For GWU Philosophy Department workshop, April 2010

Dear workshop participants – Attached are the draft Introduction and Chapter 5 of my forthcoming book, *Well-Being and Equity: A Framework for Policy Analysis* (Oxford U. Press 2011). The Introduction provides an overview of the project, which aims to integrate welfare economics and various philosophical literatures (concerning well-being, equality, and personal identity) to defend the use of "prioritarian" social welfare functions as tools for morally evaluating governmental policies and other large-scale choices. I plan to discuss Chapter 5 at the workshop. This is a *lot* to read, and readers pressed for time might want to focus on pp. 10-16, 28-50, and 59-65. But I welcome your comments and criticisms about any part of the chapter or the book project more generally.

As you'll see, the draft has no citations. The absence of a citation should not be taken as a claim of originality on my part! I've benefitted and relied upon much prior scholarship in writing this book; preparing citations is one of the (many) tasks I need to complete before publication.

I look forward to our discussion.

Best – Matt Adler

Introduction

This book aims to provide a comprehensive, philosophically grounded, defense of the use of social welfare functions as a framework for evaluating governmental policies and other large-scale choices.

The "social welfare function" (SWF) is a concept that originates in theoretical welfare economics. It is employed as a policy-analysis methodology in a number of economic literatures, such as "optimal tax" scholarship, growth theory, and environmental economics. But other methodologies –in particular, cost-benefit analysis (CBA) – are currently dominant. While CBA is defensible as a rough proxy for overall well-being,[1] it is insensitive to the distribution of well-being. By contrast, the SWF approach can incorporate distributive considerations into policy analysis in a systematic fashion.

Although I see SWFs as a practical policy-evaluation tool, the tenor of this book is theoretical. Just as the now-massive body of CBA scholarship is grounded in a theoretical literature regarding CBA, so, too, the proper design of the SWF framework raises many questions of normative theory – questions that this book will engage. In doing so, I draw upon welfare economics, social choice theory, and related formal literatures (such as utility theory and decision theory), and upon philosophical scholarship concerning a variety of topics, in particular well-being, equality, and personal identity.

Chapter 1 sets the stage. I see the SWF framework as a *moral* choice-evaluation framework. "Moral" reasoning is the species of normative reasoning characterized by a concern for human interests; by impartiality between different persons; and by a willingness to transcend and criticize existing social norms. SWFs provide a systematic tool for *morally* evaluating governmental policies and other large-scale choices. Chapter 1 explores the difference between moral evaluation and other kinds of normative evaluation, and briefly reviews questions of metaethics and normative epistemology that no work of normative theory can ignore. It also sets forth the basic argumentative strategy of this book: to take as given that a moral choice-evaluation framework should be *person-centered, consequentialist* and *welfarist* (for short, "welfarist") and to argue that the SWF approach is the most attractive framework of this sort.

In other words, this book works *within* welfarism, rather than engaging ongoing debates between welfarists and non-welfarists. Chapter 1 explains why this is a plausible strategy. However, it also takes some pains to explain why non-welfarists, too, should find the book of interest.

Chapter 1 concludes by offering a formal, generic, architecture for welfarism. The generic welfarist architecture derives a ranking of choices from a ranking of outcomes. The ranking of outcomes, in turn, depends upon individual well-being. The connection between the

[1] See Matthew D. Adler and Eric Posner, *New Foundations of Cost-Benefit Analysis* (2006).

ranking of outcomes and individual well-being is formalized via the concept of a "life-history": a pairing of a person and an outcome. Life-history (x; i) means being individual i in outcome x. A welfarist choice-evaluation framework includes an account of well-being, which at a minimum makes *intrapersonal* comparisons, ranking life-histories belonging to the same person. The Pareto principles constrain the ranking of outcomes – requiring it to be consistent with the intrapersonal ranking of life-histories in certain, basic, ways.

The SWF approach is one *specification* of this generic welfarist architecture; CBA is a competing specification.

Chapter 2 introduces the SWF framework. This approach has the distinctive feature of making *interpersonal* comparisons between life-histories – not just intrapersonal comparisons. Further, it employs a utility function (or set of such functions) to map each outcome onto a "vector" or list of numbers, representing the well-being of each individual in the population in that outcome. Outcome x is mapped by utility function $u(.)$ onto $(u_1(x), u_2(x), \ldots, u_N(x))$, where $u_i(x)$ is a numerical measure of the well-being of individual i in outcome x. A SWF, in turn, is a mathematical rule for ranking outcomes as a function of their corresponding utility vectors. One simple possibility is to add up utilities: this is the utilitarian SWF. Another possibility is to employ an outcome-ranking rule which is sensitive to the distribution of utilities. There turn out to be a multiplicity of such distribution-sensitive SWFs.

Chapter 2 explains these ideas, and also reviews the intellectual history of the SWF approach (which originates in work by Abram Bergson and Paul Samuelson some 70 years ago, and, as mentioned, is well-accepted within certain subfields of economics). The bulk of the chapter, however, focuses on criticizing the competing policy-analytic frameworks that are currently dominant. These competitors include not only CBA, but also inequality metrics, such as the well-known Gini coefficient; various other types of metrics for quantifying inequity, such as poverty metrics, "social gradient" metrics, and tax incidence metrics; and cost-effectiveness analysis (CEA). Each of these approaches is widely employed in academic work, and CBA also now has a firm legal status in governmental practice. However, each of these approaches is problematic – at least from the perspective of welfarism.[2] As Chapter 2 will show, these approaches may be vulnerable to violations of the Pareto principles, or may fail to rank outcomes in a well-behaved manner (for example, by ranking outcome x over y but y over x, or x over y and y over z but not x over z). And even if non-SWF methodologies *are* structured so as to yield a well-behaved, Pareto-respecting ranking of outcomes, they turn out to be problematic in other ways.

The analysis in Chapter 2 is meant to *motivate* the defense and elaboration of the SWF approach which occurs in subsequent chapters. Chapters 3, 4, and 5 address the central theoretical questions that must be confronted by any proponent of this approach. Chapter 3

[2] Alternatively, certain ways of employing currently dominant frameworks turn out to be variations on the SWF approach. This is true, in particular, of the use of CBA with so-called "distributive weights." See Chapter 2.

focuses on well-being. One philosophically contested issue concerns the choice between preferentialist, hedonic, and objective-good accounts of human welfare. Insofar as utility numbers are meant to quantify individual well-being in outcomes, what exactly should these numbers be measuring? A cross-cutting issue concerns interpersonal comparability. How are we to make sense of the statement that life-history (*x*; *i*) is better for well-being than life-history (*y*; *j*): that individual *i* in outcome *x* is better off than individual *j* in outcome *y*? Why believe that this statement is meaningful? What are the criteria for ranking life-histories involving different persons? Economists outside the SWF tradition are usually skeptical about the possibility of interpersonal comparisons. Many SWFs also make interpersonal comparisons of well-being *differences*, saying that the difference in well-being between life-history (*x*; *i*) and (*y*; *j*) is greater than the difference in well-being between life-history (*z*; *k*) and (*w*; *l*). But what are the criteria that would enable us to make sense of *these* sorts of comparisons?

Chapter 3 tackles these problems, proposing to analyze well-being in terms of fully-informed, fully rational, convergent extended preferences. While an ordinary preference is simply a ranking of outcomes and choices, an *extended preference* is a ranking of life-histories. To say that individual *k* has an extended preference for (*x*; *i*) over (*y*; *j*) means that *k* prefers the life-history of *i* in *x* to the life-history of *j* in *y*. The idea of an extended preference originates with John Harsanyi. More specifically, Harsanyi proposes that an interpersonally comparable metric of individual well-being be constructed by appealing to individuals' extended preferences over life-history *lotteries* – on the premise that these extended lottery preferences comply with expected utility theory. Chapter 3 will develop Harsanyi's fruitful ideas. To be sure, many challenges arise in doing so; and the account of well-being presented in Chapter 3, in a number of important respects, diverges from Harsanyi's views. In particular, my definition of extended preferences builds in a self-interest component, designed to screen out preferences for features of outcomes that have no impact on well-being; and I allow for heterogeneity in extended preferences.

The thrust of Chapter 3 is to defend the following approach for making intra- and interpersonal comparisons, and for measuring well-being via utility numbers. There is a set **U** of utility functions, pooling the fully informed, fully rational, extended preferences of everyone in the population. Life history (*x*; *i*) is at least as good for well-being as life-history (*y*; *j*) just in case $u(x; i) \geq u(y; j)$ for all $u(.)$ in **U**. A similar rule is proposed for well-being differences.[3]

Chapter 4 turns to the question of specifying the SWF. An SWF is some rule for *using* the well-being information captured in the set **U** of utility functions in order to rank outcomes. Chapter 4 argues that the most attractive such rule is a *prioritarian* SWF (more precisely, a "*continuous* prioritarian" SWF). In defending this view, Chapter 4 draws heavily on the contemporary philosophical literature concerning equality. One major theme in this literature is

[3] The well-being difference between life-history (*x*; *i*) and (*y*; *j*) is at least as great as the well-being difference between life-history (*z*; *k*) and life-history (*w*; *l*) iff, for all $u(.)$ in **U**, $u(x; i) – u(y; j) \geq u(z; k)-u(w; l)$.

the debate between those who hold a "prioritarian" conception of fair distribution, and those who reject this view. "Prioritarians" argue that well-being changes affecting worse-off individuals have greater moral significance. In other words, well-being has declining marginal moral weight. It is this proposition, and not the intrinsic value of equality, that provides the best justification for a non-utilitarian moral view – or so "prioritarians" claim. Prioritarianism corresponds to an SWF which satisfies two key axioms, explained in Chapter 4: the "Pigou-Dalton" axiom, and an axiom of separability across persons. If we add a continuity requirement, the upshot is a SWF which sums up individual utilities that have been "transformed" by a transformation function, rather than simply summing utilities in utilitarian fashion.

