The Sum of Well-Being

Jacob M. Nebel
Princeton University, USA
jnebel@princeton.edu

Is well-being the kind of thing that can be summed across individuals? This paper takes a measurement-theoretic approach to answering this question. To make sense of adding well-being, we would need to identify some natural 'concatenation' operation on the bearers of well-being that satisfies the axioms of extensive measurement and can therefore be represented by the arithmetic operation of addition. I explore various proposals along these lines, involving the concatenation of segments within lives over time, of entire lives led alongside one another or in sequence, and of evaluatively basic propositions via conjunction. All of these proposals turn out to carry highly controversial commitments about the good. I do not claim that these commitments are unacceptable. But they suggest that we cannot simply take for granted, as many philosophers do, that there is any such thing as the sum of well-being.

1. Introduction

A person's *well-being* is how good things are for that person. Philosophers often write as though people's well-beings can be summed or added together, so that there is not just your well-being, my well-being, and so on, but also our *total* well-being.

This practice is most commonly associated with the classical utilitarian view that we ought to do whatever would maximize the sum of well-being. But it is by no means exclusive to proponents of this or any other utilitarian doctrine. It is especially prevalent in population ethics (see, for example, McMahan 1981; Hurka 1983; Boonin-Vail 1996; Holtug 1999; Roberts 2002; Tännsjö 2002; Huemer 2008; Temkin 2012; Greaves 2017a; Otsuka 2018). In this literature, distributions of well-being are often represented by boxes or lists of numbers; the area of a box or sum of numbers is said to represent the total *amount* of well-being 'contained' in a population. Parfit (2017, pp. 153–4) says, for example, that 'very many people's lives might together contain a greater total sum of well-being, just as there might be a greater mass of milk in a vast heap of bottles that each contained only one drop'. Parfit and many others deny that this sum is all that matters. But they take for granted that it exists.

Mind, Vol. 132 . 528 . October 2023
https://doi.org/10.1093/mind/fzad019

I want to understand what, if anything, this means—whether well-being is even the kind of thing that can be summed across individuals, and if so, how to understand this operation. Some things don't seem to come in *amounts* that can be added together. There doesn't seem to be any such thing as the sum of the beauty of Michelangelo's *David* and the beauty of Kramer's *Fusili Jerry*, or the sum of the hardness of a diamond and the hardness of some talc, or the sum of the time at which I woke up this morning and the time at which I had my coffee. This is in contrast to intuitively additive properties like mass and length. The sum of the masses of two non-overlapping objects is just the mass of their mereological fusion (at least, in Newtonian mechanics; see McQueen 2015). What, if anything, is the analogous operation represented by addition in the case of well-being?

Measurement theory offers a way of making this question more precise. The idea of *extensive measurement* is to identify some way of 'concatenating', or combining, objects together in a way that can be represented by the arithmetic operation of addition. I motivate and explain this idea in §§2 and 3. I then explore various ways of applying extensive measurement to well-being in §§4–6. As we shall see, these proposals turn out to carry highly controversial commitments about the good. The acceptability of these commitments will depend on our substantive theory of well-being. Proponents of some theories will be happy to accept them, and thus can easily make sense of sums of well-being. They will be anathema to others. I myself am quite unsure whether they are worth accepting, so I cannot conclude either that there is or that there is not such a thing as the sum of well-being. What I want to show here is just that it is more difficult than we might have thought to make sense of adding well-being; we cannot simply take for granted, as many philosophers do, that such an operation makes sense. I conclude, in §7, by considering some possible implications of this difficulty.

2. Scale types and sums

What would it take to show that there is such a thing as the sum of well-being? According to Kagan (1998, p. 44), '[T]alk of adding up the total amount of well-being in a given outcome presupposes that it makes sense to talk about *measuring* a person's level of well-being—giving it a number that we can meaningfully compare to someone else's number—and then adding these numbers up'. It presupposes much more than this, however.

Suppose that well-being is only measurable on an *interpersonal ordinal* scale. Such a scale would tell us which things are better or worse for a person, and which people are better or worse off, but not *how much* better or worse things are for people. As Arrow (1951, p. 31) observes, it is possible (though implausible) to compare alternatives by summing the numbers assigned by arbitrarily chosen ordinal scales of well-being. But the sum of these numbers can hardly be taken to represent the *sum of well-being*. The number assigned by the Mohs scale of hardness to diamond is the sum of the numbers assigned to topaz and gypsum. But there is no sense in which the hardness of diamond is the 'sum of the hardnesses' of topaz and gypsum.

This doesn't mean that we can't add up the numbers assigned by an ordinal scale. We can add whatever numbers we like. Nor does it mean that the sum of these numbers is 'meaningless' in the sense of not representing anything at all. Trivially, given numerical scales of any kind, any operation on their values can be used to represent some relation or other (perhaps extensionally defined, as a set of ordered pairs); the question is not 'whether it means anything, but *what*' (Rozeboom 1966, p. 197).

Intuitively, well-being is measurable on more than an ordinal scale. There are not only facts about what's better or worse for you, but also facts about *how much* better or worse they are. Such information would be captured by an *interval scale* of well-being. An interval scale is one that, like the Celsius and Fahrenheit scales of temperature, represents ratios of differences in the attribute being measured. These scales are related by a *positive affine transformation*—that is, multiplication by a positive factor (changing the unit) and addition of a constant (changing the zero)—which preserves all ratios of differences.

There are two kinds of interpersonal comparisons that might be captured by interval scales of well-being. There are comparisons of *differences*—that is, gains and losses—in well-being, and there are comparisons of well-being *levels* themselves—that is, how well off each person is. Interval scales with interpersonally comparable differences allow us to make ratio comparisons of different people's gains and losses in well-being—to say, for example, that w is better for Ann than x by twice as much as y is worse for Bob than z. This would allow us to add people's gains and losses in well-being; for example, if Cat's difference in well-being is twice Ann's and Bob's, then Cat's difference in well-being is the sum of Ann's and Bob's. But if we can *only* make interpersonal comparisons of differences, then we cannot add these people's well-beings themselves, since they would not even be comparable.

Interval scales that represent interpersonal comparisons of well-being *levels* can be used to make ratio comparisons (and therefore sums) of differences between how good things are for distinct individuals. They can therefore be can be used to define the *average* well-being in a population—that is, as the level from which the sum of differences in well-being is zero (Nebel 2023, §5.1). But we cannot 'multiply' this average by the number of people to deliver the sum of people's well-beings, since interval scales don't represent ratios of well-being levels. More generally, the sum of numbers on an interval scale does not represent the sum of what those numbers represent. If each of two systems has a temperature of 2° C, there is no sense in which the 'sum of their temperatures' is 4° C. This is not because the sum of numbers on the Celsius scale represents nothing at all: it represents the sum of *differences* in temperature from the freezing point of water. But this is not the sum of their temperatures. Attributes that are only interval-scale measurable are like points along a straight line with no origin: we can add the distances between points, but not the points themselves. Sums require a *ratio scale*, such as the gram scale of mass and the metre scale of length, which have a natural zero: an object that has zero grams of mass has zero mass.

It is sometimes claimed that, if people's well-beings are measurable on interval scales with interpersonally comparable differences, then the sum of numbers on those scales represents the sum of their well-beings (Broome 1991, p. 218). The rationale for this view is that, so long as the population is fixed, the ranking of alternatives by the sum of these numbers will be invariant to admissible transformations of the scales—positive affine transformations with a common scale factor across individuals—which preserve all ratios between people's gains and losses in well-being. But it does not follow from this that the sum of these numbers represents a *sum of well-being*. By way of analogy, if we represent people's heights using positive numbers on any scale of height, the ranking of individuals by the logarithms of those numbers will be invariant to admissible transformations of our scale—converting from inches to metres, for instance—but it does not follow that there is any such thing as the logarithm of your height. What follows is only that this ordering represents *some* relation regardless of which scale of height we use: namely, the ordering of people's heights. Similarly, the ordering of alternatives by adding a fixed population's interval-scale measures of well-being represents *something* independent of the scale: namely, the sum of people's gains and losses in well-being between any two alternatives. But the sum of people's differences in well-being is not the sum of their well-beings, any more than the sum of systems' differences in temperature is the sum of their temperatures.

