbers assigned to six histories (the six possible combinations of the three bundles and the two preferences), using the Fundamental Formula—given the choice of scaling factors for utility function $u^{\overline{\overline{R}}}(\cdot)$, namely $s(u^{\overline{\overline{R}}})$ and $t(u^{\overline{\overline{R}}})$, and scaling factors for utility function $u^{\overline{\overline{R}}^+}(\cdot)$, namely $s(u^{\overline{\overline{R}}^+})$ and $t(u^{\overline{\overline{R}}^+})$. Note that the well-being level and difference comparisons between histories with the same lottery preference are independent of the choice of scaling factors. For example, as illustrated in the bottom part of the table, $w(a^{**}, \overline{\overline{R}}^+) > w(a^{***}, \overline{\overline{R}}^+) > w(a^*, \overline{\overline{R}}^+)$ and $w(a^{***}, \overline{\overline{R}}^+) - w(a^*, \overline{\overline{R}}^+) > w(a^{**}, \overline{\overline{R}}^+) - w(a^{***}, \overline{\overline{R}}^+)$. However, well-being level and difference comparisons between histories with different lottery preferences are sensitive to the choice of scaling factors. For example, using the scaling factors at the left of the bottom part of the table $(s(u^{\overline{\overline{R}}}) = 50, t(u^{\overline{\overline{R}}}) = 0, s(u^{\overline{\overline{R}}^+}) = 1, t(u^{\overline{\overline{R}}^+}) = 0)$, we have that $w(a^*, \overline{\overline{R}}) > w(a^*, \overline{\overline{R}}^+)$ and that $w(a^{**}, \overline{\overline{R}}) - w(a^*, \overline{\overline{R}}) > w(a^{**}, \overline{\overline{R}}^+) - w(a^*, \overline{\overline{R}}^+)$. But using the scaling factors in the middle part of the table $(s(u^{\overline{\overline{R}}}) = 1, t(u^{\overline{\overline{R}}}) = 0, s(u^{\overline{\overline{R}}^+}) = 1, t(u^{\overline{\overline{R}}^+}) = 10)$, these inequalities are reversed.

Although the intellectual path to the Fundamental Formula is complicated, the formula itself is reasonably intuitive. It says, simply, that the well-being number of a history is a rescaling of the vNM utility number of the attribute bundle in the history. Because vNM utility functions are only unique up to a positive affine transformation, the deliberator has flexibility to rescale the utility numbers—and that's where the choice of scaling factors comes into play.

The final component of the vNM approach is some theory as to how the scaling factors should be selected. The strong principle of deference to preferences (Deference and indeed Lottery Deference) and risk neutrality in well-being (Bernoulli) do not themselves provide guidance on this topic. Adler (2019, ch. 2) suggests a "high-low" approach: choose two attribute bundles, a "high" bundle a^{high} and a "low" bundle a^{low}, such that individuals are judged to be equally well off with a^{high} independent of their preferences, and equally well off with a^{low} independent of their preferences. This suffices to fix the scaling factors. For a different theory of the choice of scaling factors, see Fleurbaey and Zuber (2021).

4.2 The vNM Approach versus the Equivalence Approach

The two preference-based methodologies for measuring well-being that have been reviewed in this chapter are similar in many ways. But there is one key structural difference, which we have highlighted: the vNM approach insists on making well-being a function of an individual's lottery preferences, while the equivalence approach relies on ordinal preferences alone.

To see how the use of lottery rather than ordinal preferences drives a wedge between the two approaches, consider the following example. Assume that there are two lottery preferences, $\bar{\bar{R}}*$ and $\bar{\bar{R}}^{+}$, that rank attribute bundles the very same way. That is, the two lottery preferences share the very same ordinal preference (let's denote it as $\bar{R}'$). They have the very same indifference curves in the space of attribute bundles, but disagree in how they rank lotteries over attribute bundles. For each given attribute bundle a, there is the history (a, $\bar{\bar{R}}*$) and its counterpart history (a, $\bar{\bar{R}}^{+}$). The vNM approach distinguishes between the two histories. Further, and critically, the vNM approach *cannot* assign the very same well-being number to (a, $\bar{\bar{R}}*$) and (a, $\bar{\bar{R}}^{+}$) for every attribute bundle.[41] By contrast, the equivalence approach does not differentiate between (a, $\bar{\bar{R}}*$) and (a, $\bar{\bar{R}}^{+}$). As both lottery preferences share the same

[41] If we take vNM utility functions representing each preference, $u^{\bar{\bar{R}}*}(\cdot)$ and $u^{\bar{\bar{R}}^{+}}(\cdot)$, there is no way to rescale the two so that they assign the very same rescaled utilities to each and every attribute bundle. If there were scaling factors $s(u^{\bar{\bar{R}}*})$, $t(u^{\bar{\bar{R}}*})$, $s(u^{\bar{\bar{R}}^{+}})$, $t(u^{\bar{\bar{R}}^{+}})$ such that $s(u^{\bar{\bar{R}}*})u^{\bar{\bar{R}}*}(a)+t(u^{\bar{\bar{R}}*})=s(u^{\bar{\bar{R}}^{+}})u^{\bar{\bar{R}}^{+}}(a)+t(u^{\bar{\bar{R}}^{+}})$ for every a, the two utility functions would be positive affine transformations of each other—which is impossible if, as supposed, the two lottery preferences are different.

ordinal preference $\bar{R}'$, the two counterparts are represented as one and the same history, (a, $\bar{R}'$), and assigned the very same well-being number.