Formally, a continuous prioritarian SWF says: outcome *x* is morally at least as good as outcome *y* iff, for all *u*(.) belonging to **U**, $\sum_{i=1}^{N} g(u_i(x)) \geq \sum_{i=1}^{N} g(u_i(y))$, where the *g*(.) function is strictly increasing and concave (which is what ensures that this SWF both satisfies the Pareto principles *and* gives greater moral weight to well-being changes affecting worse-off individuals). Chapter 4 argues that this SWF represents the most attractive specification of welfarism.[4] A central claim in Chapter 4, and indeed throughout the book, is that welfarism and a concern for *fairness* are fully compatible. A moral view is sensitive to fairness insofar as it "respects the separateness of persons" – insofar as it sees each person as having a separate moral claim to have her interests and concerns respected. Integrating a concern for fairness into welfarism means, first, that fairness structures the ranking of *outcomes*; and, second, that the "currency" for each individual's moral claim is her well-being. These ideas, in turn, lead most directly to the prioritarian SWF. Whether the prioritarian SWF should, in addition, satisfy the continuity requirement – and I believe it should – implicates questions regarding tradeoffs that are also reviewed in Chapter 4.

Chapter 5 addresses the temporal dimension. The approach defended in Chapters 3 and 4 makes the ranking of outcomes depend upon utility numbers representing individuals' lifetime well-being. A whole-lifetime view is, indeed, adopted by the theoretical literature on SWFs; by most extant scholarship that uses SWFs to evaluate governmental policies; and by the philosophical literature on equality, which generally argues that moral norms concerning fair distribution are properly focused on the distribution of lifetime well-being. But is the whole-lifetime approach really defensible? Why not represent an outcome as a list of "sublifetime" utilities, each representing the well-being of some individual during some portion of her life (for example, her annual or momentary well-being), and then apply a continuous prioritarian SWF to these sublifetime utilities? Chapter 5 will describe and seek to respond to two arguments that challenge whole-lifetime prioritarianism, and that seem to cut in favor of "sublifetime"

[4] More precisely, Chapter 4 argues for the "Atkinsonian" SWF, which is a particular type of continuous prioritarian SWF – and one that, in fact, is fairly widely used within existing SWF scholarship.

prioritarianism or some other approach.[5] One argument, tendered by Derek Parfit, suggests that a proper understanding of personal identity undercuts a concern for the distribution of lifetime well-being. A different argument, advanced by Dennis McKerlie and other philosophers, suggests that our intuitions about equality – in particular, intuitions about the moral significance of short-term hardship and suffering – are inconsistent with a whole-lifetime view.

Chapter 6 turns to the problem of implementation. While Chapter 3 undertook the philosophical labor required to defend a particular theory of well-being and well-being measurement, the question remains: how shall we actually estimate the utility functions which the SWF approach requires as its inputs? How shall we actually construct a set **U**? Chapter 6 addresses this question at length. It begins by addressing a question left open in Chapter 3. The outcomes which are ranked by a choice-evaluation framework (be it the SWF framework or a competing framework, such as CBA) are *simplified* descriptions of reality. Simplification is necessary for the framework to be cognitively tractable. (If an outcome were a fully precise specification of a possible reality, i.e., a complete "possible world," a *human* decisionmaker would be unable to use the framework.) But what does it mean for individuals to have extended preferences regarding life-histories involving simplified outcomes – outcomes that are missing some characteristics? For example, much SWF scholarship in the "optimal tax" tradition employs outcomes that describe each individual's consumption and leisure, but fail to describe other individual attributes (health, happiness, social life, etc.). How should individual *k* think about her preference regarding (*x*; *i*) and (*y*; *j*), where she is told only that individual *i* consumes a certain amount and has a certain amount of leisure time in outcome *x*, and that individual *j* consumes a certain amount and has a certain amount of leisure time in outcome *y*?

Chapter 6 proposes an answer to this vital question, regarding the valuation of simplified outcomes.[6] With that answer in hand, it discusses how we can use information about an individual's ordinary preferences in order to make inferences about her extended preferences. And it reviews, in detail, the wealth of existing data concerning individuals' ordinary preferences that enable a policy analyst to construct a set **U**: data regarding individuals' preferences for consumption lotteries; evidence concerning intertemporal substitution and the value of statistical life; "ordinal" preference data supplied by economic research concerning labor supply and consumer demand; so-called "QALY" surveys, which reveal how individuals rank health states and lotteries over health states; and happiness surveys. This chapter also proposes novel survey formats.

[5] A third approach would be "attribute based," whereby an SWF is applied directly to individual attributes, rather than to lifetime or sublifetime utilities representing individuals' lifetime or sublifetime well-being.

[6] The answer, in short, is that the enterprise of eliciting individuals' preferences with respect to simplified life-histories rests upon an *invariance* premise: that such preferences are more or less invariant to the particular level of the missing characteristics. If the invariance premise is untrue, a fuller description of outcomes is warranted – although considerations of cognitive tractability will weigh against describing outcomes with great specificity.

Chapters 3 through 6 all focus on the ranking of outcomes. Is the well-being of a given individual in a given outcome determined by her preference-satisfaction, her mental states, or her realization of objective goods? How should her well-being be measured by utility functions? What sort of data enables us to estimate these functions? What is the appropriate SWF for ranking outcomes in light of individual utilities? Is it a utilitarian SWF, a prioritarian SWF, or some other form?

Chapter 7 turns from these questions, to the problem of generating a ranking of choices from the ranking of outcomes. A choice-evaluation framework should function to provide guidance to a decisionmaker. In particular, the SWF framework – as I conceptualize it – is a systematic methodology that should yield guidance to governmental policymakers or others confronted with large-scale choices.[7] But a human decisionmaker operates under conditions of uncertainty. She is not sure which particular outcome would result from any given choice which is available to her. How to implement a continuous prioritarian SWF under conditions of uncertainty raises thorny problems. It turns out that no methodology for doing so can simultaneously respect, on the one hand, certain axioms which seem to capture the essence of consequentialism; and, on the other hand, the ex ante versions of the Pareto and Pigou-Dalton principles.

Expected utility theory, if refined along certain lines, provides an attractive generic structure for choice under uncertainty.[8] Chapter 7 argues that a continuous prioritarian SWF should be merged with a (refined version of) expected utility theory so as to generate a ranking of choices – notwithstanding violations of the ex ante Pareto and Pigou-Dalton principles. While the SWF framework defended here satisfies the Pareto and Pigou-Dalton principles in terms of the ranking of outcomes, the ex ante versions of these principles constitute an *additional* requirement which, on balance, should be rejected. The dilemmas that arise in specifying norms of fair distribution under conditions of uncertainty have been discussed by philosophers and social choice theorists; Chapter 7 builds upon this scholarship.

Chapter 8 reviews three important problems that are connected to those addressed in this book. It describes the problems, and in a very limited way outlines tentative responses, but does not attempt anything like a full treatment.

One problem concerns future generations. My analysis throughout the book assumes a fixed and finite population. The same *N* individuals exist in each of the possible outcomes of the policy choice at hand. Scholarship on future generations relaxes this assumption, by

[7] Policy-evaluation frameworks such as the SWF approach, CBA, or the other frameworks reviewed in Chapter 2 are appropriate for governmental policies or other large-scale choices, but not for smaller choices where the expected benefits of using a systematic framework are too small to justify the decision costs of doing so. Identifying the boundary between small and large choices is very difficult. See Chapter 1.

[8] To be clear, expected utility theory surfaces at two different junctures in this book: in Chapter 3, regarding the measurement of well-being; and in Chapter 7, regarding the moral ranking of choices under conditions of uncertainty.

allowing for the possibility that choices might affect the size or identity of the population, or for an infinite future and thus infinite population. How to structure policy choice under such conditions raises new and difficult questions: the so-called "repugnant conclusion"; "non-identity" problems; and the incompatibility between the Pareto principles and an axiom of impartiality in the case of an infinite future.

A second problem concerns the optimal design of legal institutions. To say that the SWF approach is an attractive framework for morally evaluating governmental policies and other large-scale choices is not, necessarily, to say that it is optimal to structure legal institutions so that policymakers are legally instructed to employ this framework. Policy-analysis tools may be distorted by political forces. (In particular, research examining the effects of laws requiring regulatory agencies to employ CBA has reached mixed conclusions concerning whether such laws have actually produced more efficient regulations.) A cross-cutting idea is that it may be optimal to "channel" distribution through the tax system, and thus to instruct non-tax bodies to evaluate their decisions using CBA rather than using some SWF which is sensitive to distributive considerations.

A third problem concerns individual responsibility. A key deficit of welfarism is that it fails to differentiate between bad luck and irresponsibility -- between a case in which someone is badly off through no fault of her own, and a case in which someone is (wholly or partly) responsible for her well-being shortfall. Over the last several decades, the philosophical literature on equality has intensively investigated problems of responsibility; and a growing body of work in welfare economics and social choice theory is now also engaging such problems. Chapter 8 briefly reviews these literatures, and in a preliminary way suggests how a concern for responsibility might be fused with the SWF framework.

This book is, evidently, interdisciplinary. It is aimed at welfare economists who are receptive to philosophical argumentation; at philosophers who are receptive to the mathematical tools of welfare economics; and to law and policy scholars who find value in both fields. It builds upon, and draws inspiration from, the tradition of scholarly work at the intersection of philosophy and economics, exemplified by journals such as *Economics and Philosophy* or *Social Choice and Welfare*. The methodology of welfare economics is axiomatic and deductive. The focus is on clarifying the logical implications of various axioms for ranking outcomes and choices which we might be inclined to endorse. The methodology of moral philosophy is coherentist. Given a plurality of logically possible approaches to ranking outcomes and choices, which ones are most attractive in the "reflective equilibrium" sense? Which approaches fit best with our intuitive judgments about concrete cases and with general normative principles, regarding well-being, equality, and so forth?

It would be arrogant and wrongheaded to suggest that normative understanding can only be advanced by interdisciplinary work. Clearly, that is not true; there are large epistemic gains to be had from specialization. However, it seems to this author equally wrongheaded to insist that specialization is the only viable path. This book is animated by the belief that scholars can make real progress in specifying normative tools and frameworks by marrying the methodologies of economics and philosophy. I'll leave it to the reader to judge whether they are, in fact, fruitfully married here.

Chapter 5: Lifetime Prioritarianism

My presentation of the SWF approach, and analysis of competing SWFs, has presupposed that SWFs are properly applied on a *lifetime* basis. I have used the term "life-history" to refer to an item such as (x; i) – a pairing of an individual and an outcome. An outcome is a simplified possible world, i.e., a simplified description of a whole possible history of the universe; and a life-history means being some individual in some outcome. Thus a "life-history" is a simplified description of the entire life of some individual. The SWF approach – as I have presented it – assumes that the well-being ranking of life-histories can be represented by a set **U** of utility numbers. These are *lifetime* utility numbers, tracking the well-being associated with whole life-histories. Each utility function in **U** maps an outcome onto a vector of individual lifetime utility numbers. And an SWF (I have assumed) is a rule for ranking pairs of outcomes as a function of their associated lifetime utility vectors.

But why should a SWF necessarily function in this fashion? Consider any given SWF: the utilitarian SWF, the rank-weighted SWF, the leximin SWF, the continuous prioritarian SWF, or any other. The SWF might, in principle, be applied on a *non-lifetime* basis. For example, it might be applied on a *sublifetime* basis. Imagine that there is a set **V** of sublifetime utility functions, measuring the sublifetime well-being realized by individuals during temporal portions of outcomes. The SWF might rank pairs of outcomes as a function of the vectors of sublifetime utility numbers associated with the outcomes by the elements in **V**. Alternatively, the SWF might be applied on an *attribute* basis – taking as its inputs numbers measuring the levels of various individual attributes in outcomes.

This Chapter defends the lifetime approach. My defense rests upon two key premises. The first is a premise about personal identity. *Personal identity continues through a normal human lifetime*. In other words, a normal human being (a human being who possesses the psychological attributes sufficient to make her a person, and who doesn't undergo a brain transplant, suffer profound amnesia as a result of an injury, or otherwise experience a radical rupture in the intertemporal connectedness of her psychological states and physical body) remains one and the same person from birth until death. The second is that the *moral ranking of outcomes is determined by accommodating individuals' claims across outcome*. The claim-across-outcome conception of the moral ranking of outcomes was at the heart of my analysis in Chapter 4. I argued that this conception is the most attractive specification of welfarism (by contrast with a veil-of-ignorance conception or a claim-within-outcome conception); and I used it to adjudicate between different SWFs. In this chapter, too, the claim-across-outcome conception is central -- now conjoined with the premise about the continuity of personal identity over a lifetime, and used to adjudicate between lifetime versus non-lifetime versions of the SWF framework.

Chapter 4 came down in favor of a continuous prioritarian SWF. I argued, first, that the claim-across-outcome view supports two key axioms: the Pigou-Dalton axiom and an axiom of separability-across-persons. "Prioritarian" SWFs satisfy both axioms. Next, I argued against prioritarian SWFs that fail a continuity axiom: the leximin SWF and the prioritarian SWF with an absolute threshold.

This chapter therefore focuses on comparing lifetime versus non-lifetime approaches to applying a continuous prioritarian SWF. I will try to demonstrate, here, that the claim-across-outcome view, conjoined with the premise about personal identity, supports *lifetime continuous prioritarianism* rather than the application of a continuous prioritarian SWF to individual sublifetime utility numbers or to individual attribute levels.

However, the case for using SWFs on a lifetime basis is really orthogonal to the choice between prioritarian and non-prioritarian SWFs, or between prioritarian SWFs that satisfy or fail to satisfy the continuity axiom. The case for the lifetime approach *does* hinge upon the claim-across-outcome conception of the moral ranking of outcomes, but it *doesn't* hinge upon the further assertion that this view is best specified via the continuous prioritarian SWF. I am confident that the basic line of argumentation I am about to present – to the effect that a continuous prioritarian SWF is best employed on a lifetime rather than non-lifetime basis -- can be reconfigured to defend a lifetime approach to *whichever* SWF the reader believes to be justified by the claim-across-outcome view.[9]

So much for preliminaries. The chapter begins by defending the premise that personal identity continues through a normal human lifetime. It then examines the structure of lifetime well-being. What is the functional form of the lifetime utility functions in **U**?

I next compare lifetime to non-lifetime versions of the continuous prioritarian SWF, and make the basic case for the lifetime approach: because personal identity continues over a normal human lifetime, a person's *claim* in favor of one or another outcome should depend upon her lifetime well-being.

[9] My argumentation below (1) assumes that the claim-across-outcome view justifies a ranking of outcomes that satisfies the axioms of Pareto indifference, Pareto superiority, Pigou-Dalton, and separability-across-persons; (2) argues that the proper currency for claims is lifetime well-being, since personal identity continues through a normal human lifetime; (3) points out that some non-lifetime approaches to implementing a continuous prioritarian SWF can violate one or more of the four axioms just mentioned, construed in lifetime terms; and (4) argues that even non-lifetime approaches to implementing a continuous prioritarian SWF which violate none of these axioms are problematic, because they represent a misleading way to think about the ranking of outcomes. This argumentation could be used, without alteration, to defend a lifetime approach to some prioritarian SWF that fails the continuity axiom (the leximin SWF or a prioritarian SWF with an absolute threshold), as against a non-lifetime approach to employing that SWF. Moreover, if one believes that the claim-across-outcome view is best understood to justify an SWF that fails the axiom of separability across persons (as does the rank-weighted SWF) or the Pigou-Dalton axiom (as do the utilitarian and sufficientist SWF) or both, then the argumentation could be amended so as not to rely upon those axioms. For example, one could point out that certain non-lifetime approaches to using a rank-weighted SWF can violate lifetime Pareto indifference, Pareto superiority, or Pigou-Dalton; and that even non-lifetime approaches to using a rank-weighted SWF which satisfy these lifetime axioms represent a problematic way to think about the ranking of outcomes.

The chapter then considers, and attempts to rebut, two important objections to this case for lifetime prioritarianism. One objection, pressed by Derek Parfit, sounds in personal identity. As we shall see, Parfit's account of personal identity does *not* undermine the premise that personal identity continues through a normal human lifetime. However, the account is "reductionist." It reduces personal identity to psychological and physical connections. There is no "deep further fact" of personal identity. Parfit suggests that "reductionism" of this sort cuts against lifetime prioritarianism and argues in favor of either sublifetime prioritarianism or utilitarianism.

A different kind of challenge, suggested by Dennis McKerlie's work as well as that of other scholars, trades on our intuitions about equalization. The argument, here, is that we have an intuitive preference for equalizing or synchronizing individuals' attributes or sublifetime well-being; and that lifetime prioritarianism, or any other lifetime approach to applying a distribution sensitive SWF, must conflict with these intuitions. I will argue that the lifetime prioritarian can generally parry these challenges by deploying a nuanced understanding of the structure of lifetime well-being.

A terminological point: Because prioritarian SWFs that fail the continuity axiom (the leximin SWF and the prioritarian SWF with an absolute threshold) are not discussed in this chapter, I will often omit the adjective "continuous." Throughout the remainder of the chapter, when I do so, and refer simply to "prioritarianism" or "the prioritarian SWF," I mean the *continuous* prioritarian SWF.

Personal Identity over Time

There is a vast literature in contemporary philosophy concerning personhood and personal identity. Reviewing this body of work in depth would take many more pages than I have available here. Still, I think it is fair to say that the common-sense view about personal identity – that a normal human being remains one and the same person from birth to death – is well supported by the philosophical literature.

One vital question concerns the conditions under which a human being is a person. Call this the problem of human personhood. Crudely speaking, there are two quite different possibilities here that are widely defended. One, adopted by many religious traditions, and defended by some contemporary philosophers (although a minority), is that a human being is a person in virtue of being associated with something like a soul: an immaterial substance of some kind which does not supervene upon the human's physical attributes.[10] A different possibility is that a human being is a person in virtue of having certain psychological attributes (such as consciousness, rationality, or a capacity for deliberation). Note that a human's psychological

[10] Properties of type *s* "supervene" on properties of type *b* if two items identical with respect to their *b* properties must be identical with respect to their *s* properties. To say that a soul doesn't supervene on a human's physical attributes means that two human beings who are physically identical may differ in whether they possess souls or what the souls are like.

attributes may well supervene on her physical attributes; indeed this is the standard view in the philosophy of mind. A psychological account of personhood which adopts the supervenience premise is clearly distinct from the "soul" account.

A different question concerns *individuation*. What is the criterion of personal identity that differentiates between one human person and a second, distinct, human person? The focus of the literature has been on questions of personal identity over time. If *g* is a human being and a person at time *t*, and *h* is a human being and a person at time *t**, under what conditions are *g* and *h* "numerically identical": the very same particular human person?