In any case, an interval scale is clearly not sufficient to compare sums of well-being between populations of different sizes (Arrhenius 2000, p. 37). To figure out whether adding some person to the population would increase or decrease the sum of well-being, we would need to know whether her well-being would be 'positive' or 'negative'. This requires a non-arbitrary zero level of well-being, which is not captured by an interval scale. And I am looking for a way of making sense of sums of well-being that allows us to compare such sums between different-sized populations, since it is only in such cases that that one might claim to care specifically about the *sum* of well-being, as opposed to other operations, such as the average or sum of differences, that yield the same ordering of alternatives in fixed-population cases.

How can we construct a ratio scale of well-being? Some suggest that, given an interval scale, all we need to do is identify some particular level to call zero (Broome 2004, p. 254). Lives above that level are then said to have positive welfare; below it, negative welfare. Adler (2011) calls this strategy *zeroing out*. Different possible zero levels have been proposed in the literature, including the value of a life whose existence makes the world neither better nor worse (Dasgupta 1988), the value of a life that never gets better or worse over time (Broome 2004), the value to which all lives converge as they get arbitrarily short (Blackorby, Bossert and Donaldson 2005), the value of never existing (Adler 2011), and the value of the worst possible life in the outcomes under consideration (Adler and Treich 2015).

It is not enough, however, to identify *a* zero level. To get a ratio scale of well-being, we need to identify *the* zero level. If the choice of zero level were arbitrary—as Broome (2007, p. 120) himself suggests—then we could not take the ratios between numbers to be ratios of well-being, or compare sums of those numbers to determine which outcomes contain a greater sum of well-being, since we would get different results if we chose a different level to call zero. Without any reason to believe that any particular level is *the* zero level of well-being, this strategy can only be claimed to deliver a ratio scale of *differences in well-being from the chosen level*, not of well-being. By way of analogy, we could, if we wanted, insist that any scale of temperature assign the number zero to the freezing point of water. But this would not entitle us to interpret ratios or sums of degrees Celsius as ratios or sums of temperatures; they would still just be ratios or sums of temperature differences from the freezing point of water. Merely calling some temperature zero doesn't make it *zero temperature*. The same is true for well-being.

My point is not that any choice of zero level *would* be arbitrary. My point is rather that we'd need a way to *show* that lives at one's chosen zero level actually have zero well-being. While the various candidates mentioned above may be relatively natural choices, each of them rests on contentious presuppositions, and it's not obvious that all of them are extensionally equivalent.[1] Of course, this doesn't mean that there is no zero level of well-being, and it leaves open the possibility that zeroing out may be adequate or useful for purposes other than defining sums of well-being (as I consider in note 7 below). But it would seem desirable to have a more general strategy for *deriving* the zero level, in a way that explains why lives at that level, and not any other, have zero well-being, and what, more fundamentally, having 'zero well-being' even means.

Extensive measurement is just such a strategy. It is the standard approach in measurement theory for constructing a ratio scale.[2] Rather than starting with an independently constructed interval scale and declaring some level zero by fiat, the axioms of extensive measurement yield a ratio scale directly, in a way that tells us what the arithmetic operation of addition represents. This will be explained in §3.

3. Extensive structure

How can we show that it is *meaningful* to add up one person's well-being and another's? We would need to explain what is meant by 'adding up'. We cannot simply mean the arithmetic operation +, since this operation is defined on real numbers, and no one's well-being is a real number

[1] Take, for example, Adler's (2011, p. 219) suggestion involving the value of non-existence. First, it's highly controversial whether it even makes sense to compare lives with non-existence with respect to well-being (Dasgupta 1988; Broome 1993; Bykvist 2007; see Arrhenius and Rabinowicz 2015 for an influential response). Second, even if such comparisons make sense, other seemingly natural candidates could in principle come apart from this level. For instance, Fleurbaey and Voorhoeve (2015) argue that it can be better for a person to exist than not to exist, but that the level at which a person's addition to the population is a matter of indifference—that is, Dasgupta's zero level—may be above the value of non-existence. And, when some of our choices would bring about lives that are worse than non-existence, Adler's (2011) zero level comes apart from that of the worst feasible life—that is, Adler and Treich's (2015) zero level.

[2] Other possibilities for ratio-scale measurement have been devised by Luce and Narens (1985). The non-extensive structures they investigate, however, still involve concatenation operations; they are just not represented by addition. And, as Laming (1997, p. 91) observes, these structures 'constitute a family of merely mathematical possibilities with, as yet, no natural applications'. So I leave the exploration of their potential application to well-being for another occasion.

(Nebel 2021). We would need to identify some non-numerical operation that is, in a certain sense, well represented by the arithmetic operation of addition.

In measurement theory, this is called a *concatenation* operation. Such operations exist for paradigmatically 'extensive' quantities such as length and mass. There is a natural operation for combining objects in such a way that the length or mass of the combination is related to the lengths or masses of the objects combined in an intuitively addition-like way. The 'intuitively addition-like' manner of this operation will be spelled out below.

The stock example of an extensive structure in measurement theory involves the measurement of length via concatenation of rods (Krantz et al. 1971). We have a set of rods $A = \{a, b, c, \ldots\}$. There is a relation $\succcurlyeq$ defined on this set, with the interpretation that $a \succcurlyeq b$ if and only if rod a is at least as long as rod b. a and b are equally long ($a \sim b$) if and only if a is at least as long as b and b is at least as long as a. a is longer than b ($a \succ b$) if and only if a is at least as long as b and not vice versa. There is a binary operation $\circ$ on A—that is, a function from $A \times A$ to A—which takes any two rods and returns a third, which we can think of as the rod obtained by stacking the first two together from end to end. So our set of rods is closed under concatenation. We will write as though any rod a can be concatenated with itself to yield $a \circ a = 2a$. Since no one rod can be in two places at once, we can interpret this as the concatenation of two perfect copies (qualitative duplicates) of a. For any natural number n, 'na' denotes the concatenation of n perfect copies of a. (Formally, define na inductively as follows: $1a = a$, and for any $n > 1$, $na = (n-1)a \circ a$.)

The triple $(A, \succcurlyeq, \circ)$ is an *extensive structure* just in case the following conditions are satisfied for any rods in A. First, the at-least-as-long-as relation is transitive:

Transitivity. If $a \succcurlyeq b$ and $b \succcurlyeq c$, then $a \succcurlyeq c$.

Second, the relation is complete—there are no 'gaps', or rods that are incomparable in length:

Completeness. Either $a \succcurlyeq b$ or $b \succcurlyeq a$.

Third, the concatenation of some rod with the concatenation of two others is just as long as what we get by stacking together the concatenation of the first two rods with the third:

Weak Associativity. $a \circ (b \circ c) \sim (a \circ b) \circ c$

Fourth, the ordering of concatenated rods is independent of parts that they have in common; the common parts 'cancel out'. This means that one rod is as long as another if and only if the concatenation of the first rod with any third is as long as the concatenation of the second with that common third, taken in either order of concatenation:

Independence. $a \succcurlyeq b$ if and only if $a \circ c \succcurlyeq b \circ c$ if and only if $c \circ a \succcurlyeq c \circ b$.

Fifth, no rods are 'infinitely' longer or shorter than others. This possibility is ruled out by the following property:

Archimedean. If $a \succ b$, then for any $c, d \in A$, there is some natural number n such that $na \circ c \succcurlyeq nb \circ d$.

To see how the Archimedean axiom has its intended effect, suppose that d was infinitely longer than c, so that no number of c-rods could be concatenated to exceed the length of d. Then take any a and b where a is slightly longer than b. The Archimedean axiom implies that the concatenation of sufficiently many copies of a with c should be at least as long as the concatenation of that many copies of b with d. But then d cannot be infinitely longer than c after all, since their difference in length is at most n times the difference in length between a and b, for some sufficiently large n.

These five axioms are necessary and sufficient for the existence of a function u which assigns a real number to each rod in A in such a way that higher numbers are assigned to longer rods and the number assigned to the concatenation of any two rods is the sum of the numbers assigned to the two rods so concatenated:

(i) $a \succcurlyeq b$ if and only if $u(a) \geq u(b)$
(ii) $u(a \circ b) = u(a) + u(b)$

This second condition is what makes it meaningful to add lengths together: the sum of the lengths of two rods is the length of the rod obtained by concatenating them. In this sense, the qualitative operation of concatenation is represented by the arithmetic operation of addition.

Another function u' satisfies the two properties above just in case u' is a *similarity transformation* of u: for some positive k, $u'(x) = ku(x)$ for every x in A. This is the characteristic uniqueness condition of a ratio scale. The metre scale and the centimetre scale are equally good scales of length, and while they assign different numbers to objects, the ratios between numbers are preserved on both scales. The meaning of these ratios can be spelled out in terms of the concatenation operation: for example, na is n times longer than a, and is as long as n perfect copies of a stacked together from end to end.

Elements in an extensive structure can be classified as either positive, negative, or null. An element a is positive if and only if, for any element b, $a \circ b \succ b$, negative if and only if $a \circ b \prec b$, null if and only if $a \circ b \sim b$. In the case of length, all elements happen to be positive (unless we expand the structure to include points with zero extension). But the axioms of extensive measurement do not require this, and the structures I consider for well-being will feature negative and null elements.

An additive representation of an extensive structure will assign the number zero to all and only null elements. In contrast to the zeroing out strategy, the zero level is not imposed by fiat or picked as one possible zero among others. It is derived from the structure that the scale represents, and it is completely unique. Of course, if well-being is susceptible to extensive measurement, then this zero level could in principle have been identified independently, and used on behalf of the zeroing out strategy. But it would only be through extensive measurement that we would be justified in claiming that lives at that level actually have zero well-being, and that we know what this means. If it turns out, on the other hand, that the axioms of extensive measurement cannot be successfully applied to well-being, then I see no reason to assume that there *is* some one true zero level of well-being, as sums of well-being require.

With the notion of an extensive structure on the table, we can sharpen our question as follows: is there a natural concatenation operation on the bearers of well-being that satisfies the axioms of an extensive structure?[3]

I have posed our question in terms of the *bearers* of well-being—the things that are better or worse for people. Some might prefer a different approach, on which the things concatenated are degrees or magnitudes of well-being rather than objects that instantiate those degrees or magnitudes (Bykvist 2021).

To assess this approach, we would need to know more about this concatenation operation on degrees of well-being. I do not find myself to have any pre-theoretical grip on such an operation.

[3] The restriction to *natural* concatenation operations comes from Hempel (1952, p. 76), who suggests that without it, any quantity could be classified as extensive. For example, given any unbounded numerical scale u and a sufficiently rich set of objects, we can gerrymander an operation $\circ$ and relation $\succcurlyeq$ that satisfies the axioms, by defining $a \circ b$ to be some object for which $u(a \circ b) = u(a) + u(b)$ and $a \succcurlyeq b$ if and only if $u(a) \geq u(b)$. So we need some constraint to avoid making our task trivial. Some notion of naturalness seems to play a role in determining the meaning of expressions like 'addition' anyway (D. Lewis 1983). But I will otherwise leave the constraint of naturalness vague and uninterpreted.

Perhaps (as Bykvist suggests) we can try to understand it by attending to systematic connections between the envisaged operation and the concatenation of well-being bearers, just as the operation of adding mass magnitudes relates in a systematic way to the operation of mereological fusion on concrete objects. But then we would still need to be assured that there is some such operation on the bearers of well-being, even if the principles that govern that operation hold in virtue of more fundamental principles about the operation on abstract degrees of well-being. By contrast, if we first apply the axioms of extensive measurement to the bearers of well-being, we can *define* a concatenation operation on degrees of well-being—understood either as equivalence classes of well-being bearers or as abstract entities of some other kind—rather than taking such an operation as primitive: where $\mathbf{a}$ and $\mathbf{b}$ are the degrees instantiated by objects a and b respectively, define the concatenation of $\mathbf{a}$ and $\mathbf{b}$ to be the degree instantiated by $a \circ b$. This can be interpreted as the *sum* of $\mathbf{a}$ and $\mathbf{b}$ if the original structure, involving the concatenation of well-being bearers, satisfies the axioms of extensive measurement. (This approach does not require giving any *metaphysical* priority to the concatenation of well-being bearers.)

The next three sections will consider three possible ways in which well-being might have extensive structure.

4. Concatenating life-segments

The first proposal involves the concatenation of segments within a life. It generalizes a strategy used by Skyrms and Narens (2019), in turn based on Kahneman, Wakker and Sarin (1997), for extensive measurement of a single individual's pleasure and pain.

We start with a set of possible lives. Divide all of these lives up into segments of arbitrary finite duration. For every length of time and every possible life of at least that length, there will be a segment from that life—perhaps infinitely many—lasting that length of time. A is the set of all life-segments from our set of possible lives.

Our relation $\succcurlyeq$ on A will be an interpersonal betterness relation: for any life-segments a and b, $a \succcurlyeq b$ just in case it is at least as good for a person to live segment a as it is for a person to live segment b. To bracket worries about interpersonal comparisons, we assume that any given life-segment is just as good for any one person as it is for any other. This assumption could, of course, be rejected. We'll relax it in §6. For now, it will simplify things to keep this assumption.

Life-segments can be combined to yield composite life-segments. For example, the first five minutes of your life can be combined with the next ten minutes to yield a segment that is the first fifteen minutes of your life. And all the segments of your life can be combined to yield a segment that is your entire life. Perhaps some life-segments cannot be directly combined with others. For example, we might doubt whether a single life could contain a segment that starts with the last few moments of Joan of Arc's life and continues with an arbitrary snippet of Jerry Garcia's childhood. But perhaps there is a pair of life-segments that *can* be combined directly, the first of which is just as good for a person as the final moments of Joan's life and the second of which is just as good for a person as the snippet of Jerry's childhood. We can define the concatenation operation $\circ$ in terms of these surrogates: $a \circ b$ is some composite life-segment that starts with a segment just as good for a person as a (perhaps a itself) and continues with one just as good for a person as b (perhaps b itself). This allows us to think of our set A as closed under concatenation.

If our structure $(A, \succcurlyeq, \circ)$ satisfies the axioms of extensive measurement, then the value of any life can be represented as the sum of the values assigned to the segments that make it up. And the sum of well-being across lives is obtained by concatenating equivalent segments into a longer life. The sum of the value of your life and the value of mine is just the value of a life that starts with a segment just as good as your entire life and continues with a segment just as good as my entire life. As Rawls (1971, p. 24) might put it, all lives are 'fused into one'. The zero level assigned by this method will be the value of a segment that, when concatenated to any other, makes it neither better nor worse. This results in the same lifetime zero level proposed by Broome (2004).

So, does this structure satisfy the axioms?

I think it does satisfy Transitivity, Completeness, and Weak Associativity. I cannot see any reason to doubt Weak Associativity in this context. The transitivity of *at least as good for a person* is questioned by Rachels (1998) and Temkin (1996); for critiques of their arguments, see Huemer (2013), Voorhoeve (2013), Pummer (2017), and Nebel (2018). Completeness is much more controversial. Many people believe that things can be 'incommensurable' in value: neither is better than the other, nor are they equally good (Raz 1985; Chang 2002). This is a plausible view, but I believe it is mistaken, and that Completeness is true (Dorr, Nebel and Zuehl 2023; see also Broome 1997 and Regan 1997). It would take us too far afield to address the worries regarding Completeness, and I prefer to focus here on the axioms that (unlike Transitivity and Completeness) are specific to extensive measurement. But even if Completeness were false, it might still be possible to have a kind of extensive measurement of well-being, involving vectors of numbers rather than single real numbers (Carlson 2008).

Independence and the Archimedean axiom are much less plausible. Against Independence, there are familiar concerns regarding the 'shape' of a life that are ruled out by this axiom (Kamm 1988; Velleman 1991). We might think that a life that gets better over time is better than one that gets worse over time, even if those lives could be decomposed into segments of equal value and rearranged. This is ruled out by the right-hand side of Independence, which requires the order in which life-segments are lived to be irrelevant. Or consider projects and relationships over time. A life that starts with a segment in which one wants to become an astronaut and works tirelessly towards that goal and continues with a segment in which one achieves that goal and takes pleasure in that achievement may be better than one in which the first of those segments is replaced by one that is just as good but without the goal or toil towards becoming an astronaut. Or we might care about other relations between life-segments, such as learning from one's past mistakes, or enjoying a diverse array of goods over time. Or we might care especially about how good a life is at its best moment, or at its end, or think that how good it is to extend a life with some segment depends on how good it has been so far on average (for discussion of these issues, see Broome 2004, §15.3). All of these are ruled out by Independence, which requires us to care only about how good each life-segment is on its own, not about how those segments hang together.

The main problem for the Archimedean axiom is that it entails a single-life version of the *Repugnant Conclusion* (McTaggart 1927, pp. 452–3; Parfit 1986): for any life a, no matter how long and wonderful, and any barely worthwhile life-segment z, there must be some number of z-segments whose concatenation would be better than a. There are ways of generalizing extensive measurement, without the Archimedean axiom, to accommodate this sort of example—again, involving vectors of numbers rather than single real numbers (Carlson 2010). However, there would still be a problem for Independence. For suppose that we can decompose a into equally good segments $a_1, \ldots, a_n$ of arbitrarily small finite duration. Then, if a is better then any concatenation of z-segments, Independence implausibly implies (given Transitivity and

Completeness) that each of $a_1, \ldots, a_n$ must be better than any concatenation of z-segments (Jensen 2008). This seems implausible.[4]

There are, of course, responses available to these objections. If you do not find these concerns compelling, then you may happily accept the life-segments account. But I am sufficiently bothered by these worries that I would prefer another account, if there is one, that avoids them. Even if the apparent violations of Independence can be explained away, it is not clear why the additivity of well-being across individuals should depend on the (highly controversial) additivity of a single individual's well-being over time. Perhaps it does, if the life-segments approach turns out to be the best way of making sense of adding well-being. But that would be a surprising and interesting consequence of our investigation.

5. Concatenating lives

One desideratum suggested by our discussion of the life-segments account is this: we want an interpretation of adding well-being that is compatible with the unity of an individual life. Even though a life has temporal parts that can themselves be good or bad for a person, its value is not reducible to those parts (Bader MS). In this section, I want to explore the possibility of concatenating entire lives, rather than the segments that make them up. If successful, this would give us a way to make sense of adding well-being across lives without requiring intertemporal additivity within lives.

5.1. Populations

The most obvious way to concatenate lives is by imagining populations in which those lives are lived alongside one another. Consider some good life. Intuitively, we might think, a population that contains two lives of that quality contains more well-being than a population containing only one—perhaps, indeed, exactly twice as much. That is the idea behind the present proposal.

A population is a set of lives; we confine our attention to finite populations. Disjoint populations, which share no members, can be concatenated via set union: we take the population that contains all the lives in either population. But we also want to concatenate populations that overlap. To do this, we consider surrogates of these populations that do not overlap, and take their union instead. More precisely, say that

[4] If we give up Completeness, what follows instead is that none of the a_i can be worse than any concatenation of z-segments (Nebel 2022a).

two populations are *welfare-equivalent* just in case there is a one-to-one correspondence between them that maps each life in one population to an equally good life in the other population. Intuitively, welfare-equivalent populations contain the same number of lives of any given quality. The thought is that these welfare-equivalent populations are interchangeable with respect to how much well-being they contain. Our concatenation operation then takes any two populations a and b and returns the union of two disjoint populations a' and b', where a' is welfare-equivalent to a and b' is welfare-equivalent to b. So we can treat our set of populations as closed under concatenation.

The question is: what is the relevant relation over these populations? We need to know the property of these populations that is supposed to grow with concatenation. I see three natural possibilities.

The first is the *goodness* of these populations. We can define $a \succcurlyeq b$ to mean that population a is at least as good as population b. This relation will satisfy the extensive structure axioms just in case the goodness of a population can be represented as the sum of the goodness of the lives it contains. A person's well-being will then be understood as the contribution their life makes to the goodness of the overall population, and the zero level will be the value of a life whose addition to any population is a matter of indifference—that is, Dasgupta's (1988) zero level. This strategy, however, ties the meaningfulness of adding well-being to *totalist population axiology*, the view that one population is better than another if and only if it contains a greater sum of well-being. Totalism is highly controversial. Many who reject it, though, still claim to care about the sum of well-being. They cannot accept the extensive structure axioms applied to this relation. For example, egalitarians would reject Independence: one population might be better than another, but this ordering can be reversed when both are concatenated with some common third population, since that concatenation can introduce some inequality where there was none (or less) before. And many egalitarians claim to care about both equality and total well-being. They and others who reject totalism need some other way to make sense of adding well-being.

The second property we might consider is how much well-being these populations contain. We might define $a \succcurlyeq b$ to mean that a contains at least as much well-being as b. This strategy seems more appropriately neutral as regards the correct population axiology. It seems, however, very close to taking the idea of sums of well-being as primitive. If we do not already believe that there is such a thing as a sum of well-being, we may not be willing to grant that there are *amounts* of well-being that populations contain (see, for example, Bennett 1978 and

Larsson 2006). To see this, consider an analogous proposal regarding beauty. Suppose we are asked to consider whether one set of objects contains at least as much beauty as another. It is not clear that there is some stuff, *beauty*, such that in addition to objects instantiating various degrees of beauty and there being differences in how beautiful various objects are, arbitrary sets of objects carry more or less of this stuff, depending on how many beautiful objects they contain (Taurek 2021, p. 313). Similarly, there doesn't seem to be a relation of *containing at least as much density* that a set of sufficiently many feathers stands in to a set containing an anvil, or a *contains at least as much hardness as* relation that enough talc bears to a diamond. Again, if we are quite sure that there is such a thing as a sum of well-being, then perhaps this is the deepest we can go. But I would prefer an account, if there is one, that explains *why* there is a greater amount of well-being in a population that contains more good lives.

The third property is how good each set of lives is *for* the people who live those lives. We might define $a \succcurlyeq b$ to mean that a is at least as good for those who live the a-lives as b is for those who live the b-lives. Such interpopulational comparisons seem to make sense. But this relation does not satisfy the axioms of extensive measurement. Suppose that every life in a is better than every life in b. Then it would be very natural to think that a is better for the a-people than b is for the b-people. But if the extensive structure axioms were satisfied with respect to the relation in question, then if b contained sufficiently many lives, the sum of the numbers assigned to the b-lives would exceed the sum of the numbers assigned to the a-lives. So the relation in question cannot be represented as maximizing the sum of the numbers assigned to concatenated populations. Goodness *for people* does not seem to be an extensive quantity that always grows with the number of people for whom things are good.

The basic problem with this method of concatenating lives is that a person's well-being is supposed to be *goodness for that person*. The concatenation operation we have been considering takes lives, which are good for the particular people who live them, and combines them into populations, the goodness of which does not seem to belong to some particular person. That is why we had to hunt for some alternative property of interest. Is there instead a way of concatenating lives in such a way that the resulting goodness is still owned by someone?

5.2. *Life-sequences*

Rather than concatenating lives side by side, as members of a larger population, we could imagine the concatenation of lives back to back, as part of a sequence of lives. Suppose that a person could live multiple lives, one after another, without remembering previous lives or anticipating future ones. Then we might treat the concatenation of two lives to be the *life-sequence* that starts with the first and continues with the second. For example, you might first live Britney Spears's life, followed by Napoleon's.

Our set of objects A will be a set of life-sequences (including degenerate ones containing just a single life). The operation $\circ$ will concatenate life-sequences to yield compound sequences. Our relation $\succcurlyeq$ will track the goodness of these life-sequences for a person: $a \succcurlyeq b$ if and only if it is at least as good for a person to live sequence a as it is for a person to live sequence b. (I assume, as in the life-segments model, that any given life-sequence has the same value for any given person.)

If this structure satisfies the axioms of extensive measurement, then the value of a life-sequence can be represented as the sum of the values of the lives that make it up. And we can treat the sum of well-being in a world as the value of a life-sequence made up of all the lives in that world. To determine which of various populations contains more total well-being, we ask something very close to a question posed by C. I. Lewis (1946, p. 547): '[W]hich of these two objects would you prefer if the experience of all these persons were to be your own; as, for example, if you had to live the lives of each of them *seriatim*?'

Unlike the life-segments idea, this proposal does not require well-being to be additive within lives. And, unlike the population-level proposal just considered, it allows us to focus on the property we are interested in: goodness for a person. But this proposal has problems of its own.

First, we might worry about the metaphysical possibility of this concatenation operation, and thus about the existence of the life-sequences to be considered, as well as the naturalness of the operation. For example, if your survival at some time requires there to be someone psychologically continuous with you, then you could not live some other life after dying without memory of your present life. Or, if you are essentially a human organism, then we might think that a life is the kind of thing you can only have one of. I do not myself have some theory of personal identity that implies the metaphysical impossibility of these life-sequences. But it seems unfortunate to tie the meaningfulness of adding well-being to the truth of a controversial theory of personal identity.

Second, we might worry about the desirability of eternal (or not-so-eternal) recurrence. The life-sequences model implies that if a life is good, then it would be better for a person to live that life any number of times in a row: the more, the better. But it seems reasonable to judge that one's own life is good while nonetheless preferring not to live it over and over again. At the very least, one might prefer to live a sequence of lives that are very different from one's present life but just as good as it. That is ruled out by Independence.

Third, we might care about the 'shape' of a life-sequence. This concern may seem less compelling when we do not retain memory between lives. But it might seem especially tragic for a person to live their entire life working for some goal—say, the advancement of justice in their society—and then be reincarnated as a despicable person who ruins all of their previous life's work. This seems worse than a sequence in which one first brings about great injustice and then lives a life dedicated to bringing about reparations for that injustice, even if the individual lives are just as good as those in the first sequence.

Finally, there is a more general problem, which applies to all the views considered so far. It is not obvious that the goodness of a life depends only on its intrinsic properties. Perhaps it depends on what happens after it (for instance, whether one is defamed by one's descendants), before it (for instance, one's role in an intergenerational project, such as philosophy), or to the lives lived alongside it (for instance, loving relationships with or beneficent activities towards others). These features, however, are ruled out by all the views considered so far, which assume that the value of a life stays the same regardless of which other lives are lived before, after, or alongside it.

To avoid this more general problem, we might try to consider objects whose value for a person is more clearly intrinsic. I turn to such a proposal in §6.

6. Concatenating propositions

According to an influential tradition in the theory of value, the bearers of intrinsic value are the things expressed by *that*-clauses (for example, *that I am happy*). I take these things to be propositions, which for simplicity I will understand as sets of possible worlds. (We could just as well use a more fine-grained conception of propositions, or work with other entities, such as states of affairs.)

We can think of a theory of welfare as identifying, for any given person, certain propositions that are good or bad for a person in the most basic way. For example, it might be good for you that you are happy right now. It is also good for you that you are happy right now and there exist at least seven giraffes. But this latter proposition derives its goodness from the proposition, which it entails, that you are happy right now. Harman (1967) argues that the standard notion of intrinsic value requires a notion of *basic* intrinsic value; I am simply adapting that argument to welfare. I will assume that, for any individual, there is a set of *welfare-basic* propositions, which are intrinsically good or bad for that individual in the most fundamental way, and that other propositions are intrinsically good or bad for them only in a derivative way.

The general idea will be to concatenate welfare-basic propositions via conjunction. But there are several immediate problems with this suggestion (Danielsson 1997). Some welfare-basic propositions may be inconsistent, so that their conjunction is not some new proposition that is better than both. Conjunction is idempotent: $p \wedge p$ has the same value as p. And, most relevant to the axioms of extensive measurement, the value of a conjunction of (consistent and distinct) welfare-basic propositions need not intuitively be the sum of the values of its conjuncts. For example, if we take at least some of the welfare-basic propositions to be about how well things are going for you over particular periods of your life, then the apparent counterexamples to Independence will apply here, too: the conjunction *that you have a happy childhood and that you have a sad adulthood* may be worse for you than what is intuitively the sum of the values of the propositions *that you have a happy childhood* and *that you have a sad adulthood*. And the value of your taking pleasure in some excellent activity might be greater than what is intuitively the value of your experiencing pleasure and the value of your engagement in that activity.

These apparent counterexamples to Independence are prima facie cases of what Moore (1903) called *organic unities*: wholes whose value is not equal to the sum of the values of their parts. This doctrine did not lead Moore to reject the additivity of value in general. He thought the value of a whole was not the sum of the values of its parts, but the sum of that value *plus* the value that arises distinctively from the combination of those parts: 'the value which a thing possesses *on the whole* may be said to be equivalent to the sum of the value which it possesses *as a whole*, *together with* the intrinsic values which may belong to any of its parts' (Moore 1903, §129, emphasis in original). Moore's idea of value-as-a-whole can be understood in terms of basic intrinsic value (Danielsson 1997; Carlson 2001). After all, if the value of a whole is not the sum of the values of its parts, then (we might think) there must be some feature of the whole—perhaps some interaction between the parts—that is good or bad, and which makes the whole better or worse. This feature might be claimed to have basic intrinsic value. For example, the conjunction *that you have a happy childhood and that you have a sad adulthood* might entail at least three welfare-basic propositions: that you have a happy childhood, that you have a sad adulthood, and that your life goes downhill (or some other proposition regarding the narrative structure of your life that is basically good or bad for you).

How, then, can we assign values to conjunctions of welfare-basic propositions that entail new welfare-basic propositions, as seems characteristic of organic unities? Following Danielsson (1997), we can use surrogates for these propositions which are *evaluatively independent*. Suppose that p is the conjunction of welfare-basic propositions $p_1, \ldots, p_n$, and q is the conjunction of welfare-basic propositions $q_1, \ldots, q_m$. Then p and q are evaluatively independent if and only if any welfare-basic proposition entailed by $p_1 \wedge \ldots \wedge p_n \wedge q_1 \wedge \ldots \wedge q_m$ is entailed by exactly one of these conjuncts. This means that their conjunction does not introduce any new welfare-basic proposition that we did not already have before; it has no value 'as a whole' to be considered. (This definition also excludes inconsistent propositions, since $p \wedge \neg p$ classically entails arbitrary propositions that are not entailed by either conjunct. Also, p is not evaluatively independent from itself, since any proposition entailed by $p \wedge p$ is entailed by both conjuncts.)

The idea, then, will be as follows. Suppose that p and q are basically good for you, and indeed equally so. If they are evaluatively independent, then $p \wedge q$ is twice as good for you as each conjunct. If they are not evaluatively independent, then we list all of the welfare-basic propositions for you entailed by their conjunction: say, p, q and r. We then replace these propositions with suitable surrogates—welfare-basic propositions that are evaluatively independent and just as good. The value of $p \wedge q$ will then be the value of the conjunction $p' \wedge q' \wedge r'$ where p' is just as good for you as p, q' just as good for you as q, and r' just as good for you as r.

We want, however, to add up goodness for different people, and one and the same proposition can have different values for different people. For example, *that I am happy and you are sad* is good for me but bad for you. So we will need to work with richer objects: *centred propositions*, which pair a proposition with an individual. Our relation $\succcurlyeq$ on centred propositions will be interpreted as follows: for any propositions p and q, and individuals i and j, $(p, i) \succcurlyeq (q, j)$ just in case p is at least as intrinsically good for i as q is for j.

We now have all the ingredients of the proposal on the table. Our structure will be as follows. We have a finite set of individuals numbered

1, …, n. For each individual i, there is a set of welfare-basic propositions for i. Let A denote the set of all centred propositions of the form (p, i) where p is a finite, consistent conjunction of welfare-basic propositions for some person i. We have a relation $\succcurlyeq$ on this set, interpreted as above, which we assume to be transitive and complete. We assume that, for any individuals i and j, and any $(p, i), (q, i), (p', j), (q', j)$ in A, where $(p, i) \sim (p', j)$ and $(q, i) \sim (q', j)$, if p and q are evaluatively independent for i, and q and q' are evaluatively independent for j, then $(p \wedge q, i) \sim (p' \wedge q', j)$.

We can partition the set A into equivalence classes under the relation of being equally good for a person. $(\mathbf{p}, \mathbf{i})$ is the set that contains every centred proposition in A that is just as good for the individual at its centre as p is for i. $\mathbf{A}$ is the set of all equivalence classes of such centred propositions. We can then define a relation $\succeq$ on $\mathbf{A}$: $(\mathbf{p}, \mathbf{i}) \succeq (\mathbf{q}, \mathbf{j})$ if and only if $(p, i) \succcurlyeq (q, j)$.

To concatenate these equivalence classes, we need to introduce a further assumption. Assume that, for any $(p_1, 1), \ldots, (p_n, n) \in A$, there is some $(p'_1 \wedge \ldots \wedge p'_n, i) \in A$ where $(p'_1, i) \sim (p_1, 1), \ldots$, and $(p'_n, i) \sim (p_n, n)$ and $p'_1, \ldots, p'_n$ are evaluatively independent for i. This is a very strong assumption: it means that we can always find a single individual and evaluatively independent surrogates to evaluate any number of finite, consistent conjunctions of welfare-basic propositions for any number of individuals. But, with that assumption, we can concatenate any equivalence classes as follows: for any $(\mathbf{p}, \mathbf{i}), (\mathbf{q}, \mathbf{j})$ in $\mathbf{A}$, $(\mathbf{p}, \mathbf{i}) \circ (\mathbf{q}, \mathbf{j}) = (\mathbf{p}' \wedge \mathbf{q}', \mathbf{k})$, where p' is just as good for k as p is for i, q' is just as good for k as q is for j, p' and q' are evaluatively independent for k. Our assumption guarantees that this operation is well-defined for any pair of equivalence classes in $\mathbf{A}$. Essentially, we are concatenating centred (conjunctions of welfare-basic) propositions by finding equally good surrogates that are evaluatively independent, so that they do not give rise to organic unities.

Does this structure satisfy the axioms of extensive measurement? $\succeq$ inherits its transitivity and completeness from the ordering $\succcurlyeq$ on A. The independence axiom is made plausible by the use of evaluatively independent surrogates for concatenation. For Independence to fail, there would have to be elements (p, i), (q, i) and (r, i) in A such that p is at least as intrinsically good for i as q but $p \wedge r$ is not at least as intrinsically good for i as $q \wedge r$, even though these conjunctions do not entail any new welfare-basic proposition that is not entailed by the conjuncts alone. It is hard to see how this could possibly be true, since the comparison of $p \wedge r$'s and $q \wedge r$'s intrinsic values should presumably be explained by some fact entailed by one of the conjunctions that does not arise for the comparison of p and q alone.[5] So if the assumptions of this system are satisfied, then Independence should hold.

If the axioms are satisfied, then we can assign numbers to all elements of **A**, which we can call *welfare levels*. The concatenation of two welfare levels will be represented by the sum of the numbers assigned to those levels. An individual's welfare level at a world can then be identified as the equivalence class containing the strongest finite conjunction of welfare-basic propositions for that individual that are true at that world. And the total welfare at a world across finitely many people can be identified as the concatenation of all of these people's welfare levels.

This approach avoids the difficulties with the other approaches we have considered. It does not require welfare to be additive over time, and it is compatible with the existence of organic unities in a theory of welfare. It also does not require a person's welfare to depend on intrinsic properties of her life, and concatenates bearers of well-being in such a way that their goodness is still goodness for a person. It is also very flexible as to what kinds of things are good or bad for a person, and allows different kinds of things to be good for different people.

The assumptions of this approach are, however, extremely demanding, and it is completely unclear whether they should be satisfied by a reasonable theory of welfare. Even if we are perfectly comfortable with the ideology of welfare-basic propositions, and with the assumption that any given possible world entails finitely many such propositions for any given person, we would need to believe that for any collection of such propositions for any number of people, there is a collection of evaluatively independent surrogates of those propositions for some single person. This assumption seems questionable. For example, if the most important goods for a person are organic unities — for example, involving the narrative structure of one's life, or how the various goods in one's life relate — then we might doubt whether those goods could be mapped onto evaluatively independent surrogates in the way required for the set to be closed under concatenation. So while the appeal to evaluatively independent propositions allows this view to accommodate the existence of some organic unities, it is not clear that it is enough to accommodate a theory of welfare in which organic unities play a particularly central role.

[5] We cannot say, in this context, that p's value decreases when conjoined with r, since its value was supposed to be intrinsic. For independent objections to this sort of move, see Bradley (2002), Brown (2007), and Carlson (2020).

A further problem has to do with the specificity of welfare-basic propositions. On a maximally holistic view, the welfare-basic propositions might be complete possible worlds. On such a view, we cannot concatenate any welfare-basic propositions, since no two such propositions are even consistent. Now, this is an extreme view. It seems to deny that there are any general propositions in virtue of which worlds can be good or bad for a person. And we might have thought it to be the job of a theory of welfare to identify such propositions. But we can think of this view as one end of a spectrum with respect to how specific or general the welfare-basic propositions are. Even if they are not as specific as possible worlds, they might be specific enough that there are not sufficiently many consistent welfare-basic propositions to go around. And if they are instead too general, we might then worry whether worlds like ours might entail infinitely many of them for any given person.

Our discussion of this proposal has been, admittedly, quite speculative. But this is only because the proposal is so complicated that it seems unreasonable to be confident in its assumptions. It may indeed provide a way of fitting well-being into an extensive structure, without the drawbacks of the other accounts we have considered. But if this is what it takes for there to be a sum of well-being, it should leave us completely unsure whether there is any such thing.

7. Conclusion

To make sense of adding well-being, we would need to show that well-being has an extensive structure. We have considered various ways of fitting well-being into such a structure and have seen that they carry highly controversial commitments about the good. There are surely other possibilities. But I imagine that they will carry controversial commitments of their own.

It would be premature to conclude from the difficulties we have encountered that there is no such thing as the sum of well-being. But we might reasonably wonder, at this point, what reason we have to think that there is such a thing. Whether the commitments of extensive measurement are worth accepting depends, in part, on what work the sum of well-being is supposed to do in our ethical theory, and whether it can earn its keep. If the sum of well-being plays some indispensable role in our ethical lives, then we need some way of making sense of it. So, do we need sums of well-being to do ethics?

Doubts about this need might come from certain corners of social choice theory. Much of this literature asks how the ethical ranking of outcomes can be represented by some function of numbers ('utilities') that, in some sense, represent individuals' preferences or well-being (Sen 1970; d'Aspremont and Gevers 2002; Weymark 2016; Adler 2019). Various theorems show that, under certain conditions, the ethical ranking can be represented as maximizing a sum of (possibly weighted or transformed) utilities (Fleming 1952; Harsanyi 1955; d'Aspremont and Gevers 1977; Maskin 1978; Mongin 1994; Blackorby, Bossert and Donaldson 2005). Crucially, however, the nature of this representation, at both the social and individual level, is left open by the formal framework. There is no need to insist, within this framework, that there is a *sum of well-being* to be represented by the sum of these numbers.[6]

Some theorists in this tradition *do* seem to care whether their sums of numerical utilities represent 'an attribute of persons which it is meaningful to sum' (Roemer 1998, p. 130). A central critique of Harsanyi's (1955) aggregation theorem hinges on this very question (Sen 1977a; Weymark 1991). As Roemer (1998, p. 145) puts it, 'The view that Harsanyi's aggregation theorem has anything to do with the idea that society should maximize the total well-being of its members is wrong by virtue of a confusion concerning the representation of orders by cardinal utility functions'. Many followers of Harsanyi, rather than simply accepting this critique and disclaiming any concern for sums of well-being, try to answer it head-on. Broome, for instance, argues that each person's good is 'an arithmetic quantity' (1991, p. 124) and aims to show that 'the total of people's utilities will measure the total of people's good' (1991, p. 218); see also Risse (2002), Fleurbaey and Mongin (2016), and Greaves (2017b); for a measurement-theoretic critique, see Weymark (2005). Indeed, Harsanyi motivates his own contribution as an improvement on the theorem of Fleming, who interprets his results to 'yield a concept of welfare as an extensive or additive magnitude, susceptible of fundamental measurement' (Fleming 1952, p. 366).

Still, one might think that this debate is confused, or of no ethical importance: if the ethical ranking of outcomes turns out to be representable as maximizing a sum of numerical utilities, why should it matter whether this sum represents a sum of well-being?

This is a difficult and important question. I suppose that if we already had some utility function for each person and knew how to represent

[6] In a somewhat different context, Gustafsson (2021) argues that any 'utilitarian' evaluation concerning fixed populations can be justified without aggregating different people's well-beings.

the ethical ranking of outcomes with respect to the utilities assigned by those functions, then it wouldn't matter, for practical purposes, whether the sum of those utilities represents a sum of well-being, or what, more generally, those utilities represent, since we would be in a position to know the ethical ranking of outcomes (though the question may be of theoretical interest). But if we aren't already sure how to assign or compare different possible assignments of utilities, then our question—whether well-being can be summed—might bear on how we do so, if only indirectly.

In the literature on 'social welfare functionals', it is widely believed that the type of scale on which well-being is measurable, as well as its degree of interpersonal comparability, has important ethical implications (Sen 1977b). These implications are derived from *invariance conditions*, which require the ethical ranking of outcomes to be invariant to 'admissible' transformations of individual utility functions—that is, those transformations up to which the numerical representation is unique. Adler (2019) calls this requirement the *Fundamental Principle of Invariance*. For example, I have argued that well-being can be summed only if it's measurable on an interpersonal ratio scale. Suppose, however, that it's only measurable on a (fully interpersonal) interval scale. Since an interval scale is unique only up to positive affine transformation, the Fundamental Principle of Invariance would then require the ranking of outcomes to be invariant to common positive affine transformations of all people's utilities. This particular invariance condition is used in, for example, Maskin's (1978) axiomatic characterization of the sum-of-utilities social welfare functional.

Even this relatively weak invariance condition, however, has highly unappealing implications, particularly in variable-population contexts. For example, given modest assumptions, it implies that the desirability of expanding a population depends only on how additional lives compare to the population's average well-being: above-average expansions make things better, below-average worse (Blackorby, Bossert and Donaldson 2005, p. 196). This is extremely restrictive: intuitively, if everyone's life were sufficiently horrible, it would not be *good* to add only slightly less horrible lives (Parfit 1984, p. 406). Blackorby, Bossert and Donaldson (1999, p. 420) therefore conclude that 'more information than that provided by cardinal full comparability [that is, an interpersonal interval scale] is needed to generate acceptable variable-population social-welfare orderings'. This problem is avoided by the weaker invariance condition associated with an interpersonal ratio scale, which requires the ranking of outcomes to be invariant to common similarity transformations of all people's utilities. This condition still has significant ethical bite (see Nebel 2021), but of a kind that seems more conducive to making progress by narrowing down the set of eligible theories, rather than leading to absurdity.

Thus, within this framework, it matters greatly whether well-being is measurable on a ratio scale or only on an interval scale. The implications of interval-scale invariance for variable-population ethics seem to me unacceptable. I therefore believe that we must either find a way to construct a ratio scale of well-being or reject the Fundamental Principle of Invariance. Since extensive measurement is the standard strategy for deriving a ratio scale, if it turns out that well-being does not have extensive structure, that should decrease our confidence in the Fundamental Principle of Invariance.[7] The rationale behind this principle has recently been questioned on independent grounds (Morreau and Weymark 2016; Nebel 2021, 2022b). If we reject it, there is considerably more freedom in how we compare alternative distributions of well-being. However, theories that are invariant to admissible transformations seem more attractive, *ceteris paribus*, than those that aren't (Nebel 2023). So we must weigh the benefits of invariance against the difficulties posed by extensive measurement. Sorting through those difficulties would not only help us make progress in the theory of well-being; it would also allow us to use the invariance condition associated with ratio-scale measurability to rule out otherwise live options in the ethics of population and distribution.

Ultimately, I do not know whether there is such a thing as the sum of well-being, or even whether we need there to be. But I hope to have shown that we cannot simply take for granted, as many philosophers do, that there is a 'net sum of *whatever* makes life worth living' (Parfit 1984, p. 404, emphasis added). Whether or not there is such a thing as the sum

[7] The zeroing out method discussed in §2 may seem adequate for purposes of navigating this dilemma, even if it does not establish sums or ratios of well-being. We could claim that some level of well-being is relevant to the ethical ranking of outcomes, in a way that violates interval-scale invariance (see Adler 2011, p. 217 n. 124), even without claiming that lives at that level have 'zero well-being'. But this carries a risk of trivializing the Fundamental Principle of Invariance. For we could just as well, at least in principle, identify any *other* level of well-being and claim it to be ethically relevant too, in a way that violates ratio-scale invariance ('oneing out'). This, in effect, makes *all* levels 'ethically relevant', since the only transformation of utilities that preserves ratios of differences and maps two levels to themselves is the identity. So the usefulness of zeroing out for this purpose requires showing not only that some welfare level is ethically relevant in the sense alleged but also that *no other level* is ethically relevant; I find it hard to be optimistic about the prospects for such an argument.

of how good things are for one person and how good they are for another depends on what, in fact, is good for people.[8]

References

Adler, Matthew D. 2011: *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

—— 2019: *Measuring Social Welfare: An Introduction*. Oxford: Oxford University Press.

—— and N. Treich 2015: 'Prioritarianism and Climate Change'. *Environmental and Resource Economics*, 62(2), pp. 279–308.

Arrhenius, Gustaf 2000: 'Future Generations : A Challenge for Moral Theory'. Ph.D. thesis, Uppsala University.

—— and Wlodek Rabinowicz 2015: 'The Value of Existence'. In Iwao Hirose and Jonas Olson (eds.), *The Oxford Handbook of Value Theory*, pp. 424–43. New York: Oxford University Press.

Arrow, Kenneth J. 1951: *Social Choice and Individual Values*. New York: Wiley.

Bader, Ralf M. MS: 'Person-Affecting Population Ethics'. Unpublished manuscript.

Bennett, Jonathan 1978: 'On Maximising Happiness'. In R. I. Sikora and Brian Barry (eds.), *Obligations to Future Generations*, pp. 61–73. Philadelphia: Temple University Press.

Blackorby, Charles, Walter Bossert, and David Donaldson 1999: 'Information Invariance in Variable-Population Social-Choice Problems'. *International Economic Review*, 40(2), pp. 403–22.

—— 2005: *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. New York: Cambridge University Press.

Boonin-Vail, David 1996: 'Don't Stop Thinking About Tomorrow: Two Paradoxes About Duties to Future Generations'. *Philosophy and Public Affairs*, 25(4), pp. 267–307.

Bradley, Ben 2002: 'Is Intrinsic Value Conditional?' *Philosophical Studies*, 107(1), pp. 23–44.

Broome, John 1991: *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Basil Blackwell.

[8] For helpful comments and discussion, I am grateful to Jeremy Goodman, Zachary Goodsell, Ben Holguín, Dean Spears, Ralph Wedgwood, two anonymous referees for *Mind*, and audiences at Arizona State University, Princeton University, Rice University, Rutgers University–New Brunswick, and the University of Texas at Austin. I am especially indebted to Ralf Bader; I owe the general approach taken in this paper to conversations with him.

—— 1993: 'Goodness Is Reducible to Betterness: The Evil of Death Is the Value of Life'. In Peter Koslowski and Yuichi Shionoya (eds.), *The Good and the Economical: Ethical Choices in Economics and Management*, pp. 70–84. Berlin and Heidelberg: Springer-Verlag.
—— 1997: 'Is Incommensurability Vagueness?' In Ruth Chang (ed.), *Incommensurability, Incomparability, and Practical Reason*, pp. 67–89. Cambridge, MA: Harvard University Press.
—— 2004: *Weighing Lives*. Oxford and New York: Oxford University Press.
—— 2007: 'Replies'. *Economics and Philosophy*, 23(1), pp. 115–24.
Brown, Campbell 2007: 'Two Kinds of Holism About Values'. *Philosophical Quarterly*, 57(228), pp. 456–63.
Bykvist, Krister 2007: 'The Benefits of Coming into Existence'. *Philosophical Studies*, 135(3), pp. 335–62.
—— 2021: 'Taking Values Seriously'. *Synthese*, 199(3–4), pp. 6331–56.
Carlson, Erik 2001: 'Organic Unities, Non-Trade-Off, and the Additivity of Intrinsic Value'. *Journal of Ethics*, 5(4), pp. 335–60.
—— 2008: 'Extensive Measurement with Incomparability'. *Journal of Mathematical Psychology*, 52(4), pp. 250–9.
—— 2010: 'Generalized Extensive Measurement for Lexicographic Orders'. *Journal of Mathematical Psychology*, 54(4), pp. 345–51.
—— 2020: 'Organic Unities and Conditionalism About Final Value'. *Journal of Value Inquiry*, 54(2), pp. 175–81.
Chang, Ruth 2002: 'The Possibility of Parity'. *Ethics*, 112(4), pp. 659–88.
d'Aspremont, Claude, and Louis Gevers 1977: 'Equity and the Informational Basis of Collective Choice'. *Review of Economic Studies*, 44(2), pp. 199–209.
———— 2002: 'Social Welfare Functionals and Interpersonal Comparability'. In Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura (eds.), *Handbook of Social Choice and Welfare*, Vol. 1, pp. 459–541. Amsterdam: Elsevier.
Danielsson, Sven 1997: 'Harman's Equation and the Additivity of Intrinsic Value'. In Lars Lindahl, Paul Needham, and Rysiek Śliwiński (eds.), *For Good Measure: Philosophical Essays Dedicated to Jan Odelstad on the Occasion of His Fiftieth Birthday*, pp. 23–34. Uppsala: Department of Philosophy, University of Uppsala.
Dasgupta, P. 1988: 'Lives and Well-Being'. In W. Gaertner and P. K. Pattanaik (eds.), *Distributive Justice and Inequality: A Selection of Papers Given at a Conference, Berlin, May 1986*, pp. 15–38. Berlin and Heidelberg: Springer-Verlag.

Dorr, Cian, Jacob M. Nebel, and Jake Zuehl 2023: 'The Case for Comparability'. *Noûs*, 57(2), pp. 414–53.
Fleming, Marcus 1952: 'A Cardinal Concept of Welfare'. *Quarterly Journal of Economics*, 66(3), pp. 366–84.
Fleurbaey, Marc, and Philippe Mongin 2016: 'The Utilitarian Relevance of the Aggregation Theorem'. *American Economic Journal: Microeconomics*, 8(3), pp. 289–306.
Fleurbaey, Marc, and Alex Voorhoeve 2015: 'On the Social and Personal Value of Existence'. In Iwao Hirose and Andrew Reisner (eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome*, pp. 95–109. Oxford: Oxford University Press.
Greaves, Hilary 2017a: 'Population Axiology'. *Philosophy Compass*, 12(11), e12442.
—— 2017b: 'A Reconsideration of the Harsanyi–Sen–Weymark Debate on Utilitarianism'. *Utilitas*, 29(2), pp. 175–213.
Gustafsson, Johan E. 2021: 'Utilitarianism without Moral Aggregation'. *Canadian Journal of Philosophy*, 51(4), pp. 256–69.
Harman, Gilbert H. 1967: 'Toward a Theory of Intrinsic Value'. *Journal of Philosophy*, 64(23), pp. 792–804.
Harsanyi, John C. 1955: 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility'. *Journal of Political Economy*, 63(4), pp. 309–21.
Hempel, Carl G. 1952: *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press.
Holtug, Nils 1999: 'Utility, Priority and Possible People'. *Utilitas*, 11(1), pp. 16–36.
Huemer, Michael 2008: 'In Defence of Repugnance'. *Mind*, 117(468), pp. 899–933.
—— 2013: 'Transitivity, Comparative Value, and the Methods of Ethics: A Review of Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*'. *Ethics*, 123(2), pp. 318–45.
Hurka, Thomas 1983: 'Value and Population Size'. *Ethics*, 93(3), pp. 496–507.
Jensen, Karsten Klint 2008: 'Millian Superiorities and the Repugnant Conclusion'. *Utilitas*, 20(3), pp. 279–300.
Kagan, Shelly 1998: *Normative Ethics*. Boulder, CO and Oxford: Westview Press.
Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin 1997: 'Back to Bentham? Explorations of Experienced Utility'. *Quarterly Journal of Economics*, 112(2), pp. 375–405.

Kamm, F. M. 1988: 'Why Is Death Bad and Worse Than Pre-Natal Non-Existence?' *Pacific Philosophical Quarterly*, 69(2), 161–4.
Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky 1971: *Foundations of Measurement, Volume* I: *Additive and Polynomial Representations*. New York: Academic Press.
Laming, Donald 1997: *The Measurement of Sensation*. Oxford: Oxford University Press.
Larsson, Felix 2006: 'What? More Better?' In Björn Haglund and Helge Malmgren (eds.), *Kvantifikator för en Dag: Essays Dedicated to Dag Westerståhl on His Sixtieth Birthday*. Gothenburg: Philosophical Communications.
Lewis, C. I. 1946: *An Analysis Of Knowledge And Valuation*. La Salle, IL: Open Court.
Lewis, David 1983: 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy*, 61(4), pp. 343–77.
Luce, R. Duncan, and Louis Narens 1985: 'Classification of Concatenation Measurement Structures According to Scale Type'. *Journal of Mathematical Psychology*, 29(1), pp. 1–72.
McMahan, Jefferson 1981: 'Problems of Population Theory'. *Ethics*, 92(1), pp. 96–127.
McQueen, Kelvin J. 2015: 'Mass Additivity and A Priori Entailment'. *Synthese*, 192(5), pp. 1373–92.
McTaggart, John McTaggart Ellis 1927: *The Nature of Existence*, Vol. II. Edited by C. D. Broad. Cambridge: Cambridge University Press.
Maskin, Eric 1978: 'A Theorem on Utilitarianism'. *Review of Economic Studies*, 45(1), pp. 93–6.
Mongin, Philippe 1994: 'Harsanyi's Aggregation Theorem: Multi-Profile Version and Unsettled Questions'. *Social Choice and Welfare*, 11(4), pp. 331–54.
Moore, G. E. 1903: *Principia Ethica*. Cambridge: Cambridge University Press.
Morreau, Michael, and John A. Weymark 2016: 'Measurement Scales and Welfarist Social Choice'. *Journal of Mathematical Psychology*, 75, pp. 127–36.
Nebel, Jacob M. 2018: 'The Good, the Bad, and the Transitivity of Better Than'. *Noûs*, 52(4), pp. 874–99.
—— 2021: 'Utils and Shmutils'. *Ethics*, 131(3), pp. 571–99.
—— 2022a: 'Totalism Without Repugnance'. In Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit*, pp. 200–31. Oxford: Oxford University Press.

—— 2022b: 'Aggregation Without Interpersonal Comparisons of Well-Being'. *Philosophy and Phenomenological Research*, 105(1), pp. 18–41.
—— 2023: 'Ethics Without Numbers'. Forthcoming in *Philosophy and Phenomenological Research*.
Otsuka, Michael 2018: 'How It Makes a Moral Difference That One Is Worse Off than One Could Have Been'. *Politics, Philosophy and Economics*, 17(2), pp. 192–215.
Parfit, Derek 1984: *Reasons and Persons*. Oxford: Clarendon Press.
—— 1986: 'Overpopulation and the Quality of Life'. In Peter Singer (ed.), *Applied Ethics*, pp. 145–64. Oxford Readings in Philosophy. Oxford: Oxford University Press.
—— 2017: 'Future People, the Non-Identity Problem, and Person-Affecting Principles'. *Philosophy and Public Affairs*, 45(2), pp. 118–57.
Pummer, Theron 2017: 'Spectrum Arguments and Hypersensitivity'. *Philosophical Studies*, 175(7), pp. 1729–44.
Rachels, Stuart 1998: 'Counterexamples to the Transitivity of *Better Than*'. *Australasian Journal of Philosophy*, 76(1), pp. 71–83.
Rawls, John 1971: *A Theory of Justice*. Cambridge, MA: Belknap Press.
Raz, Joseph 1985: 'Value Incommensurability: Some Preliminaries'. *Proceedings of the Aristotelian Society*, 86(2), pp. 117–34.
Regan, Donald 1997: 'Value, Comparability, and Choice'. In Ruth Chang (ed.), *Incommensurability, Incomparability, and Practical Reason*, pp. 129–50. Cambridge, MA: Harvard Universioty Press.
Risse, Mathias 2002: 'Harsanyi's "Utilitarian Theorem" and Utilitarianism'. *Noûs*, 36(4), pp. 550–77.
Roberts, M. A. 2002: 'A New Way of Doing the Best That We Can: Person-Based Consequentialism and the Equality Problem'. *Ethics*, 112(2), pp. 315–50.
Roemer, John E. 1998: *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
Rozeboom, William W. 1966: 'Scaling Theory and the Nature of Measurement'. *Synthese*, 16(2), pp. 170–233.
Sen, Amartya K. 1970: *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
—— 1977a: 'Non-Linear Social Welfare Functions: A Reply to Professor Harsanyi'. In Robert E. Butts and Jaakko Hintikka (eds.), *Foundational Problems in the Special Sciences*, pp. 297–302. Dordrecht: D. Reidel.

—— 1977b: ‘On Weights and Measures: Informational Constraints in Social Welfare Analysis’. *Econometrica*, 45(7), pp. 1539–72.
Skyrms, Brian, and Louis Narens 2019: ‘Measuring the Hedonimeter’. *Philosophical Studies*, 176(12), pp. 3199–210.
Tännsjö, Torbjörn 2002: ‘Why We Ought to Accept the Repugnant Conclusion’. *Utilitas*, 14(3), pp. 339–59.
Taurek, John 2021: ‘Reply to Parfit’s “Innumerate Ethics”’. In Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (eds.), *Principles and Persons: The Legacy of Derek Parfit*, pp. 311–22. Oxford: Oxford University Press.
Temkin, Larry S. 1996: ‘A Continuum Argument for Intransitivity’. *Philosophy and Public Affairs*, 25(3), pp. 175–210.
—— 2012: *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford and New York: Oxford University Press.
Velleman, J. David 1991: ‘Well-Being and Time’. *Pacific Philosophical Quarterly*, 72(1), pp. 48–77.
Voorhoeve, Alex 2013: ‘Vaulting Intuition: Temkin’s Critique of Transitivity’. *Economics and Philosophy*, 29(3), pp. 409–23.
Weymark, John A. 1991: ‘A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism’. In Jon Elster and John E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, pp. 255–320. Cambridge: Cambridge University Press.
—— 2005: ‘Measurement Theory and the Foundations of Utilitarianism’. *Social Choice and Welfare*, 25(2/3), pp. 527–55.
—— 2016: ‘Social Welfare Functions’. In Matthew D. Adler and Marc Fleurbaey (eds.), *The Oxford Handbook of Well-Being and Public Policy*, pp. 126–59. New York: Oxford University Press.