The vNM approach's conceptualization of histories is, at first blush, quite puzzling. An outcome (as explained earlier and in Chapter 2, sec. 2.2, this volume) is a simplified model of a possible world. A possible world includes a complete description of each person's characteristics. For short, call this a "complete possible life." There is no uncertainty in a complete possible life. For any given person, and any possible characteristic that this person might have, a complete possible life will specify either that the person has or that he lacks that characteristic. In one possible life, Lia gets a Ph.D. in economics; in another possible life, she does not earn a Ph.D. In no possible life for Lia is it uncertain whether she earns a Ph.D. or, if so, what the subject is (or, if so, how many pages the dissertation is, etc., etc.).

A history is a *model* of a possible life. And just as possible lives "squeeze out" all uncertainty, so do histories. Certain types of individual attributes are selected for description; each type of attribute is modelled as having different "attainments" (e.g., different amounts of income for the income attribute, different quality levels for the health attribute). An attribute bundle specifies the individual's attainment of each of the characterized types of attributes. There is no uncertainty regarding the attainments of the attribute bundles.

Advocates of the equivalence approach argue that, in such a risk-free setting, only ordinal preferences should matter. They think of a history as a combination of an attribute bundle and an *ordinal* preference.[42] Advocates of the vNM approach disagree and argue that a history is best seen as a combination of an attribute bundle and a *lottery* preference. Why is that?

We refer the reader to Adler (2016) for an elaborated defense and to Fleurbaey (2016) for a critique. In brief, the argument of Adler runs as follows. Although histories themselves contain no uncertainty about attribute bundles, a full theory of *well-being* cannot ignore uncertainty entirely. Individuals constantly make choices under uncertainty and wonder, what is the better choice in light of my interests? What is the more *prudent* choice? It would be problematic for a theory of well-being to fail to address such questions.

Thus a full theory of well-being will include a ranking not merely of histories, but also of history lotteries, argues Adler. A very plausible axiom for the ranking of history lotteries is

[42] Although the equivalent income measure as discussed in Section 3.3.3 does not take account of individuals' lottery preferences—it is discussed there as one version of the equivalence approach, in turn understood as taking account only of individuals' ordinal preferences—it is possible to extend the equivalent-income idea so as to take account of lottery preferences. Let l denote a lottery over attribute bundles. Fleurbaey (2016, p. 462-63) discusses the possibility of computing a certainty-equivalent income, i.e., the sure income that, combined with the reference value on the non-income attributes, would place the individual in a situation that she finds equally good as l. A more risk-averse individual would request a larger risk premium and, hence, obtains a lower certainty-equivalent income. While computing such certainty-equivalent incomes is a possibility that Fleurbaey describes, he does not in fact advocate their use in social assessment. (See *id*.)

what Adler terms “Welfare Stochastic Indifference.” If two lotteries assign the same probabilities to each of a series of well-being levels, then the lotteries are equally good for well-being. The core of Adler’s argument for why histories should specify lottery preferences is this: if a history merely specifies each individual’s ordinal preference, then it’s impossible for the well-being ranking of histories to satisfy both Welfare Stochastic Indifference and the axiom of Lottery Deference. See Adler, 2016, pp. 507-08.

Pragmatically, the vNM approach is more demanding than the equivalence approach, since we need to estimate lottery preferences. It remains to be seen whether using lottery rather than ordinal preferences makes a significant difference in policy assessment. In a head-to-head comparison of a vNM measure of well-being and an equivalent-income measure, as implemented in a sample of about 20,000 German subjects (the German Socio-Economic Panel of 2010), Decancq and Neumann (2016) find the two measures to be more similar to each other—in terms of the characteristics of individuals in the bottom well-being decile and individuals’ well-being ranks in the whole population—than to well-being as measured via income, an objective index, or subjective well-being. In Chapter 5 of this volume, Schokkaert and Decancq study how the question of whether there was social progress in Russia between 1995 and 2005 depends on which well-being measure is used. They introduce equivalent-income measures of well-being, vNM measures and (standard) expenditures as arguments in SWFs of the Atkinson and Kolm-Pollak family. In this analysis, Schokkaert and Decancq do find some substantial differences between the equivalent-income and vNM measures.

4.3 Homogeneous Preferences

Both the vNM approach and the equivalence approach are methodologies for interpersonal well-being comparisons that allow for heterogeneous preferences. The set $\mathbf{R}$ of preferences (ordinal preferences in the case of the equivalence approach, lottery preferences in the case of the vNM approach) may have multiple members. A central ambition of both frameworks is to allow for heterogeneous preferences and to do so in a manner that satisfies the Deference Principle.

However, SWF scholarship often simplifies matters by positing homogeneous preferences. For example, the classic optimal-tax model, deriving from Mirrlees, posits that everyone has the same preferences regarding consumption-leisure bundles. (Tuomala 2016, ch. 4).

The equivalence approach does not operate any differently with homogeneous preferences than it does with heterogeneous preference. $\mathbf{R}$ is, specifically, the set $\bar{\mathbf{R}}$ of ordinal preferences—which in the case of homogeneous preferences is a singleton set, containing a single ordinal preference $\bar{R}$. All histories are combinations of attribute bundles with this one preference $\bar{R}$.

Well-being numbers are assigned to those histories by numbering the indifference curves of $\bar{R}$, using either reference sets or a monotone path.

With homogeneous preferences, the vNM approach simplifies significantly (and in an intuitive way). Now, **R** is, specifically, the set $\bar{\bar{\mathbf{R}}}$ of lottery preferences--which in the case of homogeneous preferences contains a single lottery preference $\bar{\bar{R}}$. There is no need to choose scaling factors—their function is to commensurate heterogeneous lottery preferences—and $w(\cdot)$ can be set equal to any vNM utility function $u^{\bar{\bar{R}}}(\cdot)$ representing the single preference. That is, $w(a,\bar{\bar{R}}) = u^{\bar{\bar{R}}}(a)$.[43]

The intellectual divide between the two approaches also becomes smaller with homogeneous preferences. Recall that, with multiple lottery preferences, the vNM approach may assign different numbers to the very same indifference curve depending upon which lottery preference that curve is associated with. (See Section 3.4.2.) However, if there is but a single lottery preference , with a single corresponding ordinal preference $\bar{R}'$, then all attribute bundles on the same indifference curve of $\bar{R}'$ are now assigned the same well-being number—namely, the vNM utility level of all bundles on that curve according to $u^{\bar{\bar{R}}}(\cdot)$.[44] In effect, in the case of homogeneous preferences, the only difference between the equivalence and vNM approaches is *how* indifference curves are numbered. If the population is modelled as having a single lottery preference $\bar{\bar{R}}$ and a single corresponding ordinal preference $\bar{R}'$, the equivalence approach ignores $\bar{\bar{R}}$ and assigns numbers to the indifference curves of $\bar{R}'$ via reference sets or a monotone path. The vNM approach, instead, numbers indifference curves by applying $u^{\bar{\bar{R}}}(\cdot)$. In both cases, the well-being number assigned to a given attribute bundle *a* is just the number of the indifference curve of ordinal preference $\bar{R}'$ on which *a* lies.

[43] In this case, because vNM utility functions are unique up to a positive affine transformation, assigning well-being numbers to histories via the formula $w(a,\bar{\bar{R}}) = u^{\bar{\bar{R}}}(a)$ produces the same well-being level and difference comparisons between histories regardless of which vNM utility function representing $\bar{\bar{R}}$ is chosen.

[44] Recall that each attribute bundle *a* is equivalent to a "degenerate" lottery: a lottery that assigns this bundle probability 1. To say that *a* and *a** lie on the same indifference curve of $\bar{R}'$, the ordinal preference corresponding to $\bar{\bar{R}}$, is equivalent to saying: lottery preference $\bar{\bar{R}}$ is indifferent between a degenerate lottery with probability 1 to *a* and a degenerate lottery with probability 1 to *a**. If that is true, then—if $u^{\bar{\bar{R}}}(\cdot)$ is indeed a vNM utility function representing $\bar{\bar{R}}$—it must be the case that $u^{\bar{\bar{R}}}(a) = u^{\bar{\bar{R}}}(a^*)$.

5 Non-Preference-Based Well-Being Measures

5.1 Objective Goods and Capabilities

We have defined a well-being measure as *preference-based* if it operates on histories, of the form (a, R)—with R either an ordinal or lottery preference—and satisfies the Deference Principle. The view that well-being depends upon preferences is very widespread in economics, of course, and is also well represented in the philosophical literature on well-being. Our definition is an attempt to make precise (within our framework of measurement) this preferentialist tradition in the conceptualization of well-being.

But the dependency of well-being on preferences has been disputed. There is also a long tradition in philosophy, going back to Aristotle, of so-called *objective-good* accounts of well-being. Such accounts deny that well-being consists in preference satisfaction. Contemporary objective-good theorists posit lists of goods as the components of individual well-being. The goods (or at least some of them[45]) are preference-independent, in the sense that an individual can have more of the good without preferring it. For example, John Finnis argues that well-being consists in the attainment of the following goods: life, knowledge, play, aesthetic experience, sociability, practical reasonableness, religion. (Finnis 1988, ch.4). James Griffin sets forth a different list: accomplishment, "the components of human existence" (roughly, autonomy and physical integrity), understanding, enjoyment, deep personal relations. (Griffin 1996, pp. 29-30). Guy Fletcher suggests: achievement, friendship, happiness, pleasure, self-respect, virtue. (Fletcher 2016.) George Sher endorses this list of goods: moral goodness, rational activity, development of abilities, having children and being a good parent, knowledge, awareness of true beauty. (Sher 1997, p. 201). (See Alkire 2002 for a survey with more examples).

There are evident linkages between this corpus of philosophical scholarship and the contemporary literature on the *capability approach,* originating with work by Amartya Sen and Martha Nussbaum. In the capability approach, well-being is assessed in terms of individual "functionings" and "capabilities" (opportunities to function) that are not reducible either to preference satisfaction or to hedonic states (Sen, 1985).[46] Capability scholars often suggest leaving the identification of functionings and capabilities to democratic discussion (Sen, 2004). Martha Nussbaum, however, has famously defended a specific list of capabilities: life; bodily health; bodily integrity; the senses, imagination and thought; emotions; practical reason;

[45] If an account includes some goods that are preference-dependent while others are not, it's still possible for an individual to be better off (according to the account) with a bundle of goods that she disprefers. (Consider two bundles that differ only with respect to the preference-independent goods.) This is the hallmark of an objective-good account.

[46] See, however, Decancq et al. (2021) for a preferentialist interpretation of the capability approach and an application with Colombian data.

affiliation; other species; play; and control over one's environment (Nussbaum 2000, 2006). This list of capabilities looks very much like a philosophical list of objective goods.

Under the impetus of the capability approach (and objective-good accounts of well-being) several international institutions have developed and implemented so-called *composite indicators of well-being*, which aggregate different aspect of well-being by means of a mathematical procedure that often takes the form of a (weighted) average. The Human Development Index (HDI) of the United Nations Development Programme or the Better Life Index (BLI) of the Organisation for Economic Cooperation and Development (OECD) are popular examples.[47] While these composite indicators are mostly measured at the country (or regional) level based on average attainments, they can also be computed at the individual level, leaving room to introduce distributional or prioritarian concerns in the social evaluation. Building on a multidimensional version of the Atkinson SWF, Alkire and Foster (2011) and Decancq (2017) provide distribution-sensitive versions of the HDI and BLI, respectively. See also Decancq and Schokkaert, chapter 5, this volume.[48]

Multidimensional composite well-being indicators have a clear advantage over so-called "dashboards" of several (one-dimensional) policy indicators, such as the Millennium or Sustainable Development Goals.[49] These dashboards are popular in empirical applications of the capability approach, for instance. They display a policy's impact on well-being with respect to a series of indicators and, hence, leave it to the decisionmaker to balance the plural impacts so that an overall assessment can be made. The difficulty is that dashboards are insensitive to the correlation between the attainments. A pragmatic advantage, however, is that dashboards can be constructed using different dimension-specific data sets, which is obviously not possible for well-being measures that consider different attainments of the same individual.

Leaving aside data issues, one important challenge to constructing a composite indicator of well-being is to select an operational procedure to aggregate the different attainments and to determine the relative weights of these attainments (see Decancq and Lugo 2013 for a critical review). This challenge brings us back to the—fundamental—question of what information can be used when making such decisions and, in particular, what role preferences can or should play therein.

---

[47] The Human Development Index is published yearly in the Human Development Report by the United Nations Development Programme, see www.undp.org. The Better Life Index can be computed by the user on www.oecdbetterlifeindex.org. Decancq (2017) compares both indices.

[48] These, and other, proposals are built on the literature on the measurement of multidimensional inequality. Measures of multidimensional inequality summarize the joint distribution of multiple attainments in a single number. However, most multidimensional inequality measures do not explicitly measure each individual's well-being as a product of her multidimensional attainments. See Aaberge and Brandolini (2015), Chakravarty and Lugo (2016) or Weymark (2006) for surveys.

[49] For the millennium development Goals, see https://www.un.org/millenniumgoals and for the Sustainable Development Goals, see https://sdgs.un.org.

Consider the simplest case of an objective-good theory that posits a list of goods, each one of which is preference-independent. Such a theory can be captured in our formalism by supposing that the well-being level of histories is wholly independent of their component preferences: $w(a, R) = w(a^\star, R^\star)$ if $a = a^\star$. Two histories with the same attribute bundle lead to the same well-being level, irrespective of the preferences. Clearly, this position does not satisfy the Deference Principle. Interestingly, however, objective well-being measures that do not take account of preferences are the only ones that can satisfy the so-called dominance principle—which states that an attribute bundle with higher attainments for all attributes leads to higher well-being, irrespective of the preferences (see Fleurbaey and Trannoy, 2003).[50]

If individuals' preferences are not used to aggregate the different attainments and to construct a measure of well-being that allows for intra- and interpersonal comparisons of levels and differences, one may wonder which information can be used to determine the well-being function $w(\cdot)$. Unfortunately, there is no off-the-shelf theory that we can borrow from the objective-good literature or the capability approach to determine the shape of this function (at least not that we are aware of). In most applied work on composite indicators of well-being, one often works with simple well-being functions that are essentially (weighted) averages of the attainments (the HDI is an example of this approach) or, alternatively, one relies on complex statistical algorithms, such as principal component analysis (Ram, 1982) or Data Envelopment Analysis (Cherchye et al. 2007) to make intra- and interpersonal comparisons of well-being.

A different approach to handling the case under consideration flows most naturally from the earlier sections of this chapter.[51] A deliberator can have her own preference with respect to attribute bundles—her own sense of how different bundles compare with respect to well-being. That is, she has her own ordinal ranking of the set **A** of attribute bundles (let's denote this as $\bar{R}^d$) which she uses to evaluate well-being using some version of the equivalence approach—either via reference sets, or via a monotone path. The well-being number of a given attribute bundle a is just $w(a, \bar{R}^d)$. In other words, we locate each attribute bundle on the deliberator's indifference curves and then apply the equivalence approach. While the Deference Principle with respect to the individual's own preferences is not satisfied, the Deference Principle with respect to the deliberator's preferences is. The deliberator's preferences may or may not depend on the individual preferences of the individuals in the society, through a process of democratic decision. The BLI of the OECD may be interpreted as an—arguably crude—operational version of an approach based on the deliberator's preferences. On an interactive website[52], users can choose some parameters of the deliberator's preferences over the eleven attributes

[50] Decancq et al. (2015a) show how equivalent incomes can be obtained by combining respect for preferences with a weaker version of the dominance principle, restricted to a monotone path.

[51] We thank Paolo Piacquadio for raising this point.

[52] http://www.oecdbetterlifeindex.org/

considered. These preferences are then used to compare the average well-being of the OECD member states in a graphical way.[53]

A more complicated case arises if the objective-good account of well-being includes *some* preference-dependent goods. Although the account overall is still objective because it denies that a given person's well-being levels with different possible attribute bundles always hinges on the person's preferences over those bundles, the theory does not ignore preferences entirely. Note, for example, that James Griffin above includes the good of "accomplishment" on his list of goods, while Guy Fletcher includes the good of "achievement." A state of affairs is an accomplishment/achievement for some individual only if she prefers that state of affairs.

In our framework, an objective-good theory with some preference-dependent goods could be represented as follows. On the one hand, the Deference Principle is rejected: this is what makes the theory objective. It is not the case that $w(a, R) \geq w(a^*, R)$ iff $a\ R\ a^*$. On the other hand, preferences are *not* ignored entirely in assigning well-being numbers. It is *not* the case that $w(a, R) = w(a^*, R^*)$ if $a = a^*$. How to construct the well-being measure for *this* type of objective-good account is beyond the scope of what we can address here.

5.2 Subjective Well-Being

A large body of work in economics and psychology now examines individuals' subjective well-being (SWB)—the term that is used by this literature to refer to happiness, pain and pleasure, satisfaction with life, and related affective or cognitive mental states. SWB is typically assessed with surveys, which ask respondents to quantify some aspect of their SWB. For example, the respondent might be asked: "All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied." Or: "Taken all together, would you say that you are very happy, pretty happy, or not too happy?" (Graham 2016; Lucas 2016.)

An emerging policy-analytic strand within this scholarship uses SWB data for policy evaluation or the assessment of social condition (See, e.g., Layard 2005, Adler 2013). For example, SWB-based cost-benefit analysis has been theorized and implemented. (Fujiwara and Dolan 2016). Inequality metrics have been used to study the inequality of SWB (see, e.g. Dutta

---

[53] Alternatively, the deliberator might have her own lottery preference with respect to the set of attribute bundles, $\bar{\bar{R}}^d$, and assign a given attribute bundle a the well-being number $w(a, \bar{\bar{R}}^d)$ as per the vNM approach. Because the vNM approach in this application is ignoring all lottery preferences except $\bar{\bar{R}}^d$, the set of lottery preferences is a singleton and we have that $w(a, \bar{\bar{R}}^d) = u^{\bar{\bar{R}}^d}(a)$, with $u^{\bar{\bar{R}}^d}(\cdot)$ any vNM utility function representing $\bar{\bar{R}}^d$. See above Section 3.4.3.

and Foster 2013). Gross-national-happiness indicators (analogous to GDP) have been calculated. Indeed, many countries now collect SWB statistics regularly.

It is therefore worth considering how SWB data might be used to inform the well-being measure in our social welfare framework. We lack space to consider the issue in detail here, so will simply sketch the main possibilities.

First, an objective-good account of well-being might include certain SWB states on its list of objective goods. For example, we included above the lists of goods endorsed by the philosophers Guy Fletcher and James Griffin. Note that Fletcher's list includes the goods of happiness and pleasure, while Griffin's includes enjoyment.

At the limit, an objective-good theory might include *only* SWB states as goods. This is true, indeed, of the classic hedonic account of well-being adopted by Jeremy Bentham. Such a view construes pleasure and pain as positive and negative *sensations*; posits pleasure and pain as preference-independent goods (namely, that an individual's well-being is increased if she feels pleasure and decreased if she feels pain, independent of her preferring pleasure and dispreferring pain); and posits these hedonic goods as the only intrinsic components of well-being. (Sumner 1996, ch. 4.)

Second, a preference-based view of well-being will take the position that pain or pleasure, happiness, or some other SWB state—like every other aspect of some individual's life—is better for that person's well-being if and any only if she has a (laundered) preference for that SWB state. Leaving aside considerations of tractability, then, attribute bundles should describe individuals' SWB states along with non-SWB attributes such as income, health, and longevity. The preference-based well-being measures we described in detail above (the equivalence approach and the vNM approach), which take the form $w(\cdot) = w(a, R)$ can be extended without difficulty to the case in which a includes both SWB and non-SWB dimensions. To be sure, tractability considerations mean that a can only include some, not all, welfare-relevant characteristics. However, if individuals tend to have strong preferences regarding some SWB state, policy modelling with the SWF framework will be significantly more accurate if it includes that SWB state in the attribute bundle.

Third, there is a plausible linkage between a particular SWB state—life-satisfaction—and preferences. Life satisfaction has a cognitive component, namely judging that one's life is going well. Let us denote the life satisfaction reported by an individual i with attribute bundle a, as $SWB_i(a)$. Note that the life satisfaction function is individual-specific because life satisfaction scores are not only determined by the attribute bundle at hand, but also by other (idiosyncratic) scaling factors such as personality traits, cultural norms, expectations and aspirations. An optimistic person with a particular attribute bundle a will report a higher life satisfaction than a pessimistic person.

Decancq et al. (2015a) study the link between life satisfaction and ordinal preferences and define the so-called *consistency assumption*, which states that $SWB_i(a) \geq SWB_i(a^{\star})$ iff $a\ \bar{R}_i\ a^{\star}$.[54] Under this assumption, life satisfaction scores provide a well-being measure which satisfies the Deference Principle in *intrapersonal* comparisons. Yet, due to the presence of the individual-specific scaling factors, interpersonal comparisons of life satisfaction are not necessarily consistent with unanimously held preferences. Indeed, the optimistic individual Anne may report a higher life satisfaction than the pessimistic Bob, even when both Anne and Bob have the same ordinal preferences which agree that Bob's attribute bundle is preferred over Anne's. This example shows that raw life satisfaction scores should not be used as preferentialist well-being measures in our social welfare framework.

Building on the consistency assumption, however, a significant body of empirical work uses life-satisfaction surveys to estimate ordinal preferences or indifference curves (see Decancq et al. (2015a) and Decancq and Neumann (2016), for instance). In these estimations one aims to control econometrically for the effect of the individual-specific scaling factors on life satisfaction scores. When this method is used to estimate preferences so as to implement the equivalence approach, there is a substantial difference between the raw life satisfaction scores and the equivalence measure of well-being. Decancq et al. (2015a), for instance, find a significant divergence between life-satisfaction scores and equivalent incomes in identifying the worst-off individuals (in a nutshell, "the subjectively least satisfied individuals have larger expenditures, better health and nicer houses than those with the lowest value of equivalent incomes" [p. 1099]); while Decancq and Neumann (2016) finds a significant divergence between life satisfaction and both equivalent-income and vNM well-being measures.

In sum: there are a number of connections between SWB and well-being. But an SWB measure of well-being (even a life satisfaction measure) is not a preference-based measure, and those who endorse a preference account of well-being should resist using it this way.

6 Conclusion

Economists traditionally define well-being in terms of preference-satisfaction. It might seem that a preference account of well-being undercuts interpersonal well-being comparisons. In particular, consider well-being comparisons between individuals with different preferences. Imagine that Muriel prefers attribute bundle a to bundle a^*, while Nate has the opposite preference. We can't say unequivocally that either bundle is better than the other. Muriel is better off with a than a^*, while Nate is better off with a^* than a. An interpersonal comparison between the two based merely on their bundles would ignore their preferences; and so perhaps

[54] Testing whether the consistency assumption holds in real-life situations is an open avenue for further research. Benjamin et al. (2012) provide some cautionary evidence.

the lesson of this sort of case is that, if we want to respect preferences, we should give up on across-person comparability of well-being. Indeed, one line of thought within economics rejects interpersonal comparisons. (Adler 2019, pp. 76-79.)

However, rejecting interpersonal comparisons is not an option for those who propose to use a social welfare function (SWF) to evaluate governmental policy—in particular a prioritarian SWF, the focus of this volume. Such rejection also flies in the face of common sense. (See Adler, chapter 2, this volume.) For those who understand well-being in terms of preferences *and* accept interpersonal well-being comparisons—as we do—the lesson of the sort of case just described is that *both* preferences *and* attributes (not attributes alone) are the inputs into well-being. A well-being measure $w(\cdot)$ should be understood as operating on "histories," each history being a combination of an attribute bundle a and a preference structure R: $w(\cdot)$ takes the form $w(a, R)$. The Same History Principle then stipulates that if two individuals have the same history—the same attributes and the same preferences—they are equally well off.

The formalism of making an individual's well-being a function of her history (a, R) does not itself imply that the account of well-being is preference-based. This formalism *allows* for preferences to be followed in determining well-being, but doesn't *require* that they be. We could use the "history" formalism together with a well-being measure $w(\cdot)$ that ignores preferences.[55] What is distinctive about a *preference-based* well-being measure is that it satisfies the Deference Principle. The Deference Principle says: $w(a, R) \geq w(a^*, R)$ iff $a\ R\ a^*$ (preference R weakly prefers a to a^*). In other words, the well-being ranking of histories that incorporate different bundles but the same preference *tracks* that preference.

This chapter, although also covering well-being measures that violate the Deference Principle, has mainly focused on two approaches that satisfy it: the equivalence approach and the vNM approach. These approaches differ in significant respects; the first focuses on ordinal preferences, the second on lottery preferences. But the two approaches are alike in more fundamental regards: each respects the Same History Principle, each respects the Deference Principle, and each constructs a well-being measure that assigns a well-being number to every history—so that the well-being of two histories is commensurable even if the two contain different preferences.[56] Both approaches, in somewhat different ways, show how to develop an interpersonally comparable well-being measure suitable for the SWF framework while also respecting preferences.

[55] That is, $w(a, R) = w(a, R^*)$ for all a, R, R^*.

[56] For every (a, R) and (a^*, R^*), either $w(a, R) > w(a^*, R^*)$; $w(a, R) = w(a^*, R^*)$; or $w(a, R) < w(a^*, R^*)$.

References

Aaberge, R. and Brandolini, A. (2015). "Multidimensional poverty and inequality." In A.B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, vol. 2A. Amsterdam: Elsevier, pp. 141-216.

Adler, M.D. (2013). "Happiness surveys and public policy. What's the use?" *Duke Law Journal*, 62: 1509-1601.

Adler, M.D. (2016). "Extended preferences." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 476-517.

Adler, M.D. (2019). *Measuring Social Welfare*: *An Introduction*. New York: Oxford University Press.

Adler, M.D. (2020). "Extended preferences and the valuation of health." In N. Eyal et al. eds., *Measuring the Global Burden of Disease: Philosophical Dimensions*. New York: Oxford University Press, pp. 86-106.

Alkire, S. (2002). "Dimensions of human development." *World Development*, 30: 181-205.

Alkire, S. and Foster, J. (2011). "Counting and multidimensional poverty measurement." *Journal of Public Economics*, 95: 476-87.

Benjamin, D. J., Heffetz, O., Kimball, M.S. and Rees-Jones, A. (2012). "What do you think would make you happier? What do you think you would choose?" *American Economic Review*, 102: 2083-2110.

Blackorby, C. and Donaldson, D. (1988). "Money metric utility: A harmless normalization?" *Journal of Economic Theory*, 46: 120-29.

Bosmans, K., Decancq, K. and Ooghe, E. (2018). "Who's afraid of aggregating money metrics?" *Theoretical Economics*, 13: 467-84.

Botzen, W.J. Wouter, and van den Bergh, J.C.J.M. (2014). "Specifications of social welfare in economic studies of climate policy: Overview of criteria and related policy insights." *Environmental and Resource Economics*, 58: 1-33.

Chakravarty, S.R. and Lugo, M.A. (2016). "Multidimensional indicators of inequality and poverty." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 246-85.

Cherchye, L., Moesen, W., Rogge, N., and Van Puyenbroeck, T. (2007). "An introduction to 'benefit of the doubt' composite indicators." *Social Indicators Research*, 82: 111-45.

Deaton, A. (1979). "The distance function in consumer behaviour with applications to index numbers and optimal taxation." *The Review of Economic Studies*, 46: 391-405.

Deaton, A. (1980). "The measurement of welfare: theory and practical guidelines." Living Standards Measurement Study (LSMS) Working Paper No. 7, World Bank.

Deaton, A. and Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

Decancq, K. (2017). "Measuring multidimensional inequality in the OECD member countries with a distribution-sensitive better life index." *Social Indicators Research*, 131: 1057-86.

Decancq, K., Fleurbaey, M. and Maniquet, F. (2019). "Multidimensional poverty measurement with individual preferences." *Journal of Economic Inequality,* 17: 29-49.

Decancq, K., Fleurbaey, M. and Schokkaert, E. (2015a). "Happiness, equivalent incomes, and respect for individual preferences." *Economica*, 82: 1082-1106.

Decancq, K., Fleurbaey, M. and Schokkaert, E. (2015b). "Inequality, income, and well-being." In A.B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, vol. 2A. Amsterdam: Elsevier, pp. 67-140.

Decancq, K. and Lugo, M. A. (2013). "Weights in multidimensional indices of well-being: An overview." *Econometric Reviews*, 32: 7-34.

Decancq, K. and Neumann, D. (2016). "Does the choice of well-being measure matter empirically?" In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 553-87.

Decancq, K. and Nys, A. (2021). "Nonparametric well-being comparisons." *European Economic Review*

Decancq, K. and Schokkaert, E. "Beyond GDP: Using equivalent incomes to measure well-being in Europe." *Social Indicators Research*, 126: 21-55.

Decancq, K., Schokkaert, E. and Zuluaga, B. (2021). "Implementing the capability approach with respect for individual valuations." In E. Chiappero-Martinetti et al. eds., *The Cambridge Handbook of the Capability Approach*. Cambridge: Cambridge University Press, pp. 307-32.

Dutta, I., and Foster, J. (2013). "Inequality of happiness in the United States: 1972-2010." *Review of Income and Wealth*, 59: 393-415.

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

Fletcher, G. (2016). "Objective list theories." In G. Fletcher, ed., *The Routledge Handbook of Philosophy of Well-Being*. Milton Park, UK: Routledge, pp. 148-60.

Fleurbaey, M. (2016). "Equivalent income." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 453-75.

Fleurbaey, M. and Blanchet, D. (2013). *Beyond GDP: Measuring Welfare and Assessing Sustainability*. New York: Oxford University Press.

Fleurbaey, M. and Maniquet, F. (2011). *A Theory of Fairness and Social Welfare*. Cambridge: Cambridge University Press.

Fleurbaey, M. and Maniquet, F. (2017). "Fairness and well-being measurement." *Mathematical Social Sciences*, 90: 119-26.

Fleurbaey, M. and Maniquet, F. (2018). "Inequality-averse well-being measurement." *International Journal of Economic Theory*, 14: 35-50.

Fleurbaey, M. and Maniquet, F. (2019) "Well-being measurement with non-classical goods." *Economic Theory*, 68: 765-86.

Fleurbaey, M. and Schokkaert, E. (2013). "Behavioral welfare economics and redistribution." *American Economic Journal: Microeconomics*, 5: 180-205.

Fleurbaey, M. and Tadenuma, K. (2014). "Universal social orderings: An integrated theory of policy evaluation, inter-society comparisons, and interpersonal comparisons." *The Review of Economic Studies*, 81: 1071-1101.

Fleurbaey, M. and Trannoy, A. (2003). "The impossibility of a Paretian egalitarian." *Social Choice and Welfare*, 21: 243-63.

Fleurbaey, M. and S. Zuber. "Fair utilitarianism." (2021). *American Economic Journal: Microeconomics*

Fujiwara, D. and Dolan, P. (2016). "Happiness-based policy analysis." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 286-317.

Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.

Goodin, R. E. (1986). "Laundering preferences." In J. Elster and A. Hylland, eds., *Foundations of Social Choice Theory*. Cambridge: Cambridge University Press, pp. 75-101.

Graham, C. (2016). "Subjective Well-Being in Economics." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 424-50.

Griffin, J. (1996). *Value Judgement: Improving our Ethical Beliefs*. Oxford: Clarendon Press.

Gul, F. and Pesendorfer, W. (2008). "The case for mindless economics." In A. Caplin and A. Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*. New York: Oxford University Press, pp. 3-42.

Kreps, D.M. (2013). *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton, NJ: Princeton University Press.

Layard, R. (2005). *Happiness: Lessons from a New Science*. New York: Penguin Press.

Lucas, R.E. (2016). "Subjective Well-Being in Psychology." In M.D. Adler and M. Fleurbaey, eds., *The Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 403-23.

McKenzie, L. (1957). "Demand theory without a utility index." *The Review of Economic Studies*, 24: 185-89.

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

Nussbaum, M.C. (2006). *Frontiers of Justice*: *Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press.

Pazner, E. A. and Schmeidler, D. (1978). "Egalitarian equivalent allocations: A new concept of economic equity." *The Quarterly Journal of Economics*, 92: 671-87.

Ram, R. (1982). "Composite indices of physical quality of life, basic needs fulfillment and income: A 'principal component' representation." *Journal of Development Economics*, 11: 227-47.

Roberts, K. (1980). "Price-independent welfare prescriptions." *Journal of Public Economics*, 13: 277-97.

Samuelson, P.A. (1974). "Complementarity: an essay on the 40th anniversary of the Hicks-

Allen revolution in demand theory." *Journal of Economic Literature*, 12: 1255-89.

Samuelson, P.A. (1977). "Reaffirming the existence of 'reasonable' Bergson-Samuelson social welfare functions." *Economica*, 44: 81-88.

Sen, A. (1979). "The welfare basis of real income comparisons: A survey." *Journal of Economic Literature*, 17: 1-45.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland: Amsterdam.

Sen, A. (2004). "Capabilities, lists, and public reason: Continuing the conversation." *Feminist Economics*, 10: 77-80.

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

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

Tuomala, M. (2016). *Optimal Redistributive Taxation*. Oxford: Oxford University Press.

Weymark, J. A. (2006). "The normative approach to the measurement of multidimensional inequality." In F. Farina and E. Savaglio, eds., *Inequality and Economic Integration*. Routledge, London, pp. 303-28