Crudely speaking, there are three different approaches to the problem of personal identity. One, which fits naturally with the "soul" account of human personhood, is that *g* and *h* are the same person if they have the same soul. Someone who adopts a psychological account of human personhood and denies the existence of souls cannot offer this answer to the question of personal identity. Interestingly, however, there are two quite different approaches that *are* open to her. One approach is to marry a psychological account of personhood with a psychological account of personal identity over time: to say that human person *g* at *t* is the very same person as human person *h* at *t** if the two are psychologically linked in a certain way. Another is to marry a psychological account of personhood with a *physical* account of personal identity over time: to say that *g* at *t* and *h* at *t** are the same person if they have the right sort of physical connection (e.g., if *g* at *t* has the very same body and brain as *h* at *t**, regardless of their psychological nexus). These latter two approaches can be hybridized. For example, one might say that *g* at *t* and *h* at *t** are the very same person only if they have both certain psychological links and certain physical connections.

This book assumes the psychological account of human personhood, rather than the "soul" view. There is a fixed population of *N* human beings, who are full human persons; and they are full persons, I assume, in virtue of having certain psychological properties.

How a welfarist moral view should be developed given a "soul" account of personhood is not a question I attempt to address here. Nor – the reader is reminded – do I consider other variations on the scenario of a fixed population of *N* human persons. In particular, I do not address how welfarism should cope with: a population of human persons that is variable rather than fixed; an infinite population of human persons; non-human persons (super-intelligent computers, angels, extraterrestrials); or humans who lack the psychological properties that are necessary for full personhood.[11]

[11] There are actually two cases here, neither of which I will attempt to address: human beings who are determinately not persons, and human beings who are indeterminate persons (at the "margins of personhood"). I leave these difficult cases aside, and focus on the case in which each of the *N* human persons is a full, determinate person, for most (if not all) of its existence as a person.

I say "most, if not all" because there is arguably a kind of temporary indeterminate personhood which arises even in the case of a normal human being, between the time when the human comes into being and the

Finally, I assume that the members of the population of *N* individuals have a *normal* psychological and physical history. Not only do they have the psychological properties that are necessary for full personhood, but they have not undergone brain transplants, traumatic brain injuries, psychological disease, or other unusual ruptures in their ongoing mental life or the physical makeup of their brains or bodies. Because the members of the population are "normal" in this sense, we can invoke a psychological account of personal identity over time, a physical account, or a hybrid account to justify the common-sense view that each such being is the very same person from birth to death.

Derek Parfit's work on personal identity in *Reasons and Persons* is worth introducing at this point. We will focus, later in the chapter, on the question whether the "reductionist" cast of this account argues against lifetime prioritarianism; and for those purposes it will be important to have a sense of the details of Parfit's view. But his work also helps illustrate the different possibilities concerning human personhood and personal identity over time. In particular, it is important to understand that Parfit's account of personhood and personal identity *confirms* the premise that personal identity continues through a normal human lifetime.

With respect to the question of personhood, Parfit pursues a psychological approach. He writes: "To be a person, a being must be self-conscious, aware of its identity and its continued existence over time." Parfit rejects the "soul" view – the view, as he puts it, that a person might be "a Cartesian Pure Ego, or spiritual substance.".

Reasons and Persons spends much more time on the question of personal identity. Here, Parfit offers a psychological criterion of personal identity, and leaves open the possibility that it might be hybridized with a physical criterion.

In constructing his criterion of personal identity, Parfit introduces the concepts of psychological "connectedness" and psychological "continuity." Consider human person *g* at time *t* and human person *h* at time *t**. There may be various direct connections between *g*'s mental states at *t* and *h*'s mental states at *t**. For example *h* at *t** may have a memory of an event that *g* experienced at *t*. Or, *g* at *t* may have the same belief, desire, or character trait as *h* at *t**. Or, *h* at *t** may consciously act on an intention that *g* at *t* formulated. If there are sufficient direct connections between *h* at *t** and *g* at *t*, then the two are *strongly connected*:

> Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person.

However, strong connectedness cannot itself be the criterion of personal identity. Personal identity is transitive. If *g* at *t* is the same particular person as *h* at *t**, and *h* at *t** the same particular person as *i* at *t***, then *g* at *t* is the same particular person as *i* at *t***. But strong

development of the psychological properties constitutive of personhood. I suggest below that the basic case for lifetime welfarism developed in this chapter is robust to this sort of indeterminacy (if it exists).

connectedness is not transitive. Consider a case in which a human being at age 75 remembers much of what that being experienced at age 50, and the human being at age 50 remembers much of what that being experienced at age 10, but the human being at age 75 remembers virtually nothing of what that being experienced at age 10.

Parfit therefore introduces the notion of psychological *continuity*. Person g at t is *continuous* with h at t^* if there is an overlapping chain of strong connectedness between the two persons. In other words, there is some series of pairs of persons and times $((j_1, t_1), (j_2, t_2), \ldots, (j_M, t_M))$, such that g at t is strongly connected with j_1 at t_1; each person in this series at the matching time is strongly connected with the next person at the matching time (so that j_1 at t_1 is strongly connected with j_2 at t_2, etc.); and j_M at t_M is strongly connected with h at t^*. And Parfit, then, offers a psychological criterion of personal identity, which appeals to psychological continuity.

> *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not taken a "branching" form. (5) Personal identity over time just consists in the holding of facts like (2) to (4). [12]

Parfit leaves open what "the right kind of cause" means: whether the psychological continuity between g at t and h at t^* must be caused by processes in a single brain shared by the two persons, or whether more esoteric causal processes are also permissible. (If "the right sort of cause" is indeed specified to require bodily continuity between g at t and h at t^*, then the upshot is a view of personal identity that hybridizes a psychological and physical criterion.) Finally, the requirement that the continuity be "non-branching" is inserted to deal with esoteric cases, for example a case in which my brain is split in half and put in two human bodies.

Because Parfit's account makes personal identity a matter of psychological links, not the sharing of a "soul," it raises the unsettling possibility that personal identity might be *indeterminate*: that one human person might be neither determinately identical to, nor determinately distinct from, another human person. Parfit argues for this possibility via discussion of a hypothetical spectrum of cases ("the Spectrum") in which a surgeon severs more and more of the psychological connections between a human being at one time and the same human being shortly thereafter. The surgeon starts, say, by removing one memory, then two, and so forth. At one end of this spectrum, the human being before and after the surgery are

[12] Lest the reader be confused by *Reasons and Persons* – a dense text, to be sure – it should be emphasized that the psychological criterion of personal identity, which is solely a matter of connectedness, should be distinguished from what Parfit calls "relation R," which is matter of *both* connectedness and continuity. Relation R is not the criterion of personal identity but, instead the criterion of "what matters" (as Parfit puts it), i.e., what humans rationally pursue. The very thrust of *Reasons and Persons* is that these two things can come apart. Although Parfit is occasionally ambiguous about this, see e.g RP p. 216, his initial presentation of the psychological criterion of personal identity makes clear that it is psychological *continuity*, not connectedness, which makes human beings at two different times the same particular person. See RP 204-09. And the rest of *Reasons and Persons*, with occasional lapses, sticks to this fairly consistently.

determinately the same person. (If only one memory has been removed, they clearly are strongly connected.) At the other end of the spectrum of possible surgeries, the human being before and after are determinately not the same person: they share no memories, desires, etc. Mustn't there, then, be some midrange of interventions where personal identity is indeterminate?[13]

I will not grapple with indeterminate identity here. Clearly, cases of indeterminate identity raise serious puzzles for the claim-across-outcome conception of fairness. If the single human being, Sam, is associated with two human persons who are neither determinately identical to each other, nor determinately distinct from each other, are those persons allocated two claims in ranking pairs of outcomes, one claim, or something in between? Some scholars have argued that, even on a psychological account, personal identity must in fact be determinate. So perhaps the puzzles are not genuine ones.

In any event, even if indeterminate identity *is* a genuine possibility in the scenario of the spectrum of surgical interventions or other scenarios, it does *not* arise in the case of the normal human life. The normal human being, Bod, at age 8 is determinately the very same person as the human being, Bod, at 85. Even though there may be few direct psychological connections between them (Bod at 85 cannot remember much of what Bod at 8 experienced, shares few of the desires that Bod had at 8, has a different emotional makeup, etc.), they are determinately continuous (Bod at 8 is strongly connected with Bod at 9, Bod at 9 with Bod at 10, and so forth), and the cause of this continuity (sharing the same, normal, human brain) is paradigmatically "the right kind of cause."

Indeed, Parfit pretty explicitly confirms that personal identity is determinate in the case of a normal human life. He writes: "In ordinary cases, questions about our identity have answers. In such cases, there is a fact about personal identity, and [the psychological criterion] is one view about what kind of fact this is. . . . In the problem cases [such as the spectrum of surgical interventions], things are different. [14]"[15]

[13] In this setup, the human being before and shortly after the surgery would be continuous only by virtue of their direct connections (there is no indirect chain linking them), and so the possibility of an intermediate mid-range of degree of connectedness raises the possibility of the two beings being indeterminately identical. Parfit also discusses a related spectrum in which the brain tissue of the two humans is less and less identical – which raises the spectre of indeterminate identity if one requires the psychological continuity constitutive of personhood to be grounded in the sharing of brain tissue.

[15] There is a different kind of indeterminacy that Parfit's account does raise. On this account, because personhood consists in psychological abilities or capacities, it seems plausible that a human being is only an indeterminate person at the beginning of human life. I am not endorsing this view, simply conceding its plausibility. See Parfit, RP, p. 322; McMahan, p. 44. In any event, it should be stressed that the potential indeterminacy, here, is *not* an indeterminacy regarding personal *identity*. It is not a matter of two human persons being neither determinately identical to each other, nor determinately distinct. Rather, it is a kind of indeterminacy concerning *personhood*: whether a particular being is a person. Moreover, unlike the case of humans with impaired psychological abilities throughout their lives, this is a case of *temporary* indeterminacy concerning personhood. Even if a normal human being is an indeterminate person for some time after the beginning of its existence as a human being (which itself might be understood to occur at conception or at some time before live birth), there is no question that the being

Lifetime Well-Being

This section clarifies the structure of the lifetime utility function. I argue, first, that the lifetime utility function is potentially quite fluid in its form: it need not be atomistic, separable with respect to attributes or times, or additive with respect to attributes or times. Because the problem of "discounting" is, in part, related to the structure of lifetime well-being, I also address that problem here – arguing, on this score, that the lifetime utility function should not incorporate a discount factor.

This Section relies upon the specific theory of well-being defended in Chapter 3: one that creates the set **U** of lifetime utility functions by pooling the utility functions that expectationally represent the fully-informed, fully rational, self-interested extended preferences of each member of the population over life-history lotteries and comparisons to nonexistence. The arguments presented in Chapter 4, in favor of a continuous prioritarian SWF, and the arguments presented in this chapter, in favor of using SWFs on a lifetime basis, do not essentially depend upon the extended-preference view of well-being. Nor, for that matter, does the point that the elements of **U** need not be atomistic, separable with respect to attributes or times, or additive with respect to attributes and times. Presumably any plausible account of lifetime well-being will have this flexibility.

Still, my discussion of the structure of lifetime well-being will be more persuasive and less abstract if undertaken with reference to a particular account of well-being. At the same time, this discussion will serve to provide a fuller understanding of the extended-preference view, and thus to lay the groundwork for Chapter 6– which discusses how to use a variety of data sources to actually estimate the lifetime utility functions in **U**. (Chapter 6 also grapples with a central question for the extended-preference framework – namely, what it means for an individual to have extended preferences regarding *simplified* possible worlds, which are missing some characteristics. The points I make in this section, regarding the functional form of the utility functions in **U**, is fully consistent with my discussion of the puzzle of simplification in Chapter 6.)

To reduce wordiness, I refer simply to an individual's "extended preferences," by which I mean her fully informed, fully rational, self-interested extended preferences.

The Structure of Lifetime Well-Being

Remember that an outcome can contain one or more periods. Each period will describe some of the attributes during that period of each of the *N* persons in the population, as well as background, "impersonal" facts (such as causal regularities). So each period *t* has the generic

eventually becomes a determinate person, remains one until death, and is determinately distinct from every other person. In this case, I see no obstacle to assigning each such person a single claim across outcomes, valenced in terms of her lifetime well-being. Remember that a claim, like lifetime well-being itself, is not temporally indexed: the person, atemporally, has a claim between two outcomes, depending on her lifetime well-being in them.

form $(\mathbf{a}_1^t, \mathbf{a}_2^t, ..., \mathbf{a}_N^t, \mathbf{a}_{imp}^t)$, with $\mathbf{a}_i^t$ the attributes of individual *i* during period *t*; and an outcome is a series of one or more such period-specific characterizations of persons' attributes plus background facts. A life-history, in turn, is a pairing of a person with an outcome. The individual attributes that figure most importantly in existing SWF scholarship or policy analysis more generally are: an individual's consumption (meaning either his consumption of particular marketed goods, or the total dollar value of his consumption); his leisure; the level of some public good he enjoys; his health state; and his hedonic state (an attribute that figures increasingly in policy analysis that draws on the happiness literature). The proponent of the SWF format is hardly committed to using these particular attributes; but it will be useful for the reader to have a sense of how the format is actually being used.

A standard philosophical distinction is the distinction between an individual's *monadic* or non-relational attributes, and her *polyadic* or relational attributes. Steve's property of being in pain is a non-relational attribute; his property of being the nephew of Julie is a relational property. As can be observed, the attributes that actually figure in policy analysis are either monadic, or at least do not express a relation between the person who possesses those attributes and other persons.[16]

Chapter 2 constructed the set **U** of utility functions by appealing to each individual's extended preferences over life-histories, life-history lotteries, and comparisons to nonexistence. More precisely, as I explained there, the format is flexible enough to allow for intertemporal change in a given individual's extended preferences.[17] Given a set of outcomes, a group of *N* individuals in the population, and some timeline, we can (in principle) ask about each of those individual's extended preferences over life-histories, lotteries, and comparisons to non-existence at each point in time. Of course, in order to simplify the task of estimation, it will be very convenient to assume that a given individual's extended preferences are the same at all points in time -- but nothing in the theory requires that.

Consider, then, the extended preferences of some individual *k* in the population, at some point in time *t** (more specifically, at the beginning of period *t**).[18] As a shorthand, we can refer to this individual as "the spectator" and the individual in any given life-history as "the subject." The spectator's extended preferences can be expectationally represented by a lifetime utility function, $u^{k,t*}(.)$, which is unique up to a positive ratio transformation. **U** is the set of all such

[16] It might be argued that some of these attributes express a relation to entities other than persons (for example, for Jim to consume a commodity is for him to stand in a particular kind of relation to it). So may be incorrect to describe them as monadic. On this issue, see below __

[17] It can be naturally generalized further to allow that an individual's preferences can be different in different outcomes. For simplicity I ignore that further generalization.

[18] Although it is possible for the timeline along which individuals' extended preferences change to be different from the periodization of outcomes, it is simplest to assume that these are the same. So each individual has one extended-preference ranking in period 1, one in period 2, and so forth.

functions, for all spectators and times. The "*k*" superscript, here, means that this is the utility function representing the extended preferences of individual *k*; the "*t**" superscript means that it represents her extended preferences at time *t**.

What will be the functional form of $u^{k,t*}(.)$? Let us say that it is *atomistic* if it depends solely on the subject's monadic attributes, relational attributes expressing a relation to entities other than persons, or background facts. In other words, $u^{k,t*}(.)$ is atomistic if it does not depend upon the attributes of other persons in the population, or on the subject's relational attributes that express a relation to other persons.

The utility functions that are actually used in SWF scholarship *are* typically atomistic. The prevalent approaches make a subject's utility a function of his own consumption (no one else's); or a function of his consumption and leisure (no one else's); or his health state (no one else's); or his level of exposure to a public good (no one else's).[19] But atomism is merely a pragmatic simplification, which eases estimation and formal analysis. For example, it may be worse, ceteris paribus, to have a given consumption level when others are a higher level, than when others are a lower level. Indeed, there is *some* SWF scholarship that employs non-atomistic utility functions.[20]

Even if the spectator's utility function *is* atomistic, it need *not* be separable or additive with respect to attributes or times.

Consider, first, the case in which outcomes have a single period. Each subject, let us assume, is characterized as having one or multiple attributes in that single period. The spectator's utility function (now assumed to be atomistic), is a function of those attributes[21] plus background facts. To say that $u^{k,t*}(.)$ is separable in those attributes means this: given two life-histories (x; i) and (y; j) in which background facts are the same, and in which certain types of attributes are identical in the two histories, the ranking of the histories does not depend on what particular level those attributes have. [22] Formally, if outcomes describe M types of attributes, and attributes of type 1…K are different as between life-histories (x; i) and (y; j), while the two histories have the same attribute of type K +1, the same attribute of type K+2, …, the same

[19] See Chapter 6 for more.

[20] See below.

[21] More precisely, of monadic attributes or attributes which do not express a relation to other persons.

[22] The literature on separability distinguishes between weaker and stronger separability conditions. See Blackorby/Primont/Russell review in Handbook of Utility Theory. The various separability conditions stated in this book, including the one in the text here, are generally variants of a stronger separability condition. Where I use "separability" to mean a weaker requirement, I will note that explicitly.

The reader might also note that background facts might be handled in different ways: we might have separability in subjects' attributes for any given specification of background facts; or background facts might be like an attribute, with the ranking of two life histories independent of the level of the subjects' attributes or impersonal facts where the same in both histories. For simplicity, and because my focus here is attributes, I have articulated the first formulation; but the second is possible as well.

attribute of type *M*, as well as the same background facts, the ranking of the histories should not depend upon what the *K*+1 attribute is, what the *K*+2 attribute is, and so forth.

The spectator's utility function need *not* be separable in this sense. To get a sense of how separability might fail, imagine that outcomes are specified in terms of the subject's health, consumption, and leisure. A spectator's ranking of different consumption/leisure packages, holding health fixed, might well depend on whether health is at a high or low level.

Even if $u^{k,t*}(.)$ is separable in the subject's attributes in the one-period case, it need not be *additive* in any given metric of the attributes.[23] In grasping the idea of "additive" utility functions, it is important to understand that a variety of metrics might be used to quantify individual levels of any given group of attributes. Thus a utility function is not "additive," simpliciter, but additive relative to some metric. In particular, $u^{k,t*}(.)$ is additive in the one-period case, relative to a given metric for measuring attributes, if $u^{k,t*}(x; i)$ is simply a linear function of the subject *i*'s levels of the attributes in *x* (as quantified with that metric). Each attribute is multiplied by some constant and these are summed.

For the simplest example of how attribute additivity can fail, consider the case in which outcomes have one period and individuals are characterized with respect to a single attribute, their consumption. Here, it is a familiar point that consumption may have "declining marginal utility" – in other words, that $u^{k,t*}(.)$ might take the form of a non-linear transformation (specifically, a concave transformation) of the subject's consumption, rather than being linear. Attribute additivity may also well fail relative to standard attribute metrics where outcomes describe a single non-consumption attribute, or where outcomes describe multiple attributes. For example, a standard utility function for outcomes characterized in terms of health and consumption makes utility the multiplicative product of consumption and health – so that utility is separable in health and consumption but not additive in health and consumption.[24] Utility as a function of leisure and consumption is sometimes represented as a non-linear function of leisure plus a non-linear function of consumption.

[23] The reader familiar with the literature regarding separability and utility functions may find this statement puzzling. A standard theorem shows that the strong sort of separability condition I am articulating here, together with a continuity condition, entails an additive representation. However,
to say that the ranking of life-histories achieved by $u^{k,t*}(.)$ can also be represented by a utility function $u^{+}(.)$ which is an additive function of the subject's attributes simply means that there is *some* attribute metric such that $u^{+}(.)$ is additive in that metric and represents the ranking of life-histories.

.

Moreover, $u^{k,t*}(.)$ has cardinal, not just ordinal properties: it expectationally represents the spectator's ranking of life history lotteries, not just her ranking of life histories. The additive function $u^{+}(.)$ might not be a linear transformation of $u^{k,t*}(.)$, and thus may fail to expectationally represent the spectator's ranking of lotteries. For example, a utility function equaling the multiplicative product of the level of health and the level of consumption is ordinally equivalent to the sum of the logarithms of those levels. However, the first utility function is not a positive linear transformation of the second, and will not expectationally represent the same preferences over life-history lotteries as the second.

[24] Another example: The utility of a health-consumption package is sometimes represented as a function of health plus a non-linear function of the consumption level.

Let us turn, now, to the case in which outcomes have multiple periods. Here, a very convenient assumption is that $u^{k,t^*}(.)$ can be expressed as a function of sublifetime utilities, each in turn a function of the subject's attributes during that period plus background facts. In other word, $u^{k,t^*}(x;i) = u^{k,t^*}[v^{k,t^*}(\mathbf{a}_i^1(x),\mathbf{a}_{imp}^1(x)), v^{k,t^*}(\mathbf{a}_i^2(x),\mathbf{a}_{imp}^2(x)),\ldots, v^{k,t^*}(\mathbf{a}_i^T(x),\mathbf{a}_{imp}^T(x))]$, where $(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$ denotes individual *i*'s attributes[25] during period *t* in outcome *x*, plus background facts in that period, and $v^{k,t^*}(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$ is the sublifetime utility of those features.

However, it should be stressed that nothing in the theoretical setup adopted here requires that $u^{k,t^*}(x; i)$ be expressible in this manner. In particular, if outcomes are described in terms of multiple attributes during each of a series of periods, and if the spectators' preferences fail to satisfy a condition of intertemporal separability in those attributes, the preferences may not be expressible in this manner. The condition of intertemporal attribute separability says: If the subject's attributes in life-history (*x*; *i*) in a given time period are the same as the subject's attributes in (*y*; *j*) in that period, for each of the *T* periods except one (*t**), the ranking of the histories should be solely a function of the subjects' attributes during *t**; it should not depend upon what the subjects' attributes are during the other period.[26]

Why believe that this intertemporal condition may fail? There is much evidence from the literature on preferences that individuals have preferences regarding the intertemporal sequencing of attributes.

> One of the most robust findings in research about assessment of experiences is the clear preference for improvement over time. … Preference for improvement has been demonstrated in many domains, including

[25] Given atomism, these would be individual *i*'s attributes that do not express a relation to other persons.

[26] The intertemporal attribute separability condition articulated here, by contrast with other separability conditions stated in this book, is a "weak" condition. It says that if attributes differ between life-histories in *one* period, and are held constant in others, then the ranking of life-histories should not be affected what the constant attributes are.

What exactly is the connection between this intertemporal attribute separability condition and the existence of a lifetime utility function *u*(.) which is decomposable into a function of sublifetime utilities (for short, a "decomposable" lifetime utility function)? Assume a finite number of periods, *T*. In the case where there are a finite or countable number of attribute combinations for each period, then it is clear that a given lifetime utility function *u*(.) in **U** will be decomposable – regardless of whether the intertemporal attribute separability condition is satisfied. Simply assign each possible combination of attributes, during each period, its own natural number. Each life history (*x*; *i*) is just a list of *T* natural numbers. The function *u*(.) can be written as a function of those numbers.

Where the number of attribute combinations for some periods is uncountable, matters become more complicated. A standard result in utility theory can be adapted to show that, if *u*(.) is continuous in attributes, and satisfies the intertemporal attribute separability condition just stated, it will be decomposable. Reciprocally, if *u*(.) is decomposable into sublifetime utilities, and *increasing* in sublifetime utility in each period, it is straightforward to see that the intertemporal attribute separability condition must be satisfied. (This is true, clearly, whether the number of attribute combinations is finite, countable, or uncountable.) But – in the uncountable cases -- mightn't there exist sublifetime utility functions, such that *u*(.) is not increasing in these, despite the failure of intertemporal attribute separability? This question raises many complex issues which I cannot pursue here. Suffice it to say that, in the uncountable case, the failure of intertemporal attribute separability may well jeopardize decomposability.

> monetary payments, experiences such as vacations, queuing events, pain, discomfort, medical outcomes and treatments, gambling, and academic performance.[27]

And various philosophers of well-being have argued that lifetime well-being is genuinely dependent on sequencing effects. Imagine, now, that the spectator's preferences for life-histories are sensitive to the intertemporal sequencing of each attribute taken individually. If so, the decomposition of $u^{k,t*}(.)$ into a function of sublifetime utilities, each in turn a function of the subject's attributes during the period plus impersonal facts, may *not* be possible.

Assume, however, that $u^{k,t*}(.)$ *is* expressible in this manner. It takes the form $u^{k,t*}(x;i) = u^{k,t*}[v^{k,t*}(\mathbf{a}_i^1(x), \mathbf{a}_{imp}^1(x)), v^{k,t*}(\mathbf{a}_i^2(x), \mathbf{a}_{imp}^2(x)), ..., v^{k,t*}(\mathbf{a}_i^T(x), \mathbf{a}_{imp}^T(x))]$. Even if this *is* true, $u^{k,t*}(.)$ may not be separable in *sublifetime utility*. To say that the spectator's lifetime utility function is separable in sublifetime utility is to say: whenever $u^{k,t*}(.)$ assigns two life-histories (*x*; *i*) and (*y*; *j*) the same level of sublifetime utility in one or more periods , the ranking of the life-histories is invariant to what that level or levels are.

Why might separability in sublifetime utility fail? Here, again, preferences for sequences emerge. Imagine that the spectator is sensitive to the sequence of the subject's sublifetime utility. Then $u^{k,t*}(.)$ will not be separable in sublifetime utility.

Finally, even if $u^{k,t*}(.)$ is expressible as a function of sublifetime utilities, and separable in these utilities, it need not be additive in sublifetime utility. The spectator's utility function is additive in sublifetime utility if it takes the form: $u^{k,t*}(x;i) = \sum_{t=1}^{T} v^{k,t*}(\mathbf{a}_i^t(x), \mathbf{a}_{imp}^t(x))$, or that form

[27] The literature regarding individual preferences for intertemporal sequences classifies a variety of phenomena under the heading of "sequencing effects." This includes a preference for improvement and other phenomena, such as preference for "spread." What these phenomena have in common, inter alia, is a failure of intertemporal separability.

To see how sequencing effects involve a failure of intertemporal separability, consider the very simple case in which individuals are ranking consumption sequence. Assume, for simplicity, that an individual ranks consumption sequences by summing amounts; as between consumption sequences with the same total amount, the individual ranks a steadily ascending sequence above a non-ascending sequence. Then the individual prefers (6, 7, 8, 9) to (9, 7, 8, 6), but (6, 20, 20, 9) and (9, 20, 20, 6) are ranked as equal, in violation of separability. [

Presumably a formal definition of sequencing effects will involve *more* than a failure of intertemporal separability. Consider the rank-weighted rule for ranking consumption sequences: consumption amounts are ordered from smallest to largest, and are weighted by fixed, declining weights. This violates intertemporal separability, but does not – intuitively – constitute a sequencing effect.

Plausibly, a sequencing effect involves *both* a failure of intertemporal separability *and* a failure of an intertemporal permutation condition. Regardless of the cogency of this definition, the important point here is that individual preferences over consumption sequences or other sequences of attributes can fail an intertemporal separability condition.

with sublifetime utility in each period adjusted by some discount factor.[28] Imagine, instead, that lifetime utility is the multiplicative product of sublifetime utility.

I have described a wide variety of different possible approaches to structuring a lifetime utility function as regards atomism versus nonatomism, separability versus nonseparability in attributes or times, and additivity versus nonadditivity in attributes or times. All of these are, in principle, possible and fully consistent with the basic idea that this function represents one or another spectator's fully informed, fully rational, self-interested extended preferences over life histories.

What, in fact, is the standard approach? (1) As mentioned, SWF scholarship and policy analysis more generally *does* typically assume that lifetime utility functions are atomistic. (2) Moreover, in evaluating multi-period outcomes, SWF scholarship and policy analysis more generally *does* assume that the lifetime utility function is expressible as a function of sublifetime utility, and indeed is additive in sublifetime utility. (3) However, SWF scholarship and policy analysis often does *not* assume that the sublifetime utility function *v*(.) is separable in the period-specific individual attributes which are the arguments for that function. And even where such separability *is* assumed, sublifetime utility may be a non-additive function of those attributes. (4) Similarly, where a one-period outcome is employed, SWF scholarship often does not assume that the lifetime utility function (which, in this case, *is* also a one-period utility function) is separable in the individual's attributes during the single period, let alone additive in those attributes.

I will adopt the standard assumptions mentioned under (1) and (2), when I turn to questions of estimations in Chapter 6. In other words, I will assume the lifetime utility function does indeed take the form $u^{k,t^*}(x;i)=\sum_{t=1}^{T}v^{k,t^*}(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$. However, it should be reiterated that these assumption are just heuristic devices. They economize on the time and expense of SWF analysis, but can be relaxed if a fuller and more nuanced analysis is desired. The concept of lifetime well-being and the extended-preference account does not necessitate them.

Does the Lifetime Utility Function Incorporate a Discount Factor?

[28] As with attribute additivity, so additivity in sublifetime utility is really relative to a metric. Even if u^{k,t^*}(.) does take the form given in the text, there may be some other metric v^+(.) of sublifetime utility, such that v^+(.) is a non-linear transformation of v^{k,t^*}(.). This point will be of some importance in the last section of the chapter.

The point made above, regarding cardinally versus ordinally equivalent lifetime utility functions, is also relevant here. See footnote ___. If u^{k,t^*}(.) is not only expressible in terms of a sublifetime utility function v^{k,t^*}(.), but separable in sublifetime utility, then (if u^{k,t^*}(.) is also continuous in v^{k,t^*}(.)), there will be a u^+(.) which is additive in some metric of sublifetime utility and represents the ranking of life-histories achieved by u^{k,t^*}(.) . However, u^+(.) may not be a linear transformation of u^{k,t^*}(.), and thus may not expectationally represent the ranking of life-history lotteries achieved by u^{k,t^*}(.). For example, if u^{k,t^*}(.) is the multiplicative product of v^{k,t^*}(.) values, and u^+(.) is the sum of the logarithms of these values, u^{k,t^*}(.) and u^+(.) rank life histories the same way but are not linear transformations of each other.

Does some sort of time preference or discount factor properly figure in moral decisionmaking? This is a complex and multifaceted problem. This section addresses *one* facet of that problem, namely this:[29] To what extent do the utility functions in **U**, which measure the well-being associated with life histories, and represent various spectators' extended preferences, properly incorporate a time preference or discount factor?

In thinking about the question, it is crucial to distinguish between three, analytically distinct, ideas. One is that a given spectator's utility function $u^{k,t*}(.)$ may or may not be *sensitive to the temporal location* of subjects' attributes. A second is that the spectator may or may not be *temporally neutral*. A third is that $u^{k,t*}(.)$ may or may not adopt the *standard discounted utility form*, very widely used in economics – meaning that $u^{k,t*}(.)$ is represented as the sum of the subject's sublifetime utility in each time period (starting either in the first period, or in period $t*$), multiplied by a discount factor which decreases with time.

Start with the first concept. Sensitivity to temporal location means that the spectator's ranking of life-histories depends, to some extent, upon whether a subject realizes a given attribute at one rather than another point in time. This feature of extended preferences might be formally expressed in various ways, but is most simply formalized via an intertemporal permutation condition: if one life-history has a certain package of attributes in each period, and another life-history simply rearranges the order of the packages, then the spectator is insensitive to temporal location if he is indifferent between the life-histories.

I see no reason to insist that $u^{k,t*}(.)$ be insensitive to the temporal location of subjects' attributes. For example, I noted in the previous subsection that $u^{k,t*}(.)$ might exhibit sequencing effects: the spectator might care about the sequence of the subject's attributes or sublifetime utility. If so, $u^{k,t*}(.)$ will (or at least may) fail the intertemporal permutation condition just stated.[30]

What about the second idea, temporal neutrality? A standard philosophical position regarding intertemporal choice and preference is that the rational, self-interested individual is temporally neutral, in the sense of giving equal weight to the different temporal periods in her life. This is called "prudence," and various philosophers (as well as some economists) have recommended it. Sidgwick famously argued in favor of this view when he wrote:

[29] A different aspect of the problem is whether the SWF itself incorporates a discount factor, for example giving priority to the current over future generations. See Chapter 8.

[30] A plausible definition of sequencing effects includes a failure of an intertemporal permutation condition as well as an intertemporal separability condition. See supra note __. In any event, paradigmatic sequencing effects that might well characterize spectators' preferences over life histories violate the permutation condition. Consider the preference for improvement. If, for example, outcomes are characterized solely in terms of consumption, the spectator might prefer a life history in which the subject's consumption in each period is less than the next, to a life history in which the subject realizes the very same consumption amounts in a decreasing sequence.

> The proposition "that one ought to aim at one's own good" is sometimes given as the maxim of Rational Self-love or Prudence: but as so stated it does not clearly avoid tautology, since we may define "good" as "what one ought to aim at." If, however, we say "one's good on the whole," the addition suggests a principle … [which is] not tautological. I have already referred to this principle as that of "impartial concern for all parts of our conscious life": -- we might express it concisely by saying that "Hereafter *as such* is to be regarded neither less nor more than Now." It is not, of course, meant that the good of the present may not reasonably be preferred to that of the future on account of its greater certainty …. All that the principle affirms is that the mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment that to that of another.

A similar position was taken by the famous economists Ramsey and Pigou, as well as by Rawls, and has been adopted by various contemporary philosophers.

One puzzle in the philosophical literature has been articulating a clean distinction between temporal bias (a failure of temporal neutrality) and sensitivity to the temporal location of attributes. On the one hand, scholars who recommend "prudence" may well recognize, indeed embrace the point that an individual's well-being over a lifetime may depend on sequencing effects. And yet they also endorse temporal neutrality. How are these both possible?

One attractive feature of my account of well-being is that it helps to answer this question. *Sensitivity to the temporal location of attributes* concerns the content of a spectator's preferences over life-histories at a given point in time. It concerns the functional form of any given $u^{k,t*}(.)$. *Temporal neutrality,* on the other hand, concerns the *basis* for the spectator's preferences. It says that the spectator shall not be influenced (in a certain manner) by his position in time in ranking life-histories. In deciding how to rank life-history (x; i) relative to life-history (y; j), the spectator may take account of his own tastes regarding attributes or their patterning (including their intertemporal patterning); he may, in particular, rank life-histories so as to prefer a particular sequence of attribute packages over a permutation of that sequence; but the spectator's ranking should not be improperly influenced by the fact that he himself is located at time $t*$ rather than some other time $t**$.[31]

One kind of temporal bias that is frequently discussed in the literature is discounting for temporal distance. Imagine that the spectator gives more weight to attributes that occur more proximately to him in time than to more distant attributes. Most simply: imagine that the spectator's preferences over life-histories are representable as an additive function of the subject's sublifetime utilities in each period, each multiplied by an adjustment factor so that periods further away from the spectator in time are deflated, and periods closer to him are inflated.

[31] A precise statement of temporal neutrality is tricky. I have stressed that it is legitimate for the spectator's extended preferences to change over time. Many such legitimate changes will be "influenced" by the spectator's position in time. (For example, imagine that the ageing of a given spectator tends to cause her to care less about material consumption, and more about non-consumption attributes.). Temporal bias involves certain *illegitimate* influences, such as discounting the past or discounting for temporal distance. However precisely these illegitimate influences are delineated, the key point is that a temporal neutrality requirement concerns the causal influence of the spectator's temporal position on his ranking of life histories, rather than the content of the ranking.

Another kind of temporal bias is ignoring the past and focusing on the future. Most simply: imagine that the spectator's preferences over life-histories at time *t* are representable as an additive function of the subject's sublifetime utilities in each period, but the periods prior to the time *t* are assigned a weight of zero.

The conjunction of these two sorts of temporal biases (plus the additional premise that $u^{k,t*}(.)$ is additive in sublifetime utility) produces one version of the standard discounted utility model. At a given point in time, the spectator ranks life histories according to the discounted sum of sublifetime utility, *beginning in period t** (where the spectator is located). In other words, $u^{k,t*}(x;i)=\sum_{t=t*}^{T}D(t)v^{k,t*}(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$, where *D*(*t*) is a discount factor that decreases as *t* increases. If the spectator's utility function at *t** has this structure, the subject's sublifetime utility in the more distant future is discounted, relative to his sublifetime utility in the nearer future; the subject's sublifetime utility in the future is discounted, relative to his sublifetime utility in the current period (period *t*); and his sublifetime utility in the past is completely ignored.

So should the spectator's preferences over life-histories be temporally neutral? We have "laundered" and idealized those preferences in various ways. Is temporal neutrality an additional requirement?

I think the answer is yes. Lifetime well-being is a feature of outcomes that lacks temporal location. It is indexed to persons, but not times. To say that one outcome is better for lifetime well-being, full stop, is a conceptual error. Individual *i* has greater lifetime well-being in outcome *x* than *y* iff outcome *x* is better *for individual i* than outcome *y*. However, it is equally a conceptual error to say that individual *i*'s lifetime well-being in outcome *x* is greater at time *t* than his lifetime well-being in outcome *y*. *Sublifetime* well-being has this temporal indexing, but lifetime well-being does not.

In other words, lifetime well-being is *atemporal* in its very structure. It is counterintuitive, then, to think that the extended preferences in light of which lifetime well-being is to be analyzed can be temporally biased.

A more pragmatic and perhaps more powerful argument for temporal neutrality is that temporal bias might well cause large-scale incomparability in the well-being ranking of life-histories. Consider, in particular, the case in which the spectator ignores the subject's past sublifetime utility, and discounts the subject's future sublifetime utility. At each point in time *t**, the spectator's utility function $u^{k,t*}(.)$ is captured by the formula $u^{k,t*}(x;i)=\sum_{t=t*}^{T}D(t)v^{k,t*}(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$. Consider now, any pair of life-histories (*x*; *i*) and (*y*; *j*), such that $u^{k,1}$ (.), at point in time 1 (the beginning of the first period), assigns (*x*; *i*) a larger subutility than (*y*; *j*) in some period t^+. Then it is quite possible that the spectator's utility functions for periods 1 through t^+ prefer (x; i) to (y; j); but that his utility functions for periods t^++1 through the final period T prefer (y; j) to (x; i), because the subjects' attributes during t^+ drop out of the formula. Pooling together all these utility functions in **U** will make the two life histories incomparable. In other words, by having a utility function that ignores the past and sums a discounted value of the subject's present and future sublifetime utility, the spectator's ranking of life histories may well "switch" as she moves forward in time – yielding incomparability when the rankings are pooled.[32]

It should be stressed that in requiring the spectator's extended preferences over life-histories to be temporally neutral, I am *not* imposing a more demanding requirement that these preferences be fixed – that they be identical at all points in time. As already explained, my account generally allows for intertemporal preference change. What the temporal neutrality requirement does is to preclude certain *causes* of such change, such as discounting for temporal distance or discounting the past. [33]

Further, I am not offering temporal neutrality as a *general* account of rational self-interested behavior. My claim is only that, *for purposes of* constructing an account of lifetime well-being in terms of preferences – specifically, extended preferences – *those* preferences will need to be temporally neutral. Like the "self-interest" and "full information" requirements, this is a kind of idealization that plays an important role *within* the account, but need not be seen as a universal feature of rational choice.

A final, interesting point is that the spectator's preferences over life-histories might assume one sort of discounted utility form *even if the spectator is temporally neutral*. Imagine that the spectator's ranking of life-histories is additive in sublifetime utility. Further, this ranking is *not* influenced by the spectator's temporal distance from the subject's attributes, or by whether those attributes are past or present. Indeed, assume that the spectator, at all times, has the very same ranking of life histories. However, the spectator *does* care about the date at which sublifetime utility occurs. The earlier it occurs, the greater weight it gets.

In other words, the spectator's extended preferences are represented by a single, intertemporally fixed utility function u^k(.), of the form: $u^k(x;i) = \sum_{t=1}^{T} D(t) v^k(\mathbf{a}_i^t(x), \mathbf{a}_{imp}^t(x))$

[32] Note also that requiring the discount factor, here, to be exponential rather than "hyperbolic' or some other non-exponential form would not resolve the problem.

[33] Further, unlike some philosophers of prudence, my account does not require that the spectator at a given point in time "step back" from his preferences at that moment, and formulate preferences over life-histories that give equal weight to his own preferences at all moments. The spectator at t^*, on my account, can fully rely on his own tastes, attitudes, and beliefs about values at t^*, ignoring his tastes, attitudes, and beliefs at some other time. To be sure, my formula for assigning well-being to life-histories does draw upon all the utility functions in **U** (including all of the utility functions, over time, of a given spectator); but sensitivity to one's extended preferences at other times is not seen, by my account, as part of what it means to have a fully rational extended preference.

There is no temporal bias here. (By contrast with the discounting formula used earlier, the spectator continues to care about the subject's attributes during periods prior in time to the spectator's own location.) Nor will pooling the spectator's utility functions at different points in time yield incomparability (since these are identical). Rather, there is a certain kind of sensitivity to temporal location. (Note that this utility function fails an intertemporal permutation condition.)

Although I generally allow that spectators' ranking of life histories can be sensitive to the temporal location of subjects' attributes, this particular sensitivity seems irrational. Why should the date at which sublifetime utility occurs, as opposed to its sequencing,[34] influence the well-being of a life-history? It is hard to see how the fully informed, fully rational, self-interested spectator would have this sort of preference – as opposed to more plausible violations of the intertemporal permutation condition.

In sum: the utility functions in **U** represent spectators' fully informed, fully rational and temporally neutral extended preferences. The temporal neutrality requirement does not, as such, limit *how* the spectator ranks life-histories. Rather, it concerns whether the spectator ranks one history over another *because of* his own particular position in time. In particular, temporal neutrality precludes the spectator at a given time *t** from giving greater weight to future attributes rather than past attributes (attributes in periods after rather than before *t**), or to attributes that occur in periods closer to *t** rather than attributes more distant from *t**. Thus $u^{k,t*}(.)$ should not assume the discounted-utility form, $u^{k,t*}(x;i) = \sum_{t=t*}^{T} D(t) v^{k,t*}(\mathbf{a}_i^t(x), \mathbf{a}_{imp}^t(x))$. Nor should it assume a different discounted-utility form, $u^{k,t*}(x;i) = \sum_{t=1}^{T} D(t) v^{k,t*}(\mathbf{a}_i^t(x), \mathbf{a}_{imp}^t(x))$, which satisfies temporal neutrality but seems irrational.

The Basic Case for Lifetime Prioritarianism

A lifetime approach to applying the continuous prioritarian SWF employs individuals' lifetime utilities as the arguments for the SWF. It uses some account of lifetime well-being *W* to generate a ranking of life-histories, differences between life-histories, and comparisons to zero in each choice situation, which are in turn measured by a set **U** of lifetime utility functions. (I have argued for a particular version of *W*, namely one that appeals to extended preferences.) The lifetime prioritarian approach then chooses some *g*(.) function and ranks outcomes using the formula: outcome *x* is at least as good as outcome *y* iff, for all *u*(.) belonging to **U**, $\sum_{i=1}^{N} g(u_i(x) \geq \sum_{i=1}^{N} g(u_i(y))$.

[34] Note that this formula, by contrast with a utility function $u^{k,t*}(.)$ that is sensitive to sequencing effects, *satisfies* intertemporal separability conditions (both separability in sublifetime utility, and intertemporal attribute separability).

This decisional procedure can be contrasted with various *non-lifetime* rules for using the continuous prioritarian SWF. As we shall see later in the chapter, substantial philosophical arguments have been advanced in favor of some version of non-lifetime prioritarianism. Although I believe that these arguments fail, non-lifetime prioritarianism is worth taking seriously. Within the SWF framework, non-lifetime prioritarianism would mean using numbers *other than lifetime utilities* as the arguments for the continuous prioritarian SWF.

One variant of non-lifetime prioritarianism is *sublifetime* prioritarianism. The general idea, here, is that the inputs to the SWF are sublifetime utility numbers, representing well-being during temporal segments of human lifetimes. Assume an outcome set in which outcomes have T periods. A simple kind of sublifetime prioritarianism would assign each individual a sublifetime well-being number for each period, so that each outcome corresponds to a grand vector of NT sublifetime utilities (or a set of such vectors), and outcomes are ranked by applying the continuous prioritarian SWF to these grand vectors.

A different version of sublifetime prioritarianism divides individual i's life into a certain number of non-overlapping segments, where each segment may consist in one or more periods. If individual i's life in outcome x is divided into $s_i(x)$ non-overlapping segments, then individual i is assigned $s_i(x)$ sublifetime utility numbers. (For example, if outcomes have 10 periods, individual i might be assigned a sublifetime utility number in x for a segment encompassing periods 1 and 2, another number for a segment encompassing periods 3 through 7, and a third number for a segment encompassing periods 8 through 10.). Outcome x then corresponds to a grand vector of $\sum_{i=1}^{N} s_i(x)$ sublifetime utilities (or a set of such vectors), and the continuous prioritarian SWF is applied to *these* vectors.

Yet a different approach divides individual i's life into a certain number of segments which may overlap, where each segment may consist in one or more periods. For example, if outcomes have 10 periods, individual i might be assigned a sublifetime utility number for a segment encompassing periods 1 through 5, a second for a segment encompassing periods 2 through 6, a third for a segment encompassing periods 3 through 7, and a fourth for a segment encompassing periods 4 through 10. If individual i's life in outcome x is divided into $s_i(x)$ potentially overlapping segments, outcome x then corresponds to a grand vector of $\sum_{i=1}^{N} s_i(x)$ sublifetime utilities (or a set of such vectors), and outcomes are then ranked by applying the continuous prioritarian SWF to these vectors of sublifetime utilities.

Actually, this last sublifetime approach (allowing for overlapping segments) can be seen as a generalization of the second, which in turn is a generalization of the first (where segments don't overlap and each segment is just a single period, so that $s_i(x) = T$ for each individual and each outcome.). So we can summarize sublifetime prioritarianism as follows. There is some set

V of sublifetime utility functions. While lifetime utility functions assign utilities to whole life histories, each *v*(.) in **V** assigns a utility number to a temporal segment of a life-history. Formally, we have not merely the set **H** of life-histories, but a set **F** of temporal segments of life-histories; and each element of **V** assigns a number to each member of **F**, representing the sublifetime well-being of that segment.[35] For each outcome *x*, the life of each individual *i* is divided by some rule into $s_i(x)$ segments. We can use the symbol (*x*; *i*; 1) to represent the first segment of individual *i*'s life in *x*, (*x*; *i*; 2) to represent the second segment, and in general (*x*; *i*; *s*) to represent segment number *s* of individual *i*'s life in outcome *x*. Each sublifetime utility function *v*(.) of **V** maps a given outcome *x* onto a grand vector of $\sum_{i=1}^{N} s_i(x)$ sublifetime utilities. That is, outcome *x* is mapped onto $v(x)=(v(x; 1;1), v(x; 1; 2), \ldots,v(x; 1; s_1(x)),\ldots, v(x; i;1), v(x; i; 2), \ldots,v(x; i; s_i(x)),\ldots, v(x; N;1), v(x; N; 2), \ldots,v(x; N; s_N(x))$. And the rule for ranking outcomes is: outcome *x* is morally at least as good as outcome *y* iff, for all *v*(.) in **V**,

$$\sum_{i=1}^{N}\sum_{s=1}^{s_i(x)} g(v(x;i;s)) \geq \sum_{i=1}^{N}\sum_{s=1}^{s_i(y)} g(v(y;i;s))$$

, with *g*(.) strictly increasing and concave.

In the simplest case where each segment in each outcome is a single period, this formula becomes: *x* is morally at least as good as *y* iff, for all *v*(.) in **V**,

$$\sum_{i=1}^{N}\sum_{t=1}^{T} g(v(x;i;t)) \geq \sum_{i=1}^{N}\sum_{t=1}^{T} g(v(y;i;t))$$

, where (*x; i; t*) denotes period *t* of individual *i*'s life in outcome *x*; and *v*(*x*; *i*; *t*) is the sublifetime utility assigned to this single period. I will generally refer to this variant of sublifetime prioritarianism as "simple" sublifetime prