Noûs

NOÛS

ORIGINAL ARTICLE OPEN ACCESS

Aggregation and the Structure of Value

Weng Kin San

Department of Philosophy, Logic, and Scientific Method, London School of Economics and Political Science, London, UK

Correspondence:
Weng Kin San (w.san@lse.ac.uk)

Received: 13 February 2024 | **Revised:** 4 November 2025 | **Accepted:** 14 November 2025

ABSTRACT

Roughly, the view I call "Additivism" sums up value across time and people. Given some standard assumptions, I show that Additivism follows from two principles. The first says that how lives align in time cannot, in itself, matter. The second says, roughly, that a world cannot be better unless it is better within some period or another. These principles, while plausible, presuppose a rich underlying structure of value—presuppositions that are implicit in the standard numerical framework of population ethics but that are often overlooked. A careful exploration of Additivism and the case for it reveals intricate connections between substantive questions about what value fundamentally consists in and structural questions about how to aggregate value.

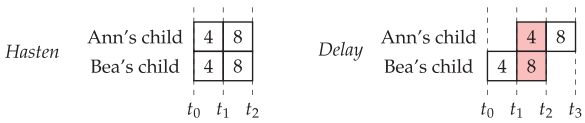
1 | Introduction

What is best for one person is not always best for another. From this arises the question of *interpersonal* aggregation central to distributive ethics: how does what is good for each person combine to determine what is best overall? But also, people's lives are extended in time. And what is best at one time need not be best at another. So, there is also the question of *intertemporal* aggregation: how does what is good at different times combine to determine what is best overall?

Considering both questions in tandem can be highly fruitful. A careful investigation of issues at the interface of the dimensions of time and people will incline us toward an "additive" picture of value.[1] To foreshadow some of the arguments to come, consider a nonadditive theory like Egalitarianism. Many think that inequality is bad—it can badly affect how well people's lives go.[2] But some think that, even once we account for inequality's bad effects, an unequal distribution of welfare among people is intrinsically bad. For such egalitarians, a question arises when we take into account the dimension of time. What is intrinsically bad: inequality *at a time* or inequality over a *lifetime*?[3]

Suppose it is inequality at each time. This kind of Timeslice Egalitarianism is objectionably sensitive to mere differences in how lives align in time. Consider:

> *Hasten vs. Delay*. Ann can induce labor and give birth earlier. This will affect neither her nor her child's well-being. But elsewhere, Bea is also pregnant and about to deliver. Although Ann and Bea's children will never meet or interact in any way, *when* Ann gives birth will affect whether their children's lives are contemporaneous.



[1] A similar theme is explored, notably, in Parfit (1984), Broome (1991) and Broome (2004b).
[2] See R. Wilkinson and Pickett (2009).
[3] See McKerlie (1989), McKerlie (2001, 2012), Temkin (1993), and Adler (2007).

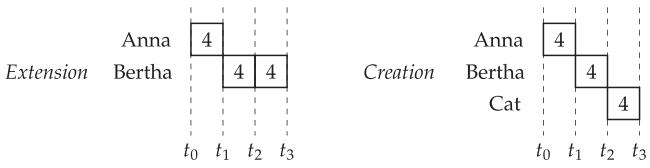
Insofar as it has no bearing on how well anyone's life goes, there is no reason to try to align the two children's timelines one way or another. And that is indeed the case on a view like Total Utilitarianism, which compares worlds by their total welfare.

But not so according to Timeslice Egalitarianism. Only in *Delay* is there any time where there is interpersonal inequality. The distribution of welfare is unequal from t_1 to t_2, when the low of one life happens to coincide with the peak of the other. So, unless *Delay* is better in some other respect that makes up for the inequality, Timeslice Egalitarians should prefer *Hasten*.

It is presumably for reasons like this that most egalitarians are Lifetime Egalitarians instead—what is bad is inequality in lifetime welfare. For instance, Nagel says that "the subject of an egalitarian principle is not the distribution of particular rewards to individuals at some time, but the prospective quality of their lives as a whole, from birth to death" (1995, 69).

But prima facie, Lifetime Egalitarians face problems of their own at the intersection of the interpersonal and intertemporal dimensions. For instance, there can sometimes be trade-offs between adding people and adding time:

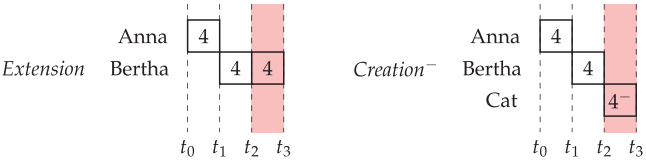
> *Extension vs. Creation*. Bertha isn't pregnant. But if she were to be, there's no scenario in which her and her child both survive childbirth. In fact, this condition is hereditary—Bertha's mother, Anna, died this way giving birth to Bertha. And Bertha's child, Cat, wouldn't even have a long life—only as long as Bertha herself would otherwise go on to live. The child's quality of life would also be only as good as Bertha's would have been.[4]



Even if Bertha has the prerogative to have a child, it surely cannot be justified here on the basis of bringing about a better world.

But it might on Lifetime Egalitarianism. Assuming lifetime welfare to be obtained by summing up welfare across time, *Creation* contains perfect lifetime equality (all three people enjoy a total lifetime welfare of 4) while *Extension* contains two people with unequal total lifetime welfare of 4 and 8. So, on egalitarian views that combine an aversion to inequality solely with a concern for total welfare, *Creation* would be preferable—an objectionable conclusion.

Worse still, if *Creation* is better, it is better by some margin. So, granting some continuity assumptions, things would presumably still be better than *Extension*, even if Bertha's child were to have a slightly worse life (of welfare 4^- slightly less than 4) as in *Creation*$^-$.



But Bertha's choice affects only what happens after t_2, prior to which the distribution of welfare is the same. But after t_2, one person is at welfare level 4 in *Extension* whereas one person is at welfare level 4^- in *Creation*$^-$. So, *Creation*$^-$ is the same as *Extension* before t_2 and worse after. How then can *Creation*$^-$ be better when there is no time when it is better?

These problems are not decisive but they gesture at the arguments to come. And the problems are not exclusive to Egalitarianism. For instance, taking into account the dimension of time, there are also at least two salient forms of Averagism.[5] A Timeslice Averagist first averages welfare across people at each time before aggregating those averages across time. Like Timeslice Egalitarianism, this runs into problems with cases like *Hasten vs. Delay*, where the only difference is in when one person lives relative to another. Both violate the following principle to be explored in *Section* 3:

Time-Shift Invariance. Worlds that differ only by a time-shift are equally good.

A Lifetime Averagist, on the other hand, first aggregates welfare across time for each person before averaging their resulting lifetime welfare. Like Lifetime Egalitarianism, this runs into problems with cases like *Extension vs. Creation*. On both theories, one possibility can be better than another even though there is no period when it is better. They violate the following principle explored in greater depth in *Section* 4:

Time-Partition Dominance.If one world is at least as good as another both before and after time t, then it's at least as good overall. If, furthermore, it's also better either before or after t, then it's better.

A view that satisfies both these principles is a family of views I call "Additivism." On those views, roughly, value can be thought of as being summed up across people and time. Of course, Egalitarianism, Averagism, and Additivism don't exhaust the space of possible theories. But it turns out that the problems for Egalitarianism and Averagism generalize to any "nonadditive" theory. The main technical result of this paper in *Section* 5 shows that given some standard background assumptions, the only kind of theory that satisfies both Time-Shift Invariance and

[4] Similar cases of prolonging vs. creating are discussed in Broome (1991), Broome (2004a) and Arrhenius (2011).
[5] See Hurka (1982a), Hurka (1982b).

Time-Partition Dominance is additive—jointly, the two principles axiomatically characterize Additivism.[6]

Presenting the case for Additivism requires a deep dive into some foundational issues in axiology. Such issues arise even just in formulating the two key axioms. Take Time-Shift Invariance. Clearly, when a person lives can make a difference—my life would be much worse were I to be magically transported, as I am, to the Middle Ages. So, the claim that it's irrelevant when someone lives is plausible only if all else besides the time difference is held constant.

To control for this, *Hasten vs. Delay* held fixed the numbers in the diagram when shifting the boxes sideways along the time axis. This is supposed to indicate that the difference is "purely temporal." But this raises the question of what these numbers represent and what exactly is involved in holding them fixed. Formulating the other axiom, Time-Partition Dominance, requires explicating what it means for things to be better or worse within a given time period. This raises similar foundational questions.

The interpretation of Additivism is similarly tangled up with some foundational issues. Additivism is not quite what a naive understanding of "adding up" value across time and people might suggest. The values that are added up can be interpreted in many different ways and they do not have to be representable by real numbers. Unpacking the exact content of Additivism helps defuse some standard objections against an additive picture of value. But it requires delving into delicate issues concerning the use of numbers in representing value—issues left for *Section* 5.

A careful examination of these foundational issues is not just an idle exercise in rigor. The standard framework of population ethics that trades freely in numbers encourages an uncritical acceptance of the legitimacy of certain numerical operations, like holding numbers fixed while shifting boxes left or right in time. But the ability to perform these formal operations depends on some substantive assumptions about the "ontology" of value. Value is implicitly assumed to be freely recombinable—the space of value forms a rich mosaic of tiles that we can freely stack, remove, and rearrange however we like. This assumption, to be codified later in *Section* 2 as "Replacement," permeates much of population ethics.

Once it is made explicit, we will see that the status of Replacement and closely related principles are deeply tangled up with the case for Additivism. The main goal of the paper is not to provide a watertight case for Additivism—any such endeavor would require a far lengthier paper. The focus will instead be on exposing ways in which arguments for Additivism and even the interpretation of Additivism itself depend on the choice of background theory of what value consists in. Our exploration of Additivism serves as a case study illustrating a more general overarching moral that is obfuscated by the common use of numbers as stand-ins for value—the inextricability of substantive questions about what constitutes value from structural questions about how those values combine.

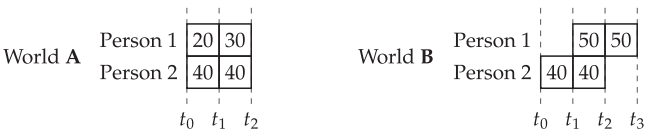
The paper proceeds as follows: *Section* 2 lays out a framework in which questions of personwise and timewise aggregation can be formulated precisely while remaining as neutral as possible on substantive questions about what exactly value consists in. *Section* 3 then introduces the first axiom of Time-Shift Invariance and provides two arguments for it. *Section* 4 does the same for the second axiom, Time-Partition Dominance. Particular emphasis will be placed on unpacking hidden assumptions about value embedded in the formulation of these principles and the arguments for them. The main formal result shows that given the framework's assumptions, these two axioms are equivalent to Additivism. Relegating the proof to the Appendix A, *Section* 5 is devoted to clarifying some subtle issues regarding the interpretation of Additivism. *Section* 6 concludes by extracting some general lessons that emerge from the preceding discussion on ways in which substantive and structural issues about value cannot always be so cleanly separated. This upshot has major methodological implications for how we conceptualize and approach the enterprise of population ethics.

2 | The Framework

Work on interpersonal aggregation often contains diagrams like the following:

World A		**World B**	
Person 1	50	Person 1	100
Person 2	80	Person 2	80

In a setting that involves not just people but time, a natural two-dimensional extension of the diagrams above are diagrams like the ones below:



Similar diagrams are found in Broome (1991), Broome (2004b) and elsewhere. But hidden in the use of such diagrams are some nontrivial assumptions about the nature and structure of value. These assumptions, though often left implicit, can seriously tip the scale in favor of particular ways of aggregating value. So, in laying out the foundational framework, it is worth proceeding slowly—assuming as little as possible from the start and properly codifying any assumptions along the way.

The first thing to note is that the diagrams above assign numbers to each person within each time interval. This assumes two things. One is that we can make sense of not just how good things are overall but how good things are for *each person* at each *fixed period* in time. Another is that the structure of value is similar

[6] This paper's key result continues in the rich tradition of deriving additive representations using separability-like principles (see Debreu (1960); Gorman (1968); Krantz et al. (1971); Wakker (2013)). Generalizations beyond real-valued representations to representations taking values in arbitrary ordered Abelian groups, applied specifically to the context of population ethics, can be found in Pivato (2014) and Thomas (2022).

enough to the structure of the real numbers that the latter can be used to faithfully represent the former. Both of these assumptions have raised suspicions. But they can be relaxed significantly.

At a very general level, the problem of value aggregation is that of deriving the value of wholes from the value of their parts. For us, the wholes of interest are *worlds*—historically complete description of all that ever happens. Now, some of the things that happen in each world are "locatable" in time and in people—they are things that can be associated with some person within some period of time or another. These are the parts of worlds I will call *life episodes*. The question of aggregating parts into wholes is then: how does a world's value depend on the value of the episodes it contains?

To expand on this notion of life episodes, they have two key features. First, they are *localizable*—each episode that is instantiated in a world is attached to some person and some time interval. The time intervals do not necessarily have to be of a fixed length. Episodes could last a millisecond or a lifetime. (This becomes important later). We could also call episodes "life events," "life stages," or "life segments"—all of which capture the localizable parts of worlds we are interested in.

But "episodes" is evocative of a second crucial feature. An episodic television series is one with a sufficiently disconnected narrative arc so that each episode is self-contained enough to be watched and judged on its own. The episodes that make up a person's life are meant to be *modular* in the same way. An episode contains everything that matters to determining its value, so that each episode's value can be evaluated in isolation from those that precede or proceed it.

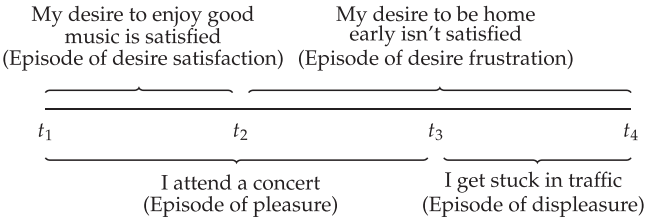
To illustrate the idea of modularity, here is an analogy. A collection of short stories is made up of individual stories. Those stories are, in turn, made up of sequences of words. In judging the literary value of a collection, there can be disagreement about how that relates to how good the stories in the collection are. Some might place more weight on holistic criteria like cohesion or thematic unity, in which case a collection could be greater or lesser than the sum of its parts.

Such disagreements notwithstanding, judgments of how good each individual story is, on its own, are intelligible. By contrast, it does not make sense to talk about how good each individual word in a short story collection is in isolation from the words that precede or follow it. The stories in a collection are standalone units of literary evaluation. The words in a collection are not. Stories but not words are modular in literary value.

Similarly, for life episodes to be modular is for there to be meaningful talk of how good they are independent of their place in relation to other episodes. No further extrinsic information is required to specify their value. Modularity is a purely structural constraint. Exactly how coarsely episodes have to be individuated for them to be modular will depend on the underlying theory of value.

I take these properties of localizability and modularity to completely characterize what I will call "episodes." Beyond these structural properties, we can remain relatively neutral on what constitutes an episode and when one episode is better than another. This leaves room for a wide range of substantive views about what is valuable.

Consider a life like the one below.



For hedonists, all that matters is pleasure and pain. So, the life can be carved into two localizable and modular bits, corresponding to the two separate hedonic episodes. One life episode spans t_1 to t_3 and the other spans t_3 to t_4. And the former (the episode is pleasure) is better than the latter (the episode of displeasure).

Similarly for desire-satisfaction theorists who care only about the satisfaction or frustration of desires. There would be two episodes corresponding to the separate instances of desire satisfaction and frustration—one from t_1 to t_2 and another from t_2 to t_4. And the former episode is better than the latter.[7]

And it does not always have to be possible to carve a life up into episodes in a nontrivial way. Sometimes, a life might consist of only a single episode. For instance, on a pluralist view where both pleasure and desire-satisfaction matter, the life above can only be decomposed into a single episode that lasts from t_1 to t_4. That is because there is no finer modular individuation of times where each period can be assigned independent value.

Hopefully, the flexibility and neutrality of the setup is clear. Episodes can occupy variably long stretches of time (possibly an entire lifetime, in some cases). This makes room for those skeptical of the possibility of comparing value at a particular instant or within fixed periods in time.[8]

And the framework can accommodate not only hedonist or desire-satisfaction theories of value. There is room for theories that value friendship, knowledge, achievement, and so on. There is even room for traditionally non-welfarist values like virtue, desert, and so on. These values can be built into whatever makes

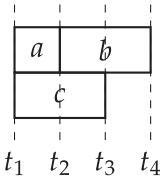
[7] Hedonists can differ on how they individuate instances of pleasure and pain. Perhaps episodes of pleasure and pain can be individuated so finely that each occupies barely a millisecond. Similar questions of individuation arise in the context of desire satisfaction too. Does the value of a satisfied desire accrue in the entire period of time between a desire being obtained and it being fulfilled? What if I want something, cease to want it, and want it again later? What if what I desire has, unbenkownst to me, already been realized before my desire was even formed? Issues to do with locating the value of satisfied desires in time are discussed by Dorsey (2013) and Sarch (2013), among others.

There are also questions to do with how to compare episodes: Is an intense basal pleasure better than a mild refined one? Is it better to satisfy a strong but uninformed desire over a weak but informed one? All of these questions have to do with the exact form of a hedonist, or desire-satisfaction, or some other kind of theory of value. They are orthogonal to the issue of aggregation and we can remain neutral on them.

[8] See Brännmark (2001), Bramble (2014, 2017), Slote (2017), King (2018), and Rosati (2021).

an episode better or worse, provided that they can be attached to some person and time interval in the same way that pleasure or desire-satisfaction can.

The life episodes across all of time for everyone who ever exists can be represented with diagrams like the one below with boxes of varying lengths:



This represents that the first person experiences episode a from t_1 to t_2 and episode b from t_2 to t_4. And the life of a second person consists of a single episode c spanning t_1 to t_3. Each such diagram corresponds to what I will call a *distribution*. A distribution is a finite collection of episodes, each of which is associated with some person and some time interval. Going forward, we can use diagrams like the one above to illustrate key concepts and ideas, leaving the precise formal descriptions for the Appendix A.

I will assume that each world can be described by a distribution. This should be fairly uncontroversial once we recall just how flexible the framework is. For instance, there can be distributions where each life can only be evaluated holistically and each consists of only a single episode. And the empty world containing no one can be described by the empty distribution.

However, the fact that every world can be described by a distribution should not be taken to imply that each world has a *unique* distribution. As we will see in *Section* 3.2, some worlds admit of descriptions by multiple distributions, since there is some arbitrariness in the choice of time coordinates.

The remainder of this section introduces three substantive assumptions which are extensions of standard assumptions in population ethics—Replacement (Section 2.1), Distributionism (*Section* 2.2), and Anonymity (Section 2.3).

2.1 | Replacement

Even if every world can be described by a distribution, not every distribution describes a world. Certain distributions might simply not be metaphysically possible. For instance, certain relationships might be necessarily symmetric. If we are the only people who ever exist, then either both of us are in a romantic relationship together at the same time or neither of us are. Suppose a and b below are life episodes that consist, among other things, of being in a romantic relationship.

Then, α might be a distribution that can be realized by some world but β is not. A romantic relationship is either doubly instantiated at the same time or not at all.

Similarly, whether some episode can be instantiated might depend on what other episodes obtain.

For instance, suppose I desire for someone in the past to have enjoyed episode c. If d is an episode that partly consists in the satisfaction of that desire, then distribution γ can be realized but δ cannot. A desire cannot be satisfied without its content obtaining.

In general, whether some episode can be had might depend on what other episodes are had, by whom, and when. So, not all combinatorially possible patterns of episodes-allocation correspond to metaphysically possible worlds.

Nevertheless, most work in population ethics that trades in numbers implicitly takes levels of goodness to be freely recombinable. Consider a diagram similar to the one in *Hasten vs. Delay*:

Suppose that in *Coincident*, how good things are for Ann and Ben is derived from their friendship. Separating them temporally in *Disjoint* deprives them of that friendship, thereby removing the original source of their welfare.

Still, it is often taken for granted that the distribution in *Disjoint* can be instantiated. The thought is that we can fix how good things are for each of them by making up the "loss" of their friendship with another source of value. This is a source that is exactly as good as being friends but that does not require their lives to overlap in time. Perhaps, it is some achievement, some amount of knowledge, or a dose of dopamine. That there is always such an equivalent source of value is a nontrivial assumption about the content of value.

Call two distributions *equivalent* if they have the same pattern of distribution but differ only in that each episode in one is replaced in the other with an equally good episode. Our qualitative framework does not assume numerical representability from the outset. In this framework, the assumption that it is always possible to make compensations like the one above can be expressed as follows:

> **Replacement**. For every distribution, there is a world with an equivalent distribution.

For instance, distribution β below might be impossible to realize because episodes a and b consist partly in being in a symmetric relationship that can only be instantiated together or not at all.

Replacement says that there is nevertheless an equivalent distribution β^* that *can* be realized. This is a distribution in which a and b are replaced, respectively, by equally good episodes a^* and b^*. These episodes, unlike a and b, can be instantiated at separate times. They contain some amount of pleasure, knowledge, achievement, or whatever else that jointly make for the same amount of value as being in a relationship.

Similarly, if d is an episode that cannot be instantiated on its own, δ might be a metaphysical impossibility.

But according to Replacement, there should be an episode d^* that is exactly as good as d that *can* be instantiated on its own so that the distribution δ^* equivalent to δ is realizable. (Going forward, the running convention would be that each episode and its primed and asterisked counterparts are equally good—that is, a, a', and a^* are equivalent, as are b, b', and b^*, and so on.)

Replacement is a kind of domain condition on the set of worlds. Though there might not be a world corresponding to every distribution, Replacement guarantees that there is at least one world for each equivalence class of distributions. Of course, what possible worlds there are is not a freely adjustable parameter—it is constrained by our best metaphysical theories. So, Replacement is better understood as a constraint on which distributions ought to *count as equivalent*. That is, in turn, a constraint on our underlying theory of value. It requires that the underlying conception of the good make for a sufficiently rich space of episodes so that for any configuration of episodes, we can find replacements for each episode to create an equivalent configuration that is metaphysically possible.

Replacement is such a widespread assumption that it is rarely ever even made explicit.[9] It is taken for granted in any work that freely makes stipulations of the kind "let this person have this level of goodness at this time, that person have that level of goodness at that time, …" without justification. But we will see over and over again that Replacement and similar principles—like Spacetime Replacement (*Section* 3.2) and Localized Choice Replacement (*Section* 4.1)—are, by no means, weak assumptions. They lay the groundwork for a strong case in favor of Additivism.

2.2 | Distributionism

The complete history of all that ever happens includes more than just facts about how well each individual's life goes. But it is common in population axiology to be presented only with facts about individual goodness and to be expected to compare different worlds solely on that basis. The implicit assumption is that there can be no difference in how good worlds are without any difference in how well individual lives go—the former supervenes on the latter.

The analogous assumption in the present context is:

Distributionism.Worlds that can be described by equivalent distributions are equally good.

Consider first the weaker assumption that worlds described by the *very same* distribution must be equally good. This precludes aspects of a world not captured by its description as a distribution, like any extratemporal or nonhuman value contained in it, from factoring into the calculation of the world's value.

For instance, a distribution does not tell us how much biodiversity or natural beauty a world contains. So, according to Distributionism, nonlocalizable value like ecological value or aesthetic value cannot affect how good a world is—except indirectly via the influence they might have on people's lives. Similarly, a distribution, in the sense introduced above, does not contain information about *where* people are located. So, according to Distributionism, spatial location cannot affect a world's value except indirectly. (This becomes important in Section 3.2.)

Note that Distributionism does not rule out the existence of what Broome calls "pattern goods" (2004b, 44)—holistic value that arises from the shape of a distribution. An example of a pattern good might be the value of equality in the distribution of episodes. An equal world has a different distribution from an unequal world. So, it is consistent with Distributionism for an unequal world to be worse.

Now, recall that *equivalent* distributions are those with the same pattern of allocation but possibly differ only in that each episode in one is replaced with an equally good episode in the other. Distributionism says that worlds are equally good not just when they share the *same* distribution but also when they have *equivalent* distributions. The idea is that for the purposes of comparing worlds, equivalent distributions can be treated as if they were the same. It matters not what each episode is exactly but only how good it is. Things are not made better or worse when an episode is replaced by one that is exactly as good. This idea is built into the standard numerical framework, since numerical levels conflates all sources of value that make for the same amount of goodness.[10]

2.3 | Anonymity

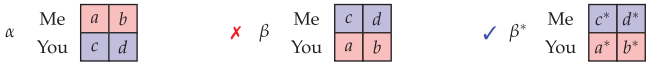
The final assumption is that it should not matter *who* has which life. If I were born into your life, going through exactly all of the things that you do from birth to death, and you mine, the world would not thereby be better or worse.

[9] Replacement can be seen as a weakening of what Broome (1991, 80–81, 115–117) calls the "rectangular field assumption." While the latter says roughly that the qualitative outcomes and alternatives are themselves freely recombinable, Replacement merely says that outcome levels or the amounts of value associated with each outcome are freely recombinable. In many contexts, Replacement is prima facie a much more plausible assumption.

[10] Believers in non-distributional value, like biodiversity and aesthetic value, will find Distributionism objectionable. For such believers, interest in what follows can still be salvaged if distributional value and non-distributional value are "separable" in that the two kinds of value can be entertained separately. If so, non-distributional value can be safely set aside. Any future occurrence of value terms like "good" or "better" can be understood as "good *qua* distribution" or "better distributionally."

Now, maybe we cannot always swap the lives of any two individuals. Perhaps the gene mutation that makes cilantro taste like soap is essential to me so that I could not be me and have the kind of life where I enjoy having cilantro daily. Perhaps episodes are things that are bound to each individual (like *Ann experiencing pleasure*) rather than things that are not (like the property of *experiencing pleasure*) so that we cannot freely assign episodes to people. Or, perhaps we want to expand our circle of moral concern to animals and there are lives that human beings can have that cows cannot.

To accommodate these and other metaphysical subtleties, let us call a sequence of episodes ordered in time a *life*. Pictorially, the lives in a distribution correspond to the rows of a diagram. The basic thought was that swapping who has which life—that is, permuting the rows of a diagram—should not change how good a world is. But because some lives might be unavailable to certain individuals, we need the notion of equivalent lives. One life is *equivalent* to another if they differ only in that each episode in one sequence is replaced by an equally good episode in the other.



So, for instance, swapping the rows in the distribution α above might result in a distribution β that is not realizable—perhaps because episodes c and d cannot be had by me. But Replacement guarantees that there is an equivalent distribution β^* that *is* realizable.

Let us call two distributions *permutation-equivalent* if one can be obtained from another by permuting the rows vertically in a diagram, while possibly replacing episodes with equally good ones. For instance, α and β^* above are permutation-equivalent. Then, our final assumption states that:

Anonymity.Worlds with permutation-equivalent distributions are equally good.

This idea that the identity of individuals does not matter should not be confused with the much more controversial idea (discussed in *Section* 4.2) that facts about personal identity and continuity or persistence over time do not matter. Anonymity allows us to permute *entire* rows vertically without thereby changing how good things are but not *parts* of rows, as in:

Anonymity, on its own, does not imply that α and γ^* have to be equally good.

Having laid out the foundational framework, we can now introduce and investigate the two axioms that jointly lead to Additivism.

3 | Time-Shift Invariance

In *Section* 1, *Hasten vs. Delay* showed how Timeslice Egalitarianism is sensitive to when one person lives relative to another. But holding fixed how much good there is within the segments of each life, a mere difference in how lives align temporally should not matter. This is the basic idea behind Time-Shift Invariance. This section spells the principle out in more detail and outlines two arguments for it, highlighting in particular the role that assumptions about what constitutes value play throughout.

The distributions below are examples of distributions that are time-shifts of each other.

Modulo the replacement of each episode with an equivalent one, the two distributions differ only in how lives align in time. In general, two distributions differ only by a time-shift roughly when one can be obtained from the other by moving rows in the diagram horizontally, possibly replacing some life episodes with equivalent ones (see Appendix A for a precise definition). Two *worlds* are then said to differ only by a time-shift when they can be described by distributions that differ only by a time-shift.

Intuitively, a pure time-shift should not affect how good things are:

Time-Shift Invariance.Worlds that differ only by a time-shift are equally good.

Of course, changing when someone lives can make things better or worse. Separate two friends in time and you deprive them of their friendship. But such knock-on effects are controlled for in the relevant kinds of time-shift. Any episode lost with a change in time has to be compensated for with an equivalent episode. It is only with Replacement that the possibility of such compensations is guaranteed.

Once the effects of a time-shift are properly controlled for, a mere change in temporal order should not matter. This sentiment echoes Parfit:

> Most of us believe that a mere difference in when something happens, if it does not affect the nature of what happens, cannot be morally significant. Certain answers to the question 'When?' are of course important. We cannot ignore the timing of events…But we aim for [a certain timing] only because of its effects. We do not believe [timing] is, as such, morally important. (Parfit 1984, 340)

Certain considerations might nevertheless seem to tell against Time-Shift Invariance. If there is intrinsic value in the longevity

of humanity's existence, then it would be better all else equal to spread lives out in time.[11] Or, the opposite would be true if, for whatever reason, there's intrinsic value in people coexisting at the same time. Strong forms of time-discounting might also favor good things being had at certain times—like earlier rather than later.[12] But Time-Shift Invariance is not just intuitive. It can be supported by two theoretical arguments—an argument from a Pareto principle (*Section* 3.1) and an argument from relativity (*Section* 3.2)—though each, again, relies on assumptions about how richly structured value is.

3.1 | The Paretian Argument

The first argument is that between two worlds that differ only by a time-shift, there is no one for whom one world is better than another. And when comparing two worlds with only the same individuals, one world cannot be better or worse unless it is better or worse *for someone*.

More precisely, recall that two lives are *equivalent* if they differ only in that each episode in one is replaced by an equivalent episode in the other. Then, plausibly:

> **Fixed-Population Pareto Equivalence**.Let α and β be distributions containing the same people. If each person's life in α is equivalent to their life in β, then worlds with distributions α and β are equally good.

The principle can be broken down into two bits. The first thought is that equivalent lives are equally good. And the second is that if two worlds contain the same people with equally good lives in either world, then the two worlds must be equally good. There can be no difference in how good worlds with a fixed population are without a difference in how well the lives in those worlds go.

Now, in any two distributions which differ only by a time-shift, the life of each person in one distribution is equivalent to their life in the other. So, Time-Shift Invariance follows from Fixed-Population Pareto Equivalence, *provided* that for any lives, there are equivalent lives located differently in time that could be lived. This presupposes Replacement.

3.2 | The Argument From Relativity

A second argument for Time-Shift Invariance arises from scientific considerations of time. In the special theory of relativity, there are not always frame-independent facts about what is simultaneous or about what precedes what. Our assessment of how good a world is often goes via picking some spacetime coordinates and representing the world as a distribution relative to those coordinates. But the choice of coordinates is somewhat arbitrary. And facts about how good things are should not vary with the choice of representation.

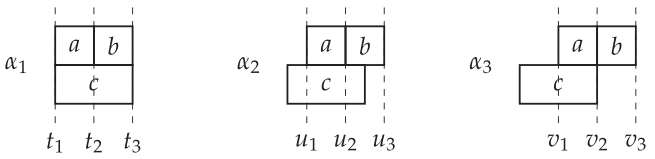
So, a first pass of the argument from relativity for Time-Shift Invariance goes as follows: How lives align in time can depend on the choice of reference frame. But changing reference frames does not change how good a world is. So, how lives align in time cannot make a moral difference. This sketch of the argument will need to be refined. But let us first visually illustrate how the argument works using spacetime diagrams.

So far, the diagrams we have used associate each episode with a person and a time interval. Spacetime diagrams like the ones below add spatial information. Each episode is associated with a spatial location (the horizontal axis), a temporal location (the vertical axis), and a person (each person's life is coded with groups of similar colors, e.g., red and pink vs. blue).[13] (For ease of visualization, we will focus on only one dimension of space instead of three.)

Heuristically, each reference frame can be associated with the point of view of a hypothetical observer. Figure 1 depicts how one and the same spacetime distribution might look relative to different reference frames. It shows how the spacetime coordinates change as the relative velocity of the observer increases. The second observer moves at a moderate velocity relative to the first. As the velocity of the observer relative to the first reference frame increases, the time and space axes gradually rotate inwards towards each other. At the limit, the axes will eventually converge on the diagonal as the relative velocity of the observer approaches the speed of light.

The spacetime diagrams show how what is simultaneous or what comes after what can depend on the choice of reference frame. Relative to the first frame, episode c begins at the same time that a begins (at t_1) and ends as b ends (at t_3). But relative to the second frame, c begins before a does (at u_1) and ends before b does (at u_3). As the velocity relative to the first frame increases, c appears to be shifted "earlier" and "earlier" compared to a and b. At the limit as the relative velocity approaches the speed of light, c ends as a ends and b begins (at v_2).

Suppressing the spatial dimension and reverting back to the diagrams we have used for most of the paper, the spacetime diagrams above correspond to the following diagrams:



Now, let **A** and **C** be worlds with distributions equivalent to α_1 and α_2, respectively. These worlds differ only by a time-shift and must therefore be equally good according to Time-Shift Invariance. Let us see how this might follow from considerations of relativity.

[11] See Frick (2017) and Scheffler (2018). For a discussion, see Fanciullo (2024).
[12] On moral reasons for time-discounting, see, for instance, Mogensen (2022).
[13] The spacetime diagrams contain some innocuous simplifications (like that a person is point-sized and has no spatial extension or that, perhaps, each person has a "center point" where we can locate the episodes attached to them).

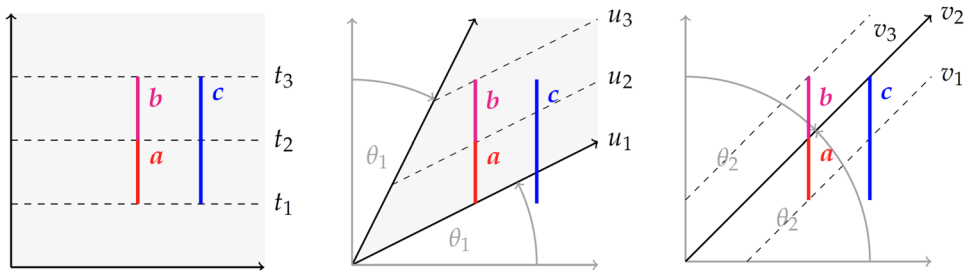


FIGURE 1 | Spacetime distribution relative to different reference frames. [Color figure can be viewed at wileyonlinelibrary.com]

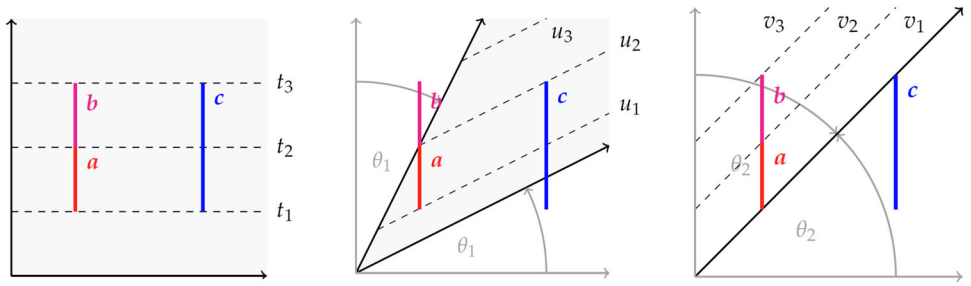
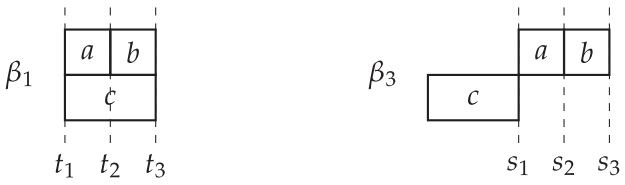


FIGURE 2 | Spacetime distribution relative to different reference frames. [Color figure can be viewed at wileyonlinelibrary.com]

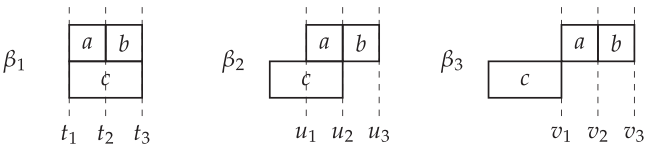
Suppose there is a world **B** with the spacetime distribution in Figure 1. (We will revisit this assumption later.) Relative to one reference frame, **B** can be described as having distribution α_1. But with a mere change in reference frame, **B** can equally well be described as having distribution α_2. So, **B** shares a distribution α_1 equivalent to **A**'s and a distribution α_2 equivalent to **C**'s. By Distributionism, worlds with equivalent distributions must be equally good. So, **A** and **B** are equally good, as are **B** and **C**. It follows by transitivity that **A** and **C** are equally good—exactly as Time-Shift Invariance requires. This is the gist of the argument from relativity. But some kinks remain to be ironed out.

One issue is that the argument, as it stands, does not yet establish Time-Shift Invariance in full generality. Time-Shift Invariance requires that a time-shift by *any* amount preserve overall value. For instance, worlds with distributions equivalent to the ones below must be equally good:



But there is no reference frame relative to which a world **B** with the spacetime distribution in Figure 1 has distribution β_3, in which c ends as early as a begins. There is a limit to how much "earlier" we can shift c relative to a and b in **B** by simply changing the reference frame. Even with a reference frame that moves at the speed of light relative to the first frame, c ends at most as early as a ends and b begins, as in α_3.

The extent of the time-shift that is possible by a mere change in reference frame depends on the lives' spatial locations. Compare the spacetime distribution in Figure 1 with the one in Figure 2, where the lives are separated further apart in space. Collapsing the spatial dimension, the spacetime distributions in Figure 2 correspond to the distributions below:



For a world with the spacetime distribution in Figure 2 (as opposed to Figure 1), there would be a reference frame relative to which it can be described in terms of a distribution β_3, where c ends as a begins. The existence of such a world would allow us to run a similar argument showing that worlds with distributions equivalent to β_1 and β_3 must be equally good.

In short, although how lives align in time can vary with the choice of reference frame, the extent of that variation is limited by how far apart the lives are in space. Considerations of relativity would support Time-Shift Invariance in full generality only given the following assumption: for any distributions α and β that differ only by a time-shift, there is always a world where the lives are sufficiently spread out such that the world has a distribution equivalent to α relative to one reference frame and a distribution equivalent to β relative to another frame. This assumption, to be codified shortly as "Spacetime Replacement," is by no means trivial.

When introducing Replacement in *Section* 2.1, we noted that certain sources of value, like friendship, might be time-sensitive. Equally, certain values might be *distance*-sensitive. Things like friendship or physical intimacy might be "proximal goods"—they cannot be had by people too far apart in space. This means

we cannot freely assume that just any spacetime distribution is instantiated by some world.

But recall the idea behind Replacement. Any "loss" of time-sensitive goods incurred by a time-shift can always be made up for with equivalent sources of value that are not similarly time-sensitive. Even if some distributions cannot be instantiated, Replacement guarantees that there are equivalent distributions that can. There is a natural spatial analog. Separating people apart in space might mean the loss of certain proximal goods. But any such loss can always be made up for with equivalent sources of value that are not distance-sensitive in the same way. Even if some spacetime distributions cannot be instantiated, there are equivalent spacetime distributions that can:

Spacetime Replacement.For every spacetime distribution, there is an equivalent spacetime distribution that is instantiated by some world.

For instance, if episode a above contains some proximal good like friendship, then there might not be a world with the spacetime distribution in Figure 2 because the lives are too far apart for a friendship to be possible. Nevertheless, Spacetime Replacement guarantees that there is an equivalent spacetime distribution—with a, b, and c replaced by some equally good episodes a', b', and c'—that is possible. The absence of friendships is compensated for by the presence of other sources of value that can be had by distantly located people.

Given transitivity and Distributionism, Spacetime Replacement suffices to entail Time-Shift Invariance. Like Replacement, Spacetime Replacement is not unassailable. Perhaps some proximal goods have no distance-insensitive substitute, in which case certain levels of goodness are necessarily unachievable by spatially isolated lives.[14]

The overarching moral of this section is that while there are strong arguments to be made for Time-Shift Invariance, these arguments—and even the formulation of Time-Shift Invariance itself—rely on replacement principles assuming that the underlying space of value is sufficiently richly structured. As we will see, the same is true of the next principle.

4 | Time-Partition Dominance

We often compare worlds not just over their complete historical timelines but also over more restricted time periods. We might say that a policy makes things worse in the *short term* but better in the *long run*. Granting the possibility of these local comparisons, it is plausible that a world that is at least as good as another in every period must be at least as good overall. And it must be better if there is also some period when it is better. We saw in *Extension vs. Creation* from *Section* 1 how some forms of Lifetime Egalitarianism run afoul of this thought. But making this principle of Time-Partition Dominance precise requires explicating what it means exactly for a world to be better or worse *within some period*. This section states the principle precisely and addresses some arguments for and against it, focusing once again on the relevance of underlying issues to do with what constitutes value.

Let us begin by assuming numerical levels of goodness and comparing two possibilities.

> *Coffee vs. Tea*. Ann could have a cup of coffee with breakfast and another after lunch. Alternatively, she could have two cups of tea instead—one with breakfast and another after lunch. Ann generally enjoys coffee more but because of her limited tolerance for caffeine, coffee has greater diminishing marginal value than tea. So, she would enjoy a first cup of coffee more than a first cup of tea but a second cup of tea more than a second cup of coffee.

Intuitively, *Tea* is better than *Coffee after t*, since Ann enjoys a second cup of tea more than a second cup of coffee. This can be understood in terms of what is better overall. We can compare worlds within some period by ignoring what happens outside of that period. For instance, we compare *Coffee* and *Tea* after t by imagining a world *Coffee-after-t* that contains 3 units of goodness for Ann after t but nothing before. Similarly, *Tea-after-t* contains 4 units of goodness for Ann after t but nothing before.

To say that *Tea* is better than *Coffee after t* is just to say that a world with a distribution like *Tea-after-t* is better than one with a distribution like *Coffee-after-t*.

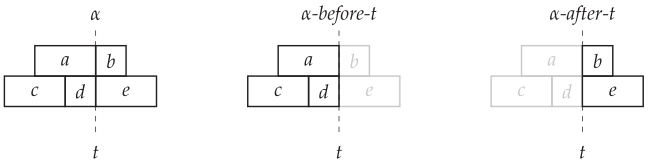
But what exactly should we picture when we try to imagine a world with distribution *Tea-after-t*? We cannot simply imagine a world where Ann simply skips her morning cup of tea. How much Ann enjoys her afternoon cup depends on whether she has already had a cup in the morning. *Tea-after-t* is supposed to hold fixed, relative to *Tea*, how good things are for Ann after t but deprive her of any prior sources of value.

To achieve this, *Tea-after-t* must be a world in which Ann undergoes something after t that is exactly as good as having a second cup of tea but that does not require her to have previously enjoyed anything of value before. (Perhaps she has just one cup of tea with an especially high caffeine content.) That

[14] Some ardent opponents of Time-Shift Invariance might insist that since mere differences in time can matter, so too can mere differences in spatial location. This would amount to rejecting Distributionism, which says that the value of a world supervenes on its two-dimensional distribution (which contains no information about spatial distribution).

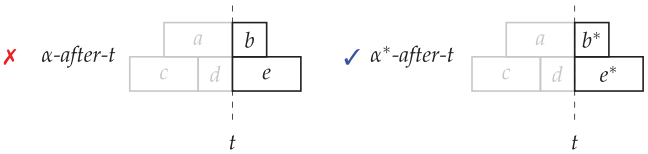
there is something fitting this description that Ann can undergo presupposes Replacement.

To formalize this idea within our framework, let us call a time a *partition* for a distribution if every episode in the distribution ends before that time or begins after it. Diagrammatically, the partitions correspond to those vertical lines that do not intersect the interior of any boxes. A partition cleanly separates the episodes in a distribution into two periods. For instance, t is a partition for distribution α below.



Then, α-*before-t* is the distribution that is exactly like α up till t and contains nothing after. And α-*after-t* is the distribution that is exactly like α after t but contains nothing before.

Even if α is a realizable distribution, α-*before-t* and α-*after-t* need not be.



Suppose episode a consists partly in the pleasure derived from having a first cup of tea and b consists partly in the pleasure from a second cup. Then, α-*after-t* describes a world in which someone enjoys a second cup of tea without ever having had a first—a metaphysical impossibility.

Nevertheless, Replacement guarantees that there is an *equivalent* distribution α^*-*after-t* that is instantiated by some world. That is a world in which episodes that become impossible to instantiate as a side-effect of "deleting" all episodes outside of some time period are made up for with equivalent episodes. And similarly with α-*before-t*.

Now, suppose worlds **A** and **B** have distributions α and β, respectively, with a common partition t. We can then define:

(i) **A** is at least as good as (or better than) **B** after t just in case a world with a distribution equivalent to α-*after-t* is at least as good as (better than) a world with a distribution equivalent to β-*after-t*;

(ii) **A** is at least as good as (or better than) **B** before t just in case a world with a distribution equivalent to α-*before-t* is at least as good as (better than) a world with a distribution equivalent to β-*before-t*.[15]

The second axiom can then be stated precisely:[16]

Time-Partition Dominance.If world **A** is at least as good as world **B** both before and after t, then **A** is at least as good as **B**. And if, furthermore, **A** is better than **B** either before or after t, then **A** is better than **B**.

Thinking of the partition as the present time, the principle roughly says that a world cannot be worse unless it is worse in the past or in the future.

The remaining subsections will focus on an argument from decision-making (*Section* 4.1) and on a controversial implication of Time-Partition Dominance (*Section* 4.2), noting in particular the under-appreciated role that replacement principles play in both cases.

4.1 | The Practical Argument From Decision-Making

When making decisions, it is sometimes practically necessary to focus only on what would change as a result of our choice and ignore the things that would remain the same regardless. Decision-making would be incredibly difficult otherwise. Fleshing out this thought gives rise to a pragmatic argument for Time-Partition Dominance, though we will see that the argument again makes some presuppositions about how rich the underlying structure of value is.

Consider a case from Nagel designed to elicit egalitarian sympathies:

> Suppose I have two children, one of which is normal and quite happy, and the other of which suffers from a painful handicap…I must decide between moving to an expensive city where the second child can receive special medical treatment and schooling…or else moving to a pleasant semi-rural suburb where the first child, who has a special interest in sports and nature, can have a free and agreeable life…the gain to the first child of moving to the suburb is substantially greater than the gain to the second child of moving to the city. If one chose to move to the city, it would be an egalitarian decision. (Nagel 1979, 123–124)

What is notable is the lack of description of how the children fared in the past. For all that is said, we could have faced many similar decisions in the past. Perhaps in all of them, the otherwise disadvantaged second child has always been favored at the expense of greater gain to the first. If so, the choice we face now might bring about something like either *City* or *Suburb*. Or

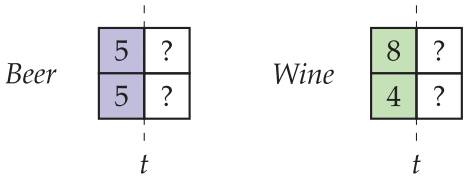
[15] More than one world might have a distribution equivalent to α-*after-t*, and similarly for β-*after-t*. But given Distributionism, the choice of worlds does not matter. If one world whose distribution is equivalent to α-*after-t* is at least as good as one world whose distribution is equivalent to β-*after-t*, then *every* world whose distribution is equivalent to α-*after-t* is at least as good as each world whose distribution is equivalent to β-*after-t*. The same applies in the case of α-*before-t* and β-*before-t*.

[16] Going forward, whenever one world is said to be at least as good or better than another before or after t, it is assumed that the worlds have distributions with t as a common partition.

maybe our previous decisions *did not* favor the second child, in which case the choice might be between something like *City** and *Suburb** instead. And so on.

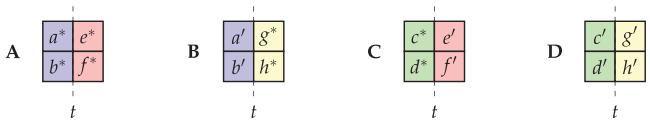
Either way, how value was distributed in the past is unaffected by the choice at hand. If the fixed past were to be ignored, then *City* is better than *Suburb* just in case *City** is better than *Suburb**.

That holds equally of the future as it does of the past. Often, the causal influence of our choice decays over time so that the choice ceases to make a difference to the distribution of value after some time. We might, for instance, be deciding between beer (which we like equally) and wine (which I like more and you less).



If the choice will have no impact on how value will be distributed past tonight, it is tempting to base the choice only on the difference in distribution in the immediate future.

More generally, consider the following worlds (recall the running convention that asterisked and primed episodes are equally good):



A and **B**'s distributions are equivalent prior to t. They differ in their distribution of value only after t. And how they differ after t is equivalent to how **C** and **D** differ. So, according to the principle to be codified shortly as Past Separability, how **A** and **B** compare and how **C** and **D** compare must align. Similarly, **A** and **C** differ in their distribution of value only before t, and that difference parallels the difference between **B** and **D**. According to Future Separability, our evaluation of the former pair must match our evaluation of the latter. More generally:

Temporal Separability. If (i) **A** and **B**'s distributions are equivalent before t; (ii) **C** and **D**'s distributions are equivalent before t; (iii) **A** and **C**'s distributions are equivalent after t; and (iv) **B** and **D**'s distributions are equivalent after t, then:

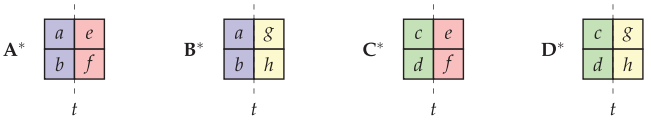
(a) (Past Separability) **A** is at least as good as **B** if and only if **C** is at least as good as **D**;

(b) (Future Separability) **A** is at least as good as **C** if and only if **B** is at least as good as **D**.

The importance of Temporal Separability is that it is logically equivalent to Time-Partition Dominance.[17] Any argument for Temporal Separability is therefore an argument for Time-Partition Dominance.

As the initial exposition suggests, the appeal of Temporal Separability stems largely from its role in streamlining decision-making. Descriptions of decision problems often neglect to specify how value would be distributed beyond the direct consequences of the choice at hand. A full description is often unrealistic and sometimes impossible. Pinning down the exact world that an action would bring about requires describing all that ever happens down to the smallest details. Temporal Separability keeps widespread decision paralysis and skepticism about what we ought to do at bay by setting reasonable informational constraints on deliberation. In its absence, even our most mundane choices could theoretically depend on information about details about the lives of ancient Egyptians or those of beings on distant planets whose lives are wholly unaffected by anything we will ever do. So goes the usual arguments for separability principles.[18]

But unpacking this argument reveals some nontrivial assumptions about the underlying structure of value. To begin, consider the following worlds (paying particular attention to how these differ from their non-asterisked counterparts above):



Here is how the argument from decision-making goes. Let $\phi_{\mathbf{X}}$ be an action that brings about world **X**. Consider a choice between $\phi_{\mathbf{A}^*}$ and $\phi_{\mathbf{B}^*}$. In common practice, our evaluation of choices depends only on what would differ as a result of picking one over another. That is to say, suppose the past had been different (c, d instead of a, b) and the choice were instead between $\phi_{\mathbf{C}^*}$ and $\phi_{\mathbf{D}^*}$. Then, whatever considerations favored $\phi_{\mathbf{A}^*}$ over $\phi_{\mathbf{B}^*}$ would equally favor $\phi_{\mathbf{C}^*}$ over $\phi_{\mathbf{D}^*}$. In either case, what differs is whether e obtains for the first person and f for the second or g and h, respectively, instead. So, we ought to $\phi_{\mathbf{A}^*}$ rather than $\phi_{\mathbf{B}^*}$ just in case we ought to $\phi_{\mathbf{C}^*}$ rather than $\phi_{\mathbf{D}^*}$. Granting that under the right conditions, we ought to choose what is best—that is, we ought to $\phi_{\mathbf{X}}$ rather than $\phi_{\mathbf{Y}}$ just in case **X** is better than **Y**—it then follows that **A*** is better than **B*** just in case **C*** is better than **D***, as Past Separability requires.

[17] If **A** is at least as good as **D** before t, then it follows from Future Separability that **A** is at least as good as **C**. And if **A** is at least as good as **D** after t, then it follows from Past Separability that **C** is at least as good as **D**. Thus, given transitivity, if **A** is at least as good as **D** both before and after t, then **A** is at least as good as **D**—exactly as Time-Partition Dominance requires. It is not difficult to show similarly that if furthermore, **A** is better than **D** either before and after t, then **A** is better than **D**. So, Temporal Separability entails Time-Partition Dominance.

Conversely, suppose **A** is at least as good as **B**. It follows from Time-Partition Dominance that **A** is at least as good as **B** after t, since **A** is exactly as good as **B** before t. This means that **C** must also be at least as good as **D** after t, since the difference in distribution between **A** and **B** after t is equivalent to the difference between **C** and **D** after t. Since **C** and **D** are equally good before t, it follows from Time-Partition Dominance that **C** is at least as good as **D**—exactly as Past Separability requires. A similar argument establishes Future Separability. So, Time-Partition Dominance entails Temporal Separability.

[18] See Parfit (1984, 420). For a discussion of separability principles, see Broome (1991), Broome (2004b), Thomas (2022), and Russell (2023).

Of course, the argument can be resisted on various fronts. Particularly contentious is the tight link it posits between what is best and what ought to done. But even setting such concerns aside, there is a gap in the argument. The argument just given regarding $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{C}^*$, and $\mathbf{D}^*$ does not apply straightforwardly to their non-asterisked counterparts. The difference is that, for instance, $\mathbf{A}^*$ and $\mathbf{B}^*$ have *exactly the same* distribution before t, whereas $\mathbf{A}$ and $\mathbf{B}$ merely share *equivalent* distributions before t (see the diagram below). It is one thing to argue that *qualitatively* unaffected periods are irrelevant and another to argue that periods that are unaffected *in value* (but possibly qualitatively affected) are irrelevant. Establishing Temporal Separability in full generality requires arguing for the latter, stronger claim.

No matter, though—provided that worlds like $\mathbf{A}^*$ to $\mathbf{D}^*$ exist, the argument does generalize. For, given Distributionism, each asterisked world above is exactly as good as its non-asterisked counterpart. So, if $\mathbf{A}^*$ is at least as good as $\mathbf{B}^*$ just in case $\mathbf{C}^*$ is at least as good as $\mathbf{D}^*$, then $\mathbf{A}$ is at least as good as $\mathbf{B}$ just in case $\mathbf{C}$ is at least as good as $\mathbf{D}$. So, it suffices to assume that for any worlds to which Temporal Separability applies, there are equivalent worlds which can figure in our choices, where the relevant periods are not just fixed in value but qualitatively fixed. More precisely:

Localized Choice Replacement. For any worlds $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ satisfying conditions (i)–(iv) in Separability, there exist worlds $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{C}^*$, $\mathbf{D}^*$ such that for all $\mathbf{X}, \mathbf{Y}$ ranging over $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$:

(a) $\mathbf{X}$ and $\mathbf{X}^*$ have equivalent distributions;

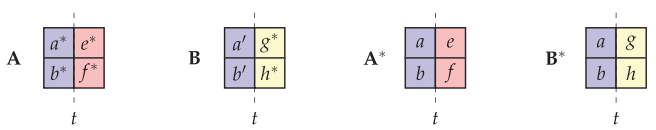
(b) if $\mathbf{X}$ and $\mathbf{Y}$ share *equivalent* distributions before/after t, then $\mathbf{X}^*$ and $\mathbf{Y}^*$ have the *same* distribution before/after t; and

(c) there are decision problems involving a choice between $\phi_{\mathbf{X}^*}$ and $\phi_{\mathbf{Y}^*}$.

As with the original Replacement principle, what worlds exist should be treated as independently fixed by our metaphysical theories. Localized Choice Replacement is best understood not as a constraint on what worlds exist or on what choices we can face but as an assumption about how rich the possible sources of value are, which then affect which distributions count as equivalent.

In some respect, the assumption can be seen as dual to the original Replacement principle. According to Replacement, even if certain qualitative sources of value are locked into rigid patterns of instantiation, *levels* of value are not similarly locked. Localized Choice Replacement, on the other hand, assumes roughly that for any levels of value, there are qualitative sources from which those levels can be derived that are themselves freely recombinable. At least some representatives from each equivalence class of episodes are amenable to free mixing-and-matching.

For instance, the levels of value before t are fixed across $\mathbf{A}$ and $\mathbf{B}$:

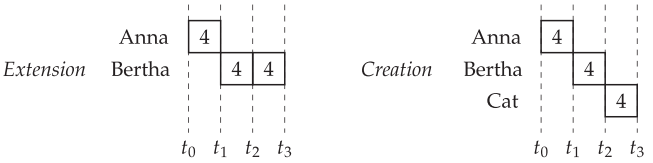


According to Localized Choice Replacement, it is possible for those fixed levels to have been derived from exactly the same episodes (a instead of a^* and a', b instead of b^* and b'). Furthermore, these episodes are ones that can be freely paired with e and f (episodes exactly as good as e^*, e' and f^*, f', respectively) as well as with g and h (episodes exactly as good as g^*, g' and h^*, h', respectively), as in $\mathbf{A}^*$ and $\mathbf{B}^*$. These are not trivial assumptions. This precludes pictures on which the goods achievable at some time is deeply constrained by what goods have or will be obtained at other times.

And even if some sources of value are freely recombinable in the way required, the argument from decision-making would still be inapplicable unless it were possible to face a choice between bringing about worlds like $\mathbf{A}^*$ and $\mathbf{B}^*$. That imposes further constraints on value. For instance, it could be that, due to the nature of the episodes a and b, any pair of worlds like $\mathbf{A}^*$ and $\mathbf{B}^*$ are such that its two inhabitants are separated far apart in space—so far apart that no action of a single agent could possibly affect how things go for both people at once. In that case, the very first step of the argument from decision-making—*consider a choice between* $\phi_{\mathbf{A}^*}$ and $\phi_{\mathbf{B}^*}$—would be a non-starter. Thus, among other things, the argument requires that value is not overly sensitive to factors, like spatial location, that affect whether a choice between the relevant worlds is possible.

4.2 | Extension-Creation Neutrality

Time-Partition Dominance has some uncomfortable implications in choices involving the prolonging of an existing life and the creation of a new one. Recall the case of *Extension vs. Creation*. Bertha can have a child but only by prematurely cutting her own life short:



Certain egalitarians might hypothetically prefer *Creation*, on account of the inequality in total lifetime welfare in *Extension*. While that seems objectionable, some might find it equally objectionable that theories like Total Utilitarianism are neutral between the two. Intuitively, *Extension* is better.

More generally, consider distributions that differ only in whether the episodes are had by a single person or by different people (*modulo* the possible substitution of episodes with equally good ones):

Call distributions (and corresponding worlds) that differ only in this respect *extension-creation equivalent*. Total Utilitarianism—

and additive theories more generally—are insensitive to where value is located and therefore indifferent between extension-creation equivalent worlds:

Extension-Creation Neutrality.Extension-creation equivalent worlds are equally good.

This indifference has been the target of much criticism. As Singer puts the objection, "[on total utilitarianism, it] is as if sentient beings are receptacles of something valuable and it does not matter if a receptacle gets broken, so long as there is another receptacle to which the contents can be transferred without any getting spilt" (1993, 121).

Such criticisms run deep, it turns out. Extension-Creation Neutrality is a key ingredient of an additive theory of value. As we will see in *Section* 5, given weak background assumptions, Extension-Creation Neutrality and Time-Partition Dominance are logically equivalent. And some of the most pressing objections against Additivism are targeted at Extension-Creation Neutrality. So, it is worth spending some time exploring the principle. In keeping with the general moral of the paper, we will see how it is difficult to disentangle the status of Extension-Creation Neutrality from foundational issues concerning value's fundamental constituents. Often, apparent reasons for rejecting Extension-Creation Neutrality can instead be seen as cause for rejecting Replacement.

Take longevity. According to Broome, "[f]or longevity to be valuable means that, given [a] fixed total of time and level of temporal wellbeing, it is better for the time to be divided up amongst fewer lives rather than amongst more lives" (2004b, 108).[19] If longer lives (as in α above) are, all else equal, better than two shorter lives (as in β), that would seem to contradict Extension-Creation Neutrality. But the culprit might instead be Replacement. Consider first the *instrumental* value of longevity. Many good things in life typically come with age—self-sufficiency, self-actualization, intellectual fulfillment, creative accomplishment. To shorten a life would be to deprive a life of these "maturity goods." It should be familiar by now that any such loss that accompanies a change in distribution must be compensated for before Extension-Creation Neutrality is applicable.

Of course, perhaps such compensation is not always possible. Perhaps the wisdom that accumulates from lived experience has no early-life surrogate. Perhaps achievement's value necessarily scales in proportion with how long one works towards it. If so, then splitting a long life into shorter ones results in lost value that cannot be recovered and Replacement fails. Consider:

Splitting a long life (α) into shorter ones (γ) might mean the loss of various maturity goods, making b^- worse than b. If there is no way to "top up" the difference to obtain an episode b' that is exactly as good as b, then, *contra* Replacement, no distribution equivalent to β is realizable. As a corollary to Replacement's failure, Extension-Creation Neutrality might fail too. But if it fails, it fails at most in the sense of a presupposition failure—without Replacement, some worlds might have no extension-creation equivalent.

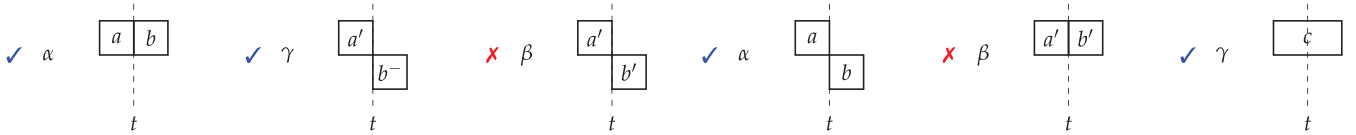
The *instrumental* value of longevity thus poses no direct threat to an additive theory of value. This much is widely recognized. For this reason, detractors of an additive theory tend to emphasize longevity's *intrinsic* value. Even then, it might still be Replacement rather than Extension-Creation Neutrality that ought to be rejected, though the dialectic is more subtle here.

Recall that our framework was set up to be neutral about what the fundamental constituents of value are. To that end, the basic units of evaluation—episodes—were characterized abstractly in terms of two structural constraints: localizability and modularity. In particular, modularity requires that the value of episodes be self-contained. This means carving episodes up in a way that renders their value intelligible independent of the distributions in which they are embedded.

What this constraint amounts to will depend on what exactly constitutes value. Hedonists might individuate episodes finely into short bouts of pleasure and pain. Others might find this myopic—it is like judging the individual words of a story. For instance, suppose there is value in achievement and the value varies with whether it is accomplished through hard work or by mere luck. Given modularity, lives then have to be segmented into coarse episodes that encompass any relevant period of strife. On an overly fine individuation, achievements would be freely detachable from what led up to it and conjoined to alternate histories, as if an achievement's value were independent of *how* it came about.

It is analogous with longevity. If longevity is fundamentally valuable, then the length of a life is a constitutive part of how well things go for a person. In that case, the value of an experience or event cannot be assessed in isolation from whether it is embedded within a long life or a short one—in the same way that the value of a goal fulfilled cannot be determined independently of what came before. Modularity would then require a "lumpier" individuation on which episodes extend across longer—perhaps, lifetime spanning—stretches of time.

The resulting picture is one in which Replacement fails. Even if some world instantiates distribution α below, *contra* Replacement, there might not be any that instantiates an extension-creation equivalent distribution β.



[19] The value of longevity is Broome's (2004b, Ch. 7) main reason for rejecting the separability of value across time and thus *ipso facto*, Time-Partition Dominance.

That is because, where longevity matters, a' and b' cannot be valued independently of the fact that one precedes the other within a single life. Given modularity, the episodes are "magnetic"—any two episodes (as in α) that we attempt to conjoin into a single life (as in β) will automatically snap together and coalesce into a single episode (as in γ). Extension-Creation Neutrality requires that a world with distribution α be exactly as good as a world with distribution β, *if there were such a world*. But it says nothing about how worlds with distribution α and γ compare. So, any intrinsic value that longevity might have could be taken as cause for rejecting Replacement, rather than Extension-Creation Neutrality.

Now, this approach of "localizing" the value of longevity by building it into how episodes are individuated is nonstandard and might thus invite apprehension. After all, similar strategies of dispersing the value of longevity into the individual temporal segments of a life have been found wanting (Broome 2004b). But problems with such dispersion strategies mostly arise from trying to cram longevity's value into predefined partitions of time into fixed intervals. The present framework, unlike standard frameworks like Broome's, is flexible in that it allows episodes to span *variably* long periods of time.

The extra flexibility is, in any case, unavoidable once we choose to admit a broad range of possible values. For, the problems with dispersion strategies arise not just in the context of longevity. Take friendship. It may not always be possible to decompose a friendship's value and determine how much of its value accrues within a given year, month, hour, or minute. Other goods, like achievement and desire-satisfaction, also plausibly resist decomposition into fixed intervals. Accommodating such goods with variable temporal extensions thus already calls for a departure from the fixed-interval framework. And once the framework is flexible enough to localize the value of achievement, friendship, and so on, then why not also longevity? If longevity has intrinsic value, then it seems to be a determinant of how good a life goes in exactly the same way that other commonly recognized sources of value are. Any differential treatment would require justification.

In my view, attempting to legislate what can or cannot be built into episodes (beyond structural constraints like localizability and modularity) will not be a fruitful endeavor. A more productive route is to recognize that being *able* to localize the value of some good, like longevity, does not mean that we always *must*. Certain considerations can sometimes call for the exclusion of certain sources of value from the episodes.

To explain, the discussions throughout the paper play out against the backdrop of some underlying theory of value. This theory is assumed to be fixed but not assumed to be fixed one way rather than another. We assume in the background some account of what constitutes value but we do not assume it to be, say, hedonism, or desire-satisfactionism, or some particular brand of an objective-list theory. But of course, *exactly how* the background theory is fixed can render some principle or view more plausible or less so. Additivism might be objectionable given hedonism but less so, the more expansive the background theory, and the more non-hedonistic considerations that are allowed to be incorporated into the episodes whose values are to be added up.

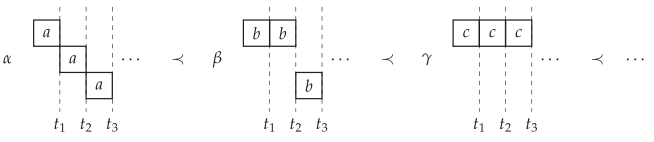
Thus, our evaluation of each principle and view is always relative to some background theory or another. And there is sometimes interest in evaluating principles relative to a background theory that is an *expansion* or *restriction* of one's preferred theory. For instance, we have seen how the value of relational goods like friendship might lead to the failure of Replacement. Even so, it can be useful to ask: Would a background theory that is restricted to the nonrelational goods validate Replacement? *Modulo* the relational goods, is value additive?

Equally, we have seen how localizing the value of longevity results in the failure of Replacement. Acknowledging this does not immediately negate interest in assessing the state of play relative to a background theory that excludes longevity from the initial bundle of goods used to individuate episodes. There can be good reason for doing so. One reason is that frameworks in distributive ethics, like ours, are designed to probe how a world's value varies with the value of its constituent episodes. In such frameworks, episodes are black boxes. Nothing can be said about how much of an episode's value comes from the friendship it contains, how much from achievement, and so on. The different sources of value are simply assumed to congeal in some way.

For goods whose value we wish to subject to individual scrutiny, leaving them out of the episodes can thus be a useful modeling choice. This does not work with all goods. Goods like friendship, achievement, or knowledge are either reflected in the episodes or not in the distribution at all. But longevity is a "pattern good." Information about pattern goods can be read off of the shape of a distribution, whether or not their value is built into the episodes. By separating longevity or other pattern goods from goods like pleasure, friendship, achievement, and so on, we can isolate out longevity's individual contribution to overall value and investigate its interaction with other components of value.

It is surprising then that while much has been said about whether longevity has intrinsic value, much less has been said about how any alleged value is to be traded off against other goods. Much work remains to be done here. I will briefly mention two issues that must be confronted by proponents of longevity's intrinsic value.

First, consider a world with lives stacked back-to-back in time (α).[20]



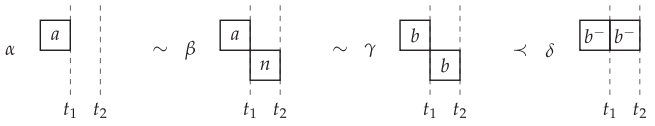
Suppose that the lives could be conjoined at some cost. (Perhaps, more resources are required to sustain a life at some fixed standard the older one gets). Suppose any cost would be borne equally across time.

[20] Going forward, the episodes are assumed to not encode any intrinsic value that longevity might have.

If longevity has value, then some sufficiently small cost would be worth bearing. That is, for some episode b that is worse than a by a small enough margin, a β-world would be better than the α-world. By the same reasoning, a γ-world, with c marginally worse than b, would be better still. Iterating, the eventual conclusion is that a single long life is better than multiple generations of lives, even when the long life is worse at all times. Some ways of filling in the details make the conclusion especially troubling. Suppose a is neutral, making worse episodes like $b, c, \ldots$ bad. The conclusion above would then be that it is worse to have many *neutral* lives than a single long life that is *bad at all times*.

Take another case. Total Utilitarianism has been much maligned for preferring a long life with greater total value across time to a shorter life of much higher *average* quality. This is a temporal analogue of the Repugnant Conclusion (see Section 5.2). Those who value longevity must contend with an even more repugnant implication—a long life that is worse *on average* but also *in total* can be better than a shorter life.

To see that, consider a world with distribution α below, where a is good. Plausibly, adding a neutral life (as in β) should not change how good things are. (One might even define a neutral life as precisely one with this property.)

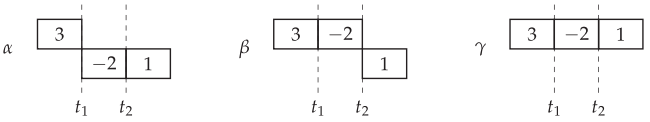


Suppose that the difference in how good the two lives in β are can be equalized. That is, worlds with distributions β and γ are equally good, for some episode b.[21] If longevity is intrinsically valuable, then conjoining the two lives in γ would be an improvement, even at the cost of a slight worsening of the episodes from b to b^- (as in δ). By transitivity, the δ-world is better than the α-world.

That seems objectionable. On any reasonable cardinalization (with the value of neutrality pegged at zero), the value of b is *at most half* that of a. If equality across temporally disjoint lives matters, then b is less than half as good as a; if it does not, then b is exactly half as good. Either way, b^- would be less than half as good as a. Thus, the life in δ is worse than the one in α both on average and in total. But the proponent of longevity's value, it seems, should prefer δ. Worse, the reasoning iterates—appending another neutral life to δ eventually leads to a life that is longer still and worse still along all other dimensions but that is, by the lights of those who value longevity, better still.

These problems are not necessarily insurmountable.[22] But they do emphasize the need for a more sophisticated story about how exactly longevity is valuable if it is and how gains in longevity are to be balanced against losses elsewhere. Developing such a theory is where I think future efforts investigating the value of longevity are most productively directed.

By way of a summary, it will be instructive to see how the dialectic concerning longevity carries over to other grounds for doubting Extension-Creation Neutrality. Granting for a moment the use of numbers for representing value, compare:



Those who think that lives with positive trajectories are, all else equal, better are apt to prefer α-worlds to β-worlds.[23] Whether a bad outcome is appended to the beginning of a life or to the end of another can be the difference between an upwards-trending life (α) and a downwards-trending one (β). Conversely, some think that extra value can be found in a life being worth living over and above the value of its constituent goods. Such people are apt to prefer β-worlds to α-worlds. Only α contains a life with overall negative value. Yet others think that there is some positive value to life itself—even one with filled only with bad experiences.[24] In that case, worlds with larger populations (like α and β) are, all else equal, better than smaller worlds (like γ). All of these are potential grounds for rejecting Extension-Creation Neutrality, since α and γ are extension-creation equivalent, as are β and γ. So, given transitivity and Extension-Creation Neutrality, worlds with any of the distributions above must be equally good.

But as with longevity, the exact upshot of taking seriously the value of a positive life trajectory, of a life worth living, or of existence is a bit subtle. It depends on whether such values are built into how episodes are individuated. On the one hand, *progressing upwards* or *being overall worth living* are features of a life that determine how good a life is, in the same way that *having fulfilling friendships* or *containing creative achievements* do. So, by default, all such values should be treated on a par. If friendships and achievements enter into considerations on how we decompose a life into modular parts, then so too should holistic determinants of value like a life's shape and its overall valence. From this point of view, the problem lies primarily with Replacement, rather than Extension-Creation Neutrality. By the lights of a background theory that countenances the importance of holistic features of a life, modularity precludes the division of a life into overly short chunks. Episodes cannot be freely detached and reattached across lives in the way that would be possible given Replacement.

[21] Alternatively: two equivalent episodes b_1 and b_2.

[22] Some of the assumptions at play in the cases above that might be contested include: richness assumptions (the space of episodes are sufficiently "dense" with many close enough in value); continuity assumptions (small changes in the value of episodes correspond to small changes in the world's value); separability assumptions (if a long life is better than two shorter ones, then a world with the former and some background population is better than a world with the latter and the same background population); assumptions about longevity (if longevity has value, then its value is not contingent on the goodness of the life extension, on the goodness of the resulting life as a whole, and so on).

[23] See Velleman (1991).

[24] For instance, Nagel (1970, 2) writes: "There are elements which, if added to one's experience, make life better; there are other elements which…make life worse. But what remains when these are set aside is not merely neutral: it is emphatically positive. Therefore life is worth living even when the bad elements of experience are plentiful, and the good ones too meager to outweigh the bad ones on their own. The additional positive weight is supplied by experience itself." See also Lee (2018) and the surrounding literature concerning the value of consciousness.

On the other hand, there is no principled barrier to restricting the class of goods used to determine how episodes are to be individuated. A non–ad hoc justification for including only nonpattern goods is that doing so would allow us to isolate out the individual contribution of pattern goods to overall value and investigate their interaction with nonpattern goods. On this picture, holistic features of a life like its trajectory and valence do not figure into how episodes are partitioned and they may thus pose genuine threats to Extension-Creation Neutrality. Still, a thorough investigation into how the alleged value of such holistic features are to be balanced against other components of value has largely been neglected, with some preliminary investigations already unearthing difficult issues that may eventually cast doubt on such holistic features having genuine value after all.[25]

5 | Additivism

We saw in the introduction how some simple forms of Egalitarianism might run into problems with at least one of Time-Shift Invariance and Time-Partition Dominance. We also noted that the same is true of various forms of Averagism. That is no accident—given the framework and its background assumptions, the only kind of theory that satisfies both principles is an additive one. The aim of this section is to state this result precisely and unpack what it means exactly for value to be additive.

To begin, value is often represented using real numbers. But this is overly restrictive. As we will see, familiar problems associated with an additive picture of value stem from properties of the real numbers—properties which are not shared by more general structures, in which we can also add and compare things in relatively well-behaved ways.

For example, we might represent how good things are using *pairs* of real numbers. Any two pairs can be added up by adding each component individually: $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$. To compare pairs, there are at least two possibilities. One is the *lexical* ordering. We begin by comparing the first components: if $x_1 > y_1$, then $(x_1, x_2) > (y_1, y_2)$. If the first components are equal, we move on to compare the second: if $x_1 = y_1$ and $x_2 \geq y_2$, then $(x_1, x_2) \geq (y_1, y_2)$. Another way to compare pairs is using the *dominance* ordering, where one pair is no less than another just in case it is no less along both components. That is, $(x_1, x_2) \geq (y_1, y_2)$ if and only if $x_1 \geq y_1$ and $x_2 \geq y_2$.

Pairs can be generalized naturally to triples, quadruples, and arbitrarily long and possibly infinite sequences of real numbers. Adding sequences up and comparing them in the ways defined above result in what I will call *lexical spaces* and *dominance spaces*. These are examples of *partially ordered Abelian groups*, which are general structures that share many of the real numbers' desirable mathematical properties while possibly lacking some of its overly restrictive features (see *Section* 5.3). Instead of restricting ourselves to the real numbers, we can allow value to be represented using any partially ordered Abelian group.

Call a reflexive and transitive ordering $\succcurlyeq$ on the set of worlds an *axiology*, where $\mathbf{A} \succcurlyeq \mathbf{B}$ means that world **A** is at least as good as world **B**. Throughout, we have also assumed background facts about which episodes are equally good—for instance, in the notion of equivalent distributions. Let us call a function V from worlds and episodes to a partially ordered Abelian group a *representation* of an axiology if it assigns better worlds greater values and vice versa—that is, for all worlds **A** and **B**:

$$\mathbf{A} \succcurlyeq \mathbf{B} \qquad \text{if and only if} \qquad V(\mathbf{A}) \geqslant V(\mathbf{B}),$$

and it assigns equally good episodes equal value. A representation V is *additive* if whenever world **A** has a distribution that contains episodes $a_1, \ldots, a_n$, then:

$$V(\mathbf{A}) \;=\; V(a_1) + \ldots + V(a_n).$$

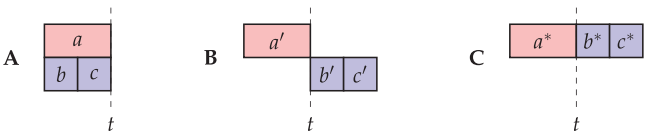
An *additive axiology* is one that has an additive representation. Additivism is the class of additive axiologies.

The main formal result of this paper is that given our framework and its background assumptions, Additivism is equivalent to the two axioms defended in the previous sections:

> **Theorem.** *Given Replacement, Distributionism, and Anonymity, an axiology satisfies Time-Shift Invariance and Time-Partition Dominance if and only if it is additive.*

The proof is a little involved and is left for the Appendix A. But the gist of it is not difficult to grasp.

It would be helpful to first establish the claim made earlier that Extension-Creation Neutrality and Time-Partition Dominance are equivalent. To illustrate, consider worlds with the distributions below:



Extension-Creation Neutrality requires that **B** and **C** be equally good. It is easy to see that this follows from Time-Partition Dominance. **B** and **C** are equally good before t. Given Anonymity, they are also equally good after t. Therefore, they must be equally good according to Time-Partition Dominance. In general, given Anonymity, Time-Partition Dominance entails Extension-Creation Neutrality.

The converse entailment also holds given some weak assumptions. First, note that given Time-Shift Invariance and Extension-Creation Neutrality, every world can be "collapsed" into an equally good *solitary world*—a world in which at most one person ever lives. For instance, Time-Shift Invariance implies that **A** and **B** above are equally good. And Extension-Creation Neutrality implies that **B** and **C** are equally good. Therefore, **A** is exactly as good as the solitary world **C**. This generalizes. Time-Shift Invariance allows us to stack multiple lives up back-to-back

[25] See, for instance, Karhu (2019) and Lee (2023).

in time and Extension-Creation Neutrality then allows us to agglomerate those lives into one long life without changing how good things are.

The equivalence of every world to some solitary world means that no loss in generality would result from restricting the scope of various principles to just the solitary worlds. In particular, the restriction of Time-Partition Dominance to solitary worlds can be decomposed into two plausible principles. The first says that a life cannot be better than another unless it is better at some time:

Life Monotonicity.A life that's at least as good both before and after t is at least as good. And if, furthermore, it's better either before or after t, then it's better.

The second says that how good a solitary world is just depends on how good the single life it contains is:

Solitary World Supervenience.For solitary worlds **A** and **B**, **A** is at least as good as **B** just in case the life in **A** is at least as good as the life in **B**.

(This is a variant of the Fixed-Population Pareto Equivalence principle introduced in Section 3.1—weakened to range only over solitary worlds but also strengthened in the substitution of "equally good" for "at least as good"). Given Time-Shift Invariance, Monotonicity, and Solitary World Supervenience, Extension-Creation Neutrality implies Time-Partition Dominance.

This intermediary observation that Time-Shift Invariance and Extension-Creation Neutrality imply that each world has a solitary equivalent is an important one. It gets us halfway toward the main theorem establishing Additivism. For solitary worlds, the question of interpersonal aggregation is trivial, since there is at most one person. So, given Time-Shift Invariance and Time-Partition Dominance (which implies Extension-Creation Neutrality), the question of how to aggregate across time and people reduces to the question of how to aggregate across just time.

To derive Additivism, all that remains is to show that the two axioms imply that timewise aggregation must be additive. This is proved in the Appendix. Without going into too much detail, the strategy is to define on (equivalence classes of) solitary distributions an addition operation and an ordering. Roughly, addition corresponds to the concatenation of lives in time and the ordering is derived from the axiology's ranking of worlds. The resulting structure can then be used to represent the value of worlds and episodes. The two key axioms guarantee that the addition operation and ordering, as defined, are well-behaved—in particular, that they satisfy the properties of a partially ordered Abelian group.

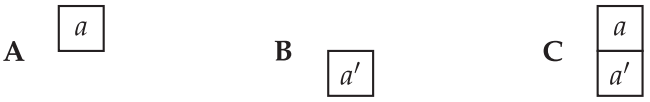
The remainder of this section is devoted to unpacking the exact content of Additivism. Doing so will help clarify the upshot of the theorem and defuse some possible worries.

5.1 | Uniqueness of Representation

First, recall that an additive axiology is one with an additive representation. This does not preclude an additive axiology from also having nonadditive representations. To illustrate, suppose V is a real-valued representation for an axiology. Then, so is the function V^2 which assigns to each world and episode the square of its value on V, since, for any real numbers x and y, x is greater than y just in case x^2 is greater than y^2. But V^2 need not be additive even if V is, since $x + y = z$ does not generally imply that $x^2 + y^2 = z^2$. In that case, V^2 would be a *nonadditive* representation of an *additive* axiology.

The nonuniqueness of a representation makes identifying an additive axiology less straightforward than one might expect. In the standard framework that assumes numbers from the outset, axiologies are often introduced via functional equations. For instance, Lifetime Averagism can be described as the ordering of worlds obtained as follows: First, we sum up the value of the episodes in a world's distribution. Then, we average that sum by the total number of people who ever exist. Better worlds are those with greater averages. But this is just one method of arriving at which worlds Lifetime Averagism deems to be better or worse. What is to say that the same ranking of worlds does not also have an additive representation?

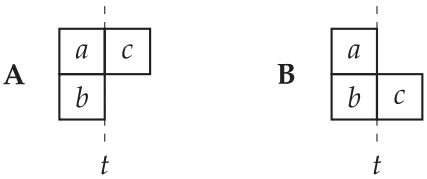
To see why not, consider three worlds with the distributions below:



A and **B** each contains only one person whose life consists of a single episode, a in one case and a' in the other. Assume that these are equally good episodes. And **C** contains both of the lives in **A** and **B**. According to Lifetime Averagism, **C** must be exactly as good as **A** and **B**.

Now, suppose V is a representation of Lifetime Averagism, in which case $V(\mathbf{A}) = V(\mathbf{B}) = V(\mathbf{C})$. If V were an additive representation, then it must be that $V(\mathbf{A}) + V(\mathbf{B}) = V(\mathbf{C})$. It is familiar in the case of real numbers that $x + x = x$ implies that $x = 0$. This is also a feature of any partially ordered Abelian group more generally. So, $V(\mathbf{A}) = V(\mathbf{B}) = V(\mathbf{C})$ and $V(\mathbf{A}) + V(\mathbf{B}) = V(\mathbf{C})$ imply that $V(\mathbf{A}) = V(\mathbf{B}) = V(\mathbf{C}) = 0$. So, Lifetime Averagism is additive only in the trivial case where every world is exactly as good as the empty world in which no one exists.

Or, consider Lifetime Egalitarianism. How do we know that it does not have an additive representation? To see how, consider worlds like the ones below:



Suppose that c is a positive episode (i.e., a world whose distribution is equivalent to one that contains only c is better than the empty world). For Lifetime Egalitarians, it is better for c to be had by whomever had the worse life prior to t. So, if a is better than b, then **B** is better than **A**.

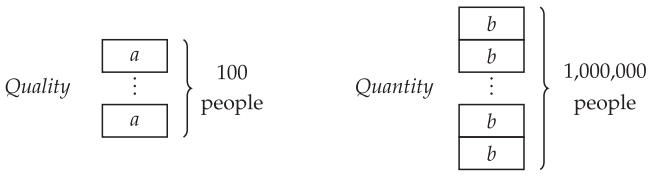
But it is clear that cannot be the case on an additive axiology, since on any additive representation, $V(\mathbf{A}) = V(\mathbf{B}) = V(a) + V(b) + V(c)$. So, Lifetime Egalitarianism is not additive. More generally, any two worlds with distributions that have exactly the same episodes must be equally good according to an additive axiology. So, Additivism rules out axiologies, like Egalitarianism, that are sensitive to pattern goods.

So, while Additivism simply requires that the true axiology have *some* additive representation—and not that every representation of it be additive—this is nevertheless a genuine constraint that rules out axiologies that are standardly thought to be nonadditive. At the same time, Additivism allows for more than just the kinds of view usually associated with labels like "Total Utilitarianism" or "Classical Utilitarianism." The ways in which Additivism is more general open up avenues for resisting some standard worries about an additive picture of value.

5.2 | Cardinal Comparisons

Total Utilitarianism is often criticized for allowing quantity to trump quality, as in:

> *Repugnant Conclusion* (Parfit 1984). Suppose that a lifetime of listening to Mozart (a) is a thousand times better than a lifetime of listening to Muzak (b)—a life that, let us suppose, would be barely worth living. Consider a world, *Quality*, in which a hundred people spend their lives listening to Mozart. Compare that to a world, *Quantity*, in which a million people spend their lives listening to Muzak. Intuitively, *Quality* is better than *Quantity*.



The problem for an additive theory is supposed to be that it implies otherwise. The thought is that if the value of b is x, then the value of a, which is a thousand times as good, must be $1000x$. But adding $1000x$ a *hundred* times over still amounts to less than adding x a *million* times over.

But what exactly does it mean for a lifetime of listening to Mozart (henceforth, a "Mozart-life") to be *a thousand times better* than a lifetime of listening to Muzak (henceforth, a "Muzak-life")? There are three salient possibilities—none of which results in an interpretation of the Repugnant Conclusion that poses a problem for Additivism.

One possibility is that cardinal facts are derived from how we make interpersonal trade-offs:

> A Mozart-life is 1000 times better $_{\text{personwise}}$ than a Muzak-life = A world with one Mozart-life is exactly as good as a world with a thousand Muzak-lives.

On this picture, cardinal comparisons simply encode facts about how many lives of a worse kind it takes to make the world as good as one with a single better life. This renders the Repugnant Conclusion almost trivial. Take the following "replication invariance" principle: if a world with one Mozart-life is exactly as good as a world with 1000 Muzak-lives, then a world with N many Mozart-lives is exactly as good as a world with $1000 \times N$ many Muzak-lives, for any natural number N. This principle is highly plausible and is satisfied by most axiologies.[26] Given the principle, the "Repugnant" Conclusion follows from the very definition of what it means for a one life to be a thousand times better.

A second possibility is that cardinal facts are derived from some independent dimension, like the strength of a pleasure or desire, or trade-offs under risk. Perhaps there is a natural cardinal scale for measuring the strength of pleasure (by the concentration of dopamine, say). A hedonist might then define:

> A Mozart-life is 1000 times better $_{\text{strengthwise}}$ than a Muzak-life = The pleasure contained in a Mozart-life is 1000 times as strong as that contained in a Muzak-life.

Or, cardinal comparisons could track considerations of risk:

> A Mozart-life is 1000 times better $_{\text{riskwise}}$ than a Muzak-life = A one-in-a-thousand chance of a Mozart-life is exactly as good as a Muzak-life for sure.

Neither of these interpretations render the Repugnant Conclusion trivial.

However, on these interpretations, Additivism does not imply that *Quantity* is better than *Quality*. Recall that an additive axiology is simply one with an additive representation, on which value can be thought of as added up over time and people. Nothing in this definition requires that the values assigned by an additive representation also be the ones that track hedonic strength or risk management. Absent any further argument, there is no requirement in Additivism that the personwise, strengthwise, and riskwise interpretations above coincide.

So, as it stands, either cardinal comparisons arise from interpersonal trade-offs, in which case the "Repugnant" Conclusion is trivial and unobjectionable. Or, they arise from some independent scale, in which case the Repugnant Conclusion is not actually a consequence of Additivism.

[26] See Blackorby et al. (2005, Ch. 4).

But there is a third possibility. Cardinal facts could be understood in terms of how value is aggregated across time:

> A Mozart-life is 1000 times better $_{\text{timewise}}$ than a Muzak-life = a Mozart-life is exactly as good as a life that consists of 1000 episodes (each equivalent to a Muzak-life) back-to-back.

Additivism does require a parity between the dimensions of people and time. For Additivists, the personwise and timewise definitions of cardinal comparisons must coincide. So, this interpretation gives rise to a version of the Repugnant Conclusion that is both nontrivial and also a genuine implication of Additivism.

But the interpretation fails to retain the intuitive force of the Repugnant Conclusion. Most of those averse to the Repugnant Conclusion would be equally troubled by its temporal analogue:

> *Temporal Repugnant Conclusion*. Consider a world, *Quality**, which consists of a single life spent listening to Mozart. Compare that to a world, *Quantity**, which consists of a single life, that's a thousand times as long, made up of a thousand episodes back-to-back, each exactly as good as listening to Muzak. Intuitively, *Quality** is better than *Quantity**.

Indeed, cases like these have also been used to argue against forms of Total Utilitarianism.[27] But to accept that *Quality** is better than *Quantity** is precisely to reject the supposition that a Mozart-life is 1000 times better $_{\text{timewise}}$ than a Muzak-life. So, this version of the Repugnant Conclusion on which cardinal comparisons are interpreted temporally would have dialectical force against Additivism only for those who are troubled by the interpersonal Repugnant Conclusion but somehow untroubled by its temporal analogue—an unlikely combination of intuitions.[28]

So far, we have focused on a very specific form of the Repugnant Conclusion. There might be a worry that clarifications on how cardinal comparisons are understood can only go so far. Some might retort that it is immaterial whether a Mozart-life is a *thousand*, or a *million*, or a *trillion* times better than a Muzak-life. What is supposed to be objectionable is the implication that there is *some* sufficiently large number of Muzak-lives that would be as good as a hundred Mozart-lives. Call this the "generalized Repugnant Conclusion."

This generalized Repugnant Conclusion is supposed to be an implication of Additivism no matter how cardinal claims are interpreted. And that is because if a Mozart-life has a value of x (however large) and a Muzak-life has a value of y (however small), there is bound to be some large enough positive integer N such that $N \times y$ is greater than x.

This is indeed the case if x and y are real numbers—it follows from what is called the "Archimedean property" of the real numbers. But recall that we allow for representations to take value in more general structures, partially ordered Abelian groups, which can lack this property. This leads us to yet another respect in which Additivism is more general than Total Utilitarianism standardly conceived—the possibility of non-real–valued representations.

5.3 | Partially Ordered Abelian Groups

Recall the lexical spaces introduced earlier, in which sequences of real numbers are compared by comparing the first components, moving on to the second only if there is a tie in the first, moving on to the third only if there is a tie in the first two components, and so on. These spaces lack the Archimedean property.

For instance, assign a Mozart-life a value of $(1, 0)$ and a Muzak-life $(0, 1)$. Then, a Mozart-life is "infinitely" better in the following sense. Any arbitrarily large finite number N of Muzak-lives would still only add up to a value of $(0, N)$. That is less than $(1, 0)$ on the lexical ordering, since 0 is less than 1 along the first components. Since an additive representation can assign values like the ones above, the generalized Repugnant Conclusion is not forced upon an Additivist.[29]

While lexical spaces relax the Archimedean property of the real numbers, the dominance spaces also introduced earlier relax its completeness. On the standard ordering of the real numbers, any two real numbers are comparable—either one must be at least as great as the other. But some values might be incommensurable.[30] Perhaps a Mozart-life and a Monet-life are so different that neither is better than the other but neither are they equally good—they are just incomparable.

Such incomparability can be accommodated using dominance spaces, on which one sequence of real numbers is at least as great as another just in case it is at least as great along *all* components. For instance, valuing a Mozart-life at $(1, 0)$ and a Monet-life at $(0, 1)$ would make them incomparable according to the dominance ordering, since neither is at least as good as the other along both components of the pairs. Intuitively, each component would represent the magnitude of value along some aspect, like musical vs. visual edification.

It is possible for an axiology to have *both* things that are infinitely better than others as well as things that are incomparable. Such axiologies might be represented using a combination of lexical and dominance spaces. For instance, instead of pairs, consider 2-by-2 matrices like:

$$x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

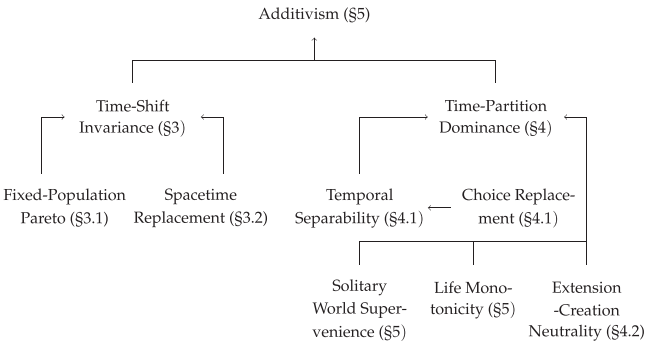
[27] See Crisp (1997, 24).
[28] For further discussion on related philosophical issues to do with the measurement of well-being and how to make sense of claims about the sum or magnitude of well-being, see Nebel (2023).
[29] See Nebel (2022).
[30] See Chang (2002), though see Dorr et al. (forthcoming).

These matrices can be added entry-wise. So, for instance, $x + y$ would be a 2-by-2 matrix with 1 everywhere except for a 0 on the bottom right. To compare matrices like the ones above, we first compare them row-wise using the lexical ordering. For instance, x is better than y along the first row (since $(1, 0)$ is better than $(0, 1)$ on the lexical ordering) but worse along the second row (since $(0, 0)$ is worse than $(0, 1)$). Then, we compare them using the dominance ordering, where one matrix is at least as good as the other if it is at least as good along every row. In the case above, x is at least as good as y only along the first row. So, x is not at least as good as y. And neither is y at least as good as x. So, they are incomparable.

These combinations of lexical and dominance spaces exhaust the partially ordered Abelian groups.[31] So, we can think of representations of axiologies as assigning values that are matrices like the one above (with possibly infinitely many rows and columns), summed up and compared in the way just described. This means that the cumulative effect of allowing representations to take values in arbitrary partially ordered Abelian groups, rather than just the real numbers, has a simple description. It allows for axiologies that are possibly lexical (on which some worlds are infinitely better than others) and possibly incomplete (on which some worlds might be incomparable).

6 | Concluding Reflections

Value can be distributed not just across people but also across time. These dimensions are not independent—how value is aggregated across time greatly constrains how it must be aggregated across people. In particular, two apparently compelling principles about time jointly entail Additivism. The first, Time-Shift Invariance, says that merely shifting lives in time should not change how good things are. The second, Time-Partition Dominance, says roughly that things cannot be better unless they are better within some period or another. Strong theoretical considerations can be provided in support of each principle. The web of relations among the various principles is summarized in the figure below.



But implicit in the formulation of these principles and in the arguments for them are certain assumptions about the underlying structure of value. The notion of a "mere" time-shift assumes that we can change when a life is lived while holding the value of its individual segments fixed. Similarly, the intelligibility of within-a-period comparisons assumes that we can hold fixed value within a given period while "deleting" all value outside of it. Shifting lives in time or deleting surrounding goods often results in the loss of certain goods. Nevertheless, value can be held fixed if such losses can always be made up for with other equally good sources of value. Replacement codifies the assumption that the underlying space of value is rich enough to always allow for such compensations.

More generally, some sources of value might be impossible to instantiate under certain conditions—like when people are too far apart in time or in space, when agents face particular choices, or when certain necessary connections hold among different sources of value. Replacement principles are principles to the effect that under such conditions, those sources can always be substituted with equivalent sources of value that *can* be instantiated in some given configuration or another. Such principles underlie various arguments for the two main principles—like Spacetime Replacement in the argument from relativity (Section 3.2) and Localized Choice Replacement in the argument from decision-making (Section 4.1). Some of the most salient objections to Extension-Creation Neutrality—and ipso facto, Time-Partition Dominance—based on holistic features of a life are also cause for rejecting Replacement (Section 4.2).

Recognizing the centrality of replacement principles in the case for Additivism has major methodological ramifications for how we conceptualize and engage in the enterprise of population ethics. First, two kinds of questions in axiology are often entertained separately. On the one hand are *substantive* questions about the nature and content of value—what things make for value and when is one such thing better than another? On the other are *structural* questions to do with aggregation—how does the pattern in which value (whatever that may be) is distributed determine overall value?

Much of population axiology takes itself to be concerned with structural questions, while attempting to remain as neutral as possible about substantive ones. We find, for instance, in seminal works of population ethics statements like the following:

> "welfare" is a term that will be used often in this essay. This concept has, not surprisingly, acquired a number of different meanings. On the one hand, we need to narrow down the possible meanings of this expression so that we know what the examples and principles that

[31] The theorem that underlies this fact is a generalization of the Hahn Embedding Theorem (see Hausner and Wendel (1952); Conrad (1953); Clifford (1954)). The Hahn Embedding Theorem shows that any linearly ordered Abelian group is isomorphic to an ordered subgroup of a lexicographically ordered additive group $\mathbb{R}^I$, which contains sequences of reals (indexed by a well-ordered set I) which vanish outside of a well-ordered subset of I. The theorem can be generalized to partially ordered Abelian groups whose ordering may not be complete. Specifically, for any well-ordered sets I and J, let an arbitrary element of $\left(\mathbb{R}^I\right)^J$ be a sequence of sequence of reals of the form $((x_i^j)_{i\in I})_{j\in J}$, where for each j, the sequence $(x_i^j)_{i\in I}$ vanishes outside of a well-ordered subset of I. The lexical-dominance ordering can be defined as follows. For each $j \in J$, we define a lexical ordering $\leq_j$ on which $(x_i^j)_{i\in I} \leq_j (y_i^j)_{i\in I}$ just in case either $x_{i_1}^j < y_{i_1}^j$ (where i_1 is the smallest member of I) or $x_{i_1}^j = y_{i_1}^j$ but $x_{i_2}^j < y_{i_2}^j$ (where i_2 is the second smallest member of I), and so on. Then, the group ordering $\leq$ is defined by letting $((x_i^j)_{i\in I})_{j\in J} \leq ((y_i^j)_{i\in I})_{j\in J}$ just in case $(x_i^j)_{i\in I} \leq_j (y_i^j)_{i\in I}$ for all $j \in J$. $\left(\mathbb{R}^I\right)^J$ with componentwise addition and the dominance-lexical ordering is a partially ordered Abelian group. And it can be shown that every partially ordered Abelian group can be embedded as an ordered subgroup of some such group $\left(\mathbb{R}^I\right)^J$.

> we shall discuss involve. On the other hand, we want to avoid taking a stand on controversial issues about welfare which don't affect the nature of the problems that we are going to discuss—we don't want to narrow the scope of our discussion unnecessarily. (Arrhenius 2000, 6)
>
> I do not assume any particular account of temporal wellbeing. The wellbeing of a person at a time is how well her life goes at that time. It takes into account everything that is good or bad for her at the time. There are narrow accounts of temporal wellbeing, such as hedonism, and broader accounts. I am happy to accommodate any of them. This book is about the aggregation of wellbeing, and I prefer to remain as uncommmitted as possible about the nature of the wellbeing that is aggregated. If you think you know some sort of good that does not appear in the distribution of temporal wellbeing, you may need to broaden your conception of temporal wellbeing. Once you have done that, the good may appear in the distribution after all. (Broome 2004b, 45)

But if replacement principles and the case for Additivism are indeed so intertwined, then the structural and substantive questions are not so easily separated. For, the plausibility of replacement principles turn on substantive questions about exactly what value consists in. For instance, are there substitutes for friendship under conditions that make friendship impossible, like when people are too far apart in time or in space? An affirmative answer is much more plausible on a theory like hedonism than on one that places overwhelming importance on our social relationships. It is no accident then that certain substantive theories of value (like hedonism) tend to incline ethicists toward certain structural views (like an additive picture of value)—a combination of views whose history traces back to Bentham (1977) and Mill (1867).

The inseparability of the substantive from the structural should not be surprising. Claims about aggregation often lack precise content until they are supplemented with some theory of value. For instance, the claim that *value* is additive is not a concrete view but a schema—"value" is a placeholder whose scope requires further specification. Additivism means something very different coming from a hedonist and from a non-hedonist.

Various other structural claims about aggregation similarly contain "value" as a free parameter to which some specific theory is to be plugged in. So, there may be no real disagreement—at least not about aggregation—between someone who rejects a structural principle because they accept one theory of value and another who rejects the principle because they accept a different theory of value. The exact content of a principle (like that *value* is additive or that holding *value* fixed, this or that difference in distributional pattern is irrelevant) varies with what is meant by "value."

Some issues might have obvious answers regardless of how the "value" parameter is set. But in most cases, "value" is simply too nebulous and flexible a concept for structural questions to have definitive answers without further clarification of what "value" encompasses. Much of this paper is an exercise in illustrating this point: If friendship is of *sui generis* value, then principles like Replacement and Spacetime Replacement fail, and the case for Time-Shift Invariance and Additivism thereby weakened…If "value" includes value associated with the length and shape of a life, then Replacement fails; if not, then perhaps Extension-Creation Neutrality fails…Short of suspending inquiry into structural questions entirely until a consensus is reached about what constitutes value, the best that can be hoped for are conditional claims like these about the plausibility of various structural claims relative to different ways of specifying the scope of "value."

It is worth drawing a parallel to decision theory, where the dependence of the substantive and the structural is somewhat better appreciated. On one view, the role of decision theory is simply to identify general structural constraints that rational preference ought to satisfy *whatever the contents of preferences may be*. The problem is that purely structural constraints are vacuous without some minimal specification of the objects of preferences. For instance, take the constraint of transitivity. Alleged counterexamples to transitivity (a rational agent strictly prefers x to y, y to z, but also z to x) can always be accommodated by reindividuating the objects (the agent's preference is not for x over y, y over z, and so on but for x-when-the-alternative-is-y over y-when-the-alternative-is-x, for y-when-the-alternative-is-z over z-when-the-alternative-is-y, and so on).[32] In the complete absence of ground rules about the objects of preferences, structural constraints like transitivity have no bite.[33] The overarching aim of this paper is to drive home a similar lesson in axiology: a theory of aggregation cannot be developed in total abstraction from some constraints on value's contents.

Another important methodological upshot of our discussion concerns a reformation of the standard framework of population ethics. The use of numbers obfuscates many important foundational issues and must thus be approached with care. Mathematical representations of value can be justified—for instance, via a representation theorem, like the one proved in Appendix A. Though, even then, subtle issues of interpretation remain, as we saw in *Section* 5.

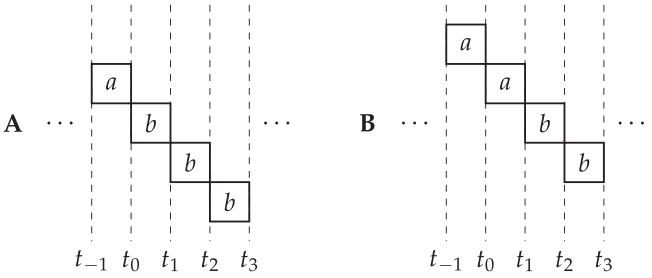
Those who reject replacement principles should be especially wary of the use of numbers that conflate different qualitative sources of value. A value of "10" could stand for many equally good things—an adrenaline rush, a feeling of peace, an intellectual achievement, a bonding moment. Some of these "10s" might be possible to instantiate at some time but not others, for some people but not others, at some spatial locations but not others, in conjunction with some other sources of value but not others. Those who subscribe to theories which result in the failure of replacement principles should carefully specify and keep track of what each instance of "10" in a distribution stands for, since there is no guarantee that there is always some source that makes for that amount of value. In that case, the

[32] See Sen (1993).
[33] This point is convincingly argued for by Broome (1999, Ch. 5).

standard numerical framework of population ethics is perhaps best abandoned entirely.

This paper focused on highlighting under-appreciated connections between substantive issues and structural issues. Nevertheless, it is worth emphasizing that this does not mean one's view about aggregation always just boils down to one's substantive view about what is valuable. Certain assumptions about what makes for value can stack the deck in favor of, say, an additive theory of value. But while assumptions like various replacement principles lay the groundwork for *one possible route* toward Additivism, accepting Additivism does not require accepting such principles. Nor does rejecting Additivism require rejecting replacement principles.

One of the most pressing challenges for Additivism that is, in some sense, purely structural and has to do simply with how addition behaves concerns its infinitary generalization. In a generalized framework containing distributions with infinitely many episodes, our two key axioms are inconsistent. Consider two distributions with lives extending infinitely into the past and into the future:



B differs from **A** only in that the episodes are uniformly shifted rightward by one unit of time. According to Time-Shift Invariance, the two worlds must be equally good. But supposing episode b to be better than a, **A** and **B** are equally good before t_0 and after t_1. And **B** is better between t_0 and t_1. So, **B** must be better than **A** according to Time-Partition Dominance—contradiction. Such issues are left for future work to address.[34]

Acknowledgments

For incredibly helpful comments and discussion, I am grateful to Odilon Camara, Pietro Cibinel, Zach Goodsell, John Hawthorne, Philip Li, Ralph Wedgwood, Matt Wiseman, and audiences at the Yale Early Career Ethics Workshop and USC's Speculative Society. Thanks to two referees for this journal whose incisive comments have greatly improved the paper. Special thanks are owed to Jake Nebel and especially Jeff Russell for going above and beyond as advisors—the time and effort they have dedicated toward this and related projects over the years cannot be overstated.

REFERENCES

Adler, M. D. 2007. "*Well-Being, Inequality and Time: The Time-Slice Problem and Its Policy Implications*." Technical Report. Institute for Law & Economics, University of Pennsylvania Law School.

Arrhenius, G. 2000. "Future Generations: A Challenge for Moral Theory." PhD diss., Uppsala University.

Arrhenius, G. 2011. "The Impossibility of a Satisfactory Population Ethics." In *Descriptive and Normative Approaches to Human Behavior*, edited by E. Dzhafarov, and L. Perry, 1–26. World Scientific.

Askell, A. 2018. "Pareto Principles in Infinite Ethics." PhD diss., New York University.

Bentham, J. 1977. *A Comment on the Commentaries and a Fragment on Government*. Humanities Press.

Blackorby, C., W. Bossert and D. J. Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge University Press.

Bostrom, N. 2011. "Infinite Ethics." *Analysis and Metaphysics* 10: 9–59.

Bramble, B. 2014. "Whole-Life Welfarism." *American Philosophical Quarterly* 51, no. 1: 63–74.

Bramble, B. 2017. *The Passing of Temporal Well-Being*. Routledge.

Brännmark, J. 2001. "Good Lives: Parts and Wholes." *American Philosophical Quarterly* 38, no. 2: 221–231.

Broome, J. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell.

Broome, J. 1999. *Ethics Out of Economics*. Cambridge University Press.

Broome, J. 2004a. "The Value of Living Longer." In *Public Health, Ethics, and Equity*, edited by S. Anand, F. Peter, and A. Sen, 243–260. Oxford University Press.

Broome, J. 2004b. *Weighing Lives*. Oxford University Press.

Chang, R. 2002. "The Possibility of Parity." *Ethics* 112, no. 4: 659–688.

Clifford, A. H. 1954. "Note on Hahn's Theorem on Ordered Abelian Groups." *Proceedings of the American Mathematical Society* 5, no. 6: 860–863.

Conrad, P. F. 1953. "Embedding Theorems for Abelian Groups With Valuations." *American Journal of Mathematics* 75, no. 1: 1–29.

Crisp, R. 1997. *Mill on Utilitarianism*. Routledge.

Debreu, G. 1960. "Topological Methods in Cardinal Utility Theory." In *Mathematical Methods in the Social Sciences*, 16–26. Proceedings of the First Stanford Symposium. Stanford University Press.

Dorr, C., J. M. Nebel, and J. Zuehl. Forthcoming. "The Case for Comparability." Noûs.

Dorsey, D. 2013. "Desire-Satisfaction and Welfare as Temporal." *Ethical Theory and Moral Practice* 16, no. 1: 151–171.

Fanciullo, J. 2024. "Why Prevent Human Extinction?" *Philosophy and Phenomenological Research* 109, no. 2: 650–662.

Frick, J. 2017. "On the Survival of Humanity." *Canadian Journal of Philosophy* 47, no. 2–3: 344–367.

Gorman, W. M. 1968. "The Structure of Utility Functions." *Review of Economic Studies* 35, no. 4: 367–390.

Hausner, M., and J. G. Wendel. 1952. "Ordered Vector Spaces." *Proceedings of the American Mathematical Society* 3, no. 6: 977–982.

Hurka, T. M. 1982a. "Average Utilitarianisms." *Analysis* 42, no. 2: 65–69.

Hurka, T. M. 1982b. "More Average Utilitarianisms." *Analysis* 42, no. 3: 115–119.

Karhu, T. N. 2019. "Matters of Life and Death." PhD diss., London School of Economics and Political Science.

King, O. C. 2018. "Pulling Apart Well-Being at a Time and the Goodness of a Life." *Ergo: An Open Access Journal of Philosophy* 5: 349–370.

[34] See, for instance, Vallentyne and Kagan (1997), Lauwers and Vallentyne (2004), Bostrom (2011), Askell (2018) and Wilkinson (2020).

Krantz, D., D. Luce, P. Suppes, and A. Tversky, eds. 1971. *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*. New York Academic Press.

Lauwers, L., and P. Vallentyne. 2004. "Infinite Utilitarianism: More Is Always Better." *Economics and Philosophy* 20, no. 2: 307–330.

Lee, A. Y. 2018. "Is Consciousness Intrinsically Valuable?" *Philosophical Studies* 175, no. 1: 1–17.

Lee, A. Y. 2023. "The Neutrality of Life." *Australasian Journal of Philosophy* 101, 685–703.

McKerlie, D. 1989. "Equality and Time." *Ethics* 99, no. 3: 475–491.

McKerlie, D. 2001. "Dimensions of Equality." *Utilitas* 13, no. 3: 263.

McKerlie, D. 2012. *Justice Between the Young and the Old*. Oxford University Press USA.

Mill, J. S. 1867. *Utilitarianism*. World Pub. Co.

Mogensen, A. L. 2022. "The Only Ethical Argument for Positive δ? Partiality and Pure Time Preference." *Philosophical Studies* 179, no. 9: 2731–2750.

Nagel, T. 1970. "Death." *Noûs* 4, no. 1: 73–80.

Nagel, T. 1979. *Mortal Questions*. Cambridge University Press.

Nagel, T. 1995. *Equality and Partiality*. Oxford University Press.

Nebel, J. M. 2022. "Totalism Without Repugnance." In *Ethics and Existence: The Legacy of Derek Parfit*, edited by J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan, 200–231. Oxford University Press.

Nebel, J. M. 2023. "The Sum of Well-Being." *Mind* 132, no. 528: 1074–1104.

Parfit, D. 1984. *Reasons and Persons*. Oxford University Press.

Pivato, M. 2014. "Additive Representation of Separable Preferences Over Infinite Products." *Theory and Decision* 77, no. 1: 31–83.

Rosati, C. S. 2021. "The Normative Significance of Temporal Well-Being." *Res Philosophica* 98, no. 1: 125–139.

Russell, J. S. 2023. "On Two Arguments for Fanaticism." *Noûs* 58, no. 3: 565–595.

Sarch, A. 2013. "Desire Satisfactionism and Time." *Utilitas* 25, no. 2: 221–245.

Scheffler, S. 2018. *Why Worry About Future Generations?* Oxford University Press.

Sen, A. 1993. "Internal Consistency of Choice." *Econometrica* 61: 495–521.

Singer, P. 1993. *Practical Ethics*. 2nd ed. Cambridge University Press.

Slote, M. 2017. "Goods and Lives." *Pacific Philosophical Quarterly* 63, no. 4: 311–326.

Temkin, L. S. 1993. *Inequality*. Oxford University Press.

Thomas, T. 2022. "Separability and Population Ethics." In *The Oxford Handbook of Population Ethics*, 271–295. Oxford University Press.

Vallentyne, P., and S. Kagan. 1997. "Infinite Value and Finitely Additive Value Theory." *Journal of Philosophy* 94, no. 1: 5–26.

Velleman, J. D. 1991. "Well-Being and Time." *Pacific Philosophical Quarterly* 72, no. 1: 48–77.

Wakker, P. P. 2013. *Additive Representations of Preferences: A New Foundation of Decision Analysis*. Vol. 4. Springer Science & Business Media.

Wilkinson, H. 2020. "Infinite Aggregation: Expanded Addition." *Philosophical Studies* 178, no. 6: 1917–1949.

Wilkinson, R., and K. Pickett. 2009. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. Bloomsbury Press.

Appendix A

The goal of this appendix is to show that given the background assumptions, Time-Shift Invariance and Time-Shift Dominance entail Additivism. We already saw, roughly, how given the two axioms, every world is equivalent to some world with a solitary distribution, in which at most one person ever lives. In a nutshell, the proof strategy will be to construct a partially ordered Abelian group whose members are equivalence classes of solitary distributions, whose addition operation is the concatenation of solitary distributions in time, and whose ordering is the ordering of solitary distributions according to the axiology. The axioms and assumptions of the framework guarantee that the algebraic structure and ordering structure, defined this way, have the desired properties.

A.1 | The Framework

Let $\mathbf{E}$ be a set of episodes, including an empty episode 0. And let $\mathbf{E}^* = \mathbf{E}/\{0\}$. Let $\mathbb{N}$ be an index set for possible individuals. A distribution is a set $\{\alpha_n : \mathbb{R} \to \mathbf{E}\}_{n\in\mathbb{N}}$ of functions such that the union of their images $\bigcup_{n\in\mathbb{n}} \alpha_n(\mathbb{R})$ is a finite subset $\{a_1, \ldots, a_k\}$ of $\mathbf{E}$ and for each such $a \neq 0 \in \{a_1, \ldots, a_k\}$, the preimage $(\alpha_n)^{-1}(a)$ is a non-empty interval of $\mathbb{R}$ for a particular $n = i$ and is empty for $n \neq i$. Intuitively, $a_1, \ldots, a_k$ are the episodes contained in the distribution and $(\alpha_n)^{-1}(a)$ is the interval of time for which episode a is had by individual n.

Let D be the set of all distributions. Formally, we can identify a world with the set of all distributions that describe it. So, the set $\mathcal{W}$ of all worlds is some subset of the powerset of D. And $\alpha \in \mathbf{A}$ means that world $\mathbf{A}$ can be described by distribution α.

An axiology $\succcurlyeq$ is a preorder on $\mathcal{W}$. Intuitively, $\mathbf{A} \succcurlyeq \mathbf{B}$ means that world $\mathbf{A}$ is at least as good as world $\mathbf{B}$. We say that $\mathbf{A}$ and $\mathbf{B}$ are equally good ($\mathbf{A} \sim \mathbf{B}$) when both $\mathbf{A} \succcurlyeq \mathbf{B}$ and $\mathbf{B} \succcurlyeq \mathbf{A}$. And we say that $\mathbf{A}$ is strictly better than $\mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) when $\mathbf{A} \succcurlyeq \mathbf{B}$ but not $\mathbf{B} \succcurlyeq \mathbf{A}$.

An axiology induces a preorder on the set of episodes. Intuitively, some worlds are singleton worlds, whose distributions are such that exactly one episode is ever instantiated. So, for each episode a, we can find a singleton world $\mathbf{A}$ whose distribution instantiates only a. We can then define episodes a and b to be equally good ($a \equiv b$) when the singleton world $\mathbf{A}$ associated with a and the singleton world $\mathbf{B}$ associated with b are equally good.[35] Two distributions α, β are then said to be equivalent (written "$\alpha \cong \beta$") if for each $n \in \mathbb{N}$ and $a \in A \subseteq \mathbf{E}$ such that $\bigcup_{a\in A}(\alpha_n)^{-1}(a) = \mathbb{R}$, there exists $b \in \mathbf{E}$ where $(\alpha_n)^{-1}(a) = (\beta_n)^{-1}(b)$ and $a \equiv b$. Informally, distributions are equivalent if they have the same pattern but with episodes possibly replaced by equally good ones.

The basic assumptions of our framework can then be stated as follows:

Replacement. For any $\alpha \in D$, there exist $\beta \in D$ and $\mathbf{B} \in \mathcal{W}$ such that $\alpha \cong \beta$ and $\beta \in \mathbf{B}$.

Distributionism. For any $\mathbf{A}, \mathbf{B} \in \mathcal{W}$ and $\alpha, \beta \in D$, if $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$ and $\alpha \cong \beta$, then $\mathbf{A} \sim \mathbf{B}$.

Anonymity. For any $\alpha \in D$ and bijection $\pi : \mathbb{N} \to \mathbb{N}$, if $\alpha = \{\alpha_n\} \in \mathbf{A}$ and $\alpha_\pi = \{\alpha_{\pi(n)}\} \in \mathbf{B}$, then $\mathbf{A} \sim \mathbf{B}$.

Given Replacement and Distributionism, an axiology $\succcurlyeq$ induces a ranking on distributions as follows (which we can also represent with "$\succcurlyeq$" without any risk of confusion). For any $\alpha, \beta \in D$, let $\alpha \succcurlyeq \beta$ if and only if there exist $\alpha' \cong \alpha$ and $\beta' \cong \beta$ with $\alpha' \in \mathbf{A}$ and $\beta' \in \mathbf{B}$ such that $\mathbf{A} \succcurlyeq \mathbf{B}$. The ordering of distributions, defined this way, can easily be checked to be a well-defined preorder given Replacement and Distributionism.

A.2 | Time-Shift Invariance

We now have the requisite vocabulary to state our first axiom precisely. Call a function $f : \mathbb{R} \to \mathbb{R}$ a translation function if it is of the form $f : t \mapsto t + x$ for some translation constant $x \in \mathbb{R}$. Say that distributions $\alpha, \beta \in D$ differ only by a time-shift if for each $n \in \mathbb{N}$, there exist a translation function $f_n : \mathbb{R} \to \mathbb{R}$ such that $\alpha_n = f_n \circ \beta_n$ (i.e., α_n is the result of applying β_n followed by f_n). Two worlds $\mathbf{A}, \mathbf{B}$ are then said to

[35] Of course, there are many possible choices of singleton worlds for each episode but given our axioms later, the choice will turn out not to matter.

differ only by a time-shift if there exist distributions α, α', β with $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$ such that α and α' differ only by a time-shift and $\alpha' \cong \beta$. Then, according to the first axiom:

Time-Shift Invariance. For any $\mathbf{A}, \mathbf{B} \in \mathcal{W}$, if $\mathbf{A}$ and $\mathbf{B}$ differ only by a time-shift, then $\mathbf{A} \sim \mathbf{B}$.

It follows that any two distributions that differ only by a time-shift must be equally good. For suppose $\alpha, \beta \in \mathcal{D}$ differ only by a time-shift. Given Replacement, there exist $\alpha', \beta' \in \mathcal{D}$ such that $\alpha' \in \mathbf{A}$ and $\beta' \in \mathbf{B}$ for some $\mathbf{A}, \mathbf{B} \in \mathcal{W}$. By definition, $\mathbf{A}$ and $\mathbf{B}$ differ only by a time-shift and so by Time-Shift Invariance, $\mathbf{A} \sim \mathbf{B}$. And so, by the definition of preorder on the distributions, it follows that $\alpha \sim \beta$.

A.3 | Time-Partition Dominance

For the second axiom, let us first define a binary concatenation operation $\oplus_t$ on functions $\alpha_n, \beta_n : \mathbb{R} \to \mathbf{E}$ as follows: For each $x \in \mathbb{R}$:

$$\alpha_n \oplus_t \beta_n(x) = \begin{cases} \alpha_n(x) & \text{if } x \leq t; \\ \beta_n(x) & \text{if } x > t. \end{cases}$$

We can then define the distribution that is a concatenation of distributions at t by letting $\alpha \oplus_t \beta = \{\alpha_n \oplus_x \beta_n\}_{n \in \mathbb{N}}$. Intuitively, this is the spliced distribution that looks like α before time t and β after t. Let $\varnothing$ be the empty distribution, where $\varnothing(n, t) = 0$ for all $n \in \mathbb{N}$ and $t \in \mathbb{R}$.

We call $t \in \mathbb{R}$ a partition for a distribution α if for each $n \in \mathbb{N}$, there exists $a \in \mathbf{E}/\{0\}$ such that either $\min\{(\alpha_n^{-1}(a)\} \geq t$ or $\max\{(\alpha_n)^{-1}(a)\} \leq t$. In other words, no one enjoys a nonempty episode whose interval spans a time that includes t. We say that world $\mathbf{A}$ is at least as good as $\mathbf{B}$ before time t ($\mathbf{A} \succcurlyeq_{<t} \mathbf{B}$) if there exist $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$ such that:

(i) $t \in \mathbb{R}$ is a common partition for α, β; and

(ii) for all $\alpha' \cong (\alpha \oplus_t \varnothing)$ and $\beta' \cong (\beta \oplus_t \varnothing)$, if $\alpha' \in \mathbf{A}'$ and $\beta' \in \mathbf{B}'$, then $\mathbf{A}' \succcurlyeq \mathbf{B}'$.

Similarly, $\mathbf{A}$ is at least as good as $\mathbf{B}$ after time t ($\mathbf{A} \succcurlyeq_{>t} \mathbf{B}$) if there exist $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{B}$ such that:

(i) $t \in \mathbb{R}$ is a common partition for α, β; and

(ii) for all $\alpha' \cong (\varnothing \oplus_t \alpha)$ and $\beta' \cong (\varnothing \oplus_t \beta)$, if $\alpha' \in \mathbf{A}'$ and $\beta' \in \mathbf{B}'$, then $\mathbf{A}' \succcurlyeq \mathbf{B}'$.

The second axiom then says that:

Time-Partition Dominance. For all $\mathbf{A}, \mathbf{B} \in \mathcal{W}$:
(i) if $\mathbf{A} \succcurlyeq_{>t} \mathbf{B}$ and $\mathbf{A} \succcurlyeq_{<t} \mathbf{B}$, then $\mathbf{A} \succcurlyeq \mathbf{B}$; and
(ii) if, furthermore, either $\mathbf{A} \succ_{>t} \mathbf{B}$ or $\mathbf{A} \succ_{<t} \mathbf{B}$, then $\mathbf{A} \succ \mathbf{B}$.

It follows that for any common partition $t \in \mathbb{R}$ of distributions $\alpha, \beta \in \mathcal{D}$:

(i) if $(\alpha \oplus_t \varnothing) \succcurlyeq (\beta \oplus_t \varnothing)$ and $(\varnothing \oplus_t \alpha) \succcurlyeq (\varnothing \oplus_t \beta)$, then $\alpha \succcurlyeq \beta$; and

(ii) if, furthermore, either $(\alpha \oplus_t \varnothing) \succ (\beta \oplus_x \varnothing)$ or $(\varnothing \oplus_x \alpha) \succ (\varnothing \oplus_x \beta)$, then $\alpha \succ \beta$.

A.4 | The Result

A valuation $V : \mathcal{W} \to \mathcal{G}$ is a function from the set of worlds to a partially ordered Abelian group $\mathcal{G} = \langle G, +, \geq \rangle$. A valuation $V : \mathcal{W} \to \mathcal{G}$ represents an axiology $\succcurlyeq$ if for all $\mathbf{A}, \mathbf{B} \in \mathcal{W}$:

$$\mathbf{A} \succcurlyeq \mathbf{B} \quad \iff \quad V(\mathbf{A}) \geq V(\mathbf{B}).$$

As already mentioned, we can conflate each episode and some singleton world whose distribution contains only that episode. Identifying each episode a with an arbitrary singleton world $\mathbf{A}$ in this way, we can abuse notation slightly and let $V(a) = V(\mathbf{A})$. An axiology $\langle \mathcal{W}, \succcurlyeq \rangle$ is then said to be additive if there exists a valuation $V : \mathcal{W} \to \mathcal{G}$ representing it such that for all $\mathbf{A} \in \mathcal{W}$, if $\alpha \in \mathbf{A}$ and the episodes contained in the distribution is $\bigcup_{n \in \mathbb{N}} \alpha_n(\mathbb{R}) = \{a_1, \ldots, a_k\}$, then:

$$V(\mathbf{A}) = V(a_1) + \ldots + V(a_k).$$

We want to show:

> **Theorem.** *Given Replacement, Distributionism, and Anonymity, an axiology satisfies Time-Shift Invariance and Time-Partition Dominance if and only if it is additive.*

The right-to-left direction is easy to check. For the converse direction, fix an axiology $(\mathcal{W}, \succcurlyeq)$ and suppose it satisfies Time-Shift Invariance and Time-Partition Dominance. Our proof strategy will be to, first, construct a partially ordered commutative monoid $\mathcal{M} = \langle M, +, \geq \rangle$ from equivalence classes of distributions containing at most one person and then showing that the constructed monoid has the properties required to be embedded into a partially ordered Abelian group.

Step 1. The monoid set M

Let $\mathcal{D}_0 \subset \mathcal{D}$ be the subset of solitary distributions in which no one besides possibly person 0 enjoys any nonempty episodes. That is, $\alpha \in \mathcal{D}_0$ just in case for all $n \neq 0$, $\alpha_n(t) = 0$ for all $t \in \mathbb{R}$. For any $\alpha \in \mathcal{D}_0$, let $[\alpha] = \{\beta \in \mathcal{D}_0 : \alpha \sim \beta\}$ be the equivalence class of distributions in $\mathcal{D}_0$ that are exactly as good as α. We define the underlying set of the monoid to be $M = \{[\alpha] : \alpha \in \mathcal{D}_0\}$.

Step 2. The monoid operation $+$

Let $\alpha, \beta \in \mathcal{D}_0$. First, we define $\overleftarrow{\alpha}$ to be the time-shifted distribution where the final nonempty episode in α ends at time 0, that is, $\overleftarrow{\alpha} = f_\alpha \circ \alpha$, where $f_\alpha : r \mapsto r - \max\{\alpha_0^{-1}[E^*]\}$. Similarly, $\vec{\beta}$ is the time-shifted distribution where the first nonempty episode in β starts at 0, that is, $\vec{\beta} = g_\beta \circ \beta$, where $g_\beta : r \mapsto r - \min\{\beta_0^{-1}[E^*]\}$. Now, we define the concatenation operation $\odot$ by letting $\alpha \odot \beta = \overleftarrow{\alpha} \oplus_0 \vec{\beta}$. Informally, this is the distribution in which person 0 lives the life they would have in α followed by the life they would have in β. Now, we define the monoid operation $+$, where for all $[\alpha], [\beta] \in M$:

$$[\alpha] + [\beta] = [\alpha \odot \beta].$$

First, we need to check that this is well-defined in that it does not depend on the choice of representatives. That is, suppose $\alpha' \in [\alpha]$ and $\beta' \in [\beta]$. We need to show that $[\alpha \odot \beta] = [\alpha' \odot \beta']$. Now, α and $\alpha \odot \varnothing$ differ only by a time-shift and so, by Time-Shift Invariance, $\alpha \sim (\alpha \odot \varnothing)$. Similarly, $\alpha' \sim (\alpha' \odot \varnothing)$. By assumption, $\alpha \sim \alpha'$. So, $(\alpha \odot \varnothing) \sim (\alpha' \odot \varnothing)$. By the same reasoning, $(\varnothing \odot \beta) \sim (\varnothing \odot \beta')$. So, by Time-Partition Dominance, $(\alpha \odot \beta) \sim (\alpha' \odot \beta')$, which means that $[\alpha \odot \beta] = [\alpha' \odot \beta']$.

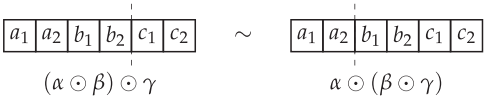
Next, we need to show that the operation, as defined above, satisfies the properties required for it to be the operation of a commutative monoid:

(i) *Identity*. There is an identity element $[\varnothing] \in M$ such that $[\alpha] + [\varnothing] = [\alpha]$ for all $[\alpha] \in M$. This follows from Time-Shift Invariance, since $(\alpha \odot \varnothing) = \overleftarrow{\alpha}$ differs from α only by a time-shift, and so $[\alpha] + [\varnothing] = [\alpha \odot \varnothing] = [\alpha]$.

(ii) *Associativity*. For all $[\alpha], [\beta], [\gamma] \in M$,

$$([\alpha] + [\beta]) + [\gamma] = [\alpha] + ([\beta] + [\gamma]).$$

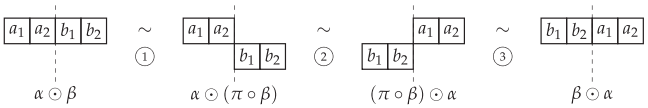
Again, this follows from Time-Shift Invariance. It is easy to see that $(\alpha \odot \beta) \odot \gamma$ and $\alpha \odot (\beta \odot \gamma)$ differ only by a time-shift, and so $([\alpha] + [\beta]) + [\gamma] = [(\alpha \odot \beta) \odot \gamma] = [\alpha \odot (\beta \odot \gamma)] = [\alpha] + ([\beta] + [\gamma])$.



(iii) *Commutativity*. For all $[\alpha], [\beta] \in M$,

$$[\alpha] + [\beta] = [\beta] + [\alpha].$$

This amounts to showing that $(\alpha \odot \beta) \sim (\beta \odot \alpha)$. Let $\pi : \mathbb{N} \to \mathbb{N}$ be a bijection such that $\pi(0) \neq 0$. Given Anonymity, $\pi \circ \beta \sim \beta$. So, by Time-Partition Dominance, $(\alpha \odot \beta) \sim (\alpha \odot (\pi \circ \beta))$ and $(\beta \odot \alpha) \sim ((\pi \circ \beta) \odot \alpha)$. By Time-Shift Invariance, $(\alpha \odot (\pi \circ \beta)) \sim ((\pi \circ \beta) \odot \alpha)$. So, $(\alpha \odot \beta) \sim (\beta \odot \alpha)$.



Step 3. The monoid preorder $\geq$

Define $\geq$ on M as follows:

$$[\alpha] \geq [\beta] \quad \text{iff} \quad \alpha \succcurlyeq \beta, \qquad \text{for all } [\alpha], [\beta] \in M.$$

This is well-defined in that it does not depend on the choice of representatives. For, suppose $\alpha' \in [\alpha]$ and $\beta' \in [\beta]$. Then, $\alpha' \sim \alpha$ and $\beta' \sim \beta$. Since $\succcurlyeq$ is a preorder, this means that $\alpha \succcurlyeq \beta$ if and only if $\alpha' \succcurlyeq \beta'$. It is also clear that $\geq$, as defined, is a preorder, since $\succcurlyeq$ is.

Finally, we need to check that the monoid preorder is compatible with the monoid operation previously defined in the sense that for all $[\alpha], [\beta], [\gamma] \in M$:

$$[\alpha] \geq [\beta] \quad \text{iff} \quad [\alpha] + [\gamma] \geq [\beta] + [\gamma].$$

To see this, first note that α and $\alpha \odot \varnothing$ differ only by a time-shift and so, by Time-Shift Invariance, $\alpha \succcurlyeq (\alpha \odot \varnothing)$. Similarly, $\beta \succcurlyeq (\beta \odot \varnothing)$. So, $\alpha \succcurlyeq \beta$ if and only if $(\alpha \odot \varnothing) \succcurlyeq (\beta \odot \varnothing)$. Furthermore, by reflexivity, $(\varnothing \odot \gamma) \sim (\varnothing \odot \gamma)$. So, by Time-Partition Dominance, $(\alpha \odot \varnothing) \succcurlyeq (\beta \odot \varnothing)$ if and only if $(\alpha \odot \gamma) \succcurlyeq (\beta \odot \gamma)$. Putting all these together:

$$\begin{aligned}
[\alpha] \geq [\beta] &\iff \alpha \succcurlyeq \beta \\
&\iff (\alpha \odot \varnothing) \succcurlyeq (\beta \odot \varnothing) \\
&\iff (\alpha \odot \gamma) \succcurlyeq (\beta \odot \gamma) \\
&\iff [\alpha \odot \gamma] \geq [\beta \odot \gamma] \\
&\iff [\alpha] + [\gamma] \geq [\beta] + [\gamma].
\end{aligned}$$

This completes our construction of the partially ordered commutative monoid.

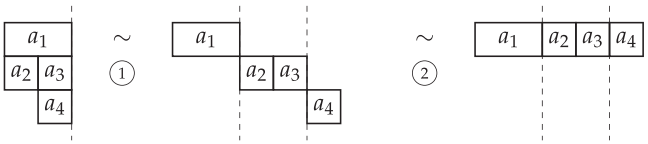
Now, our partially ordered commutative monoid can embedded into a partially ordered Abelian group if it is cancellative, that is, for all $[\alpha], [\beta], [\gamma] \in M$:

$$[\alpha] + [\gamma] = [\beta] + [\gamma] \quad \Rightarrow \quad [\alpha] = [\beta]$$

Suppose $[\alpha] + [\gamma] = [\beta] + [\gamma]$. Then, $[\alpha \odot \gamma] = [\beta \odot \gamma]$ and so, $(\alpha \odot \gamma) \sim (\beta \odot \gamma)$. By Time-Partition Dominance, $\alpha \sim \beta$, which implies that $[\alpha] \sim [\beta]$, as desired.

Step 4. The valuation function V

First, note that every distribution is exactly as good as some distribution in $\mathcal{D}_0$. For instance:



where ① follows by Time-Shift Invariance and ② by Time-Partition Dominance and Anonymity. So, we can define $V : \mathcal{W} \to \mathcal{G}$, where $\mathcal{G}$ is a partially ordered Abelian group in which the partially ordered commutative monoid $\mathcal{M}$ constructed above embeds, as follows:

$$V(\mathbf{A}) = [\alpha'], \quad \text{where } \alpha \in \mathbf{A}, \alpha \sim \alpha', \text{ and } \alpha' \in \mathcal{D}_0.$$

This is an additive representation for the axiology, since if $\mathbf{A}$ contains episodes $a_1, \ldots, a_n$:

$$V(\mathbf{A}) = [\alpha] = [\alpha_1 \odot \ldots \odot \alpha_n] = [\alpha_1] + \ldots + [\alpha_n] = V(a_1) + \ldots + V(a_n),$$

where each singleton distribution α_i contains only episode a_i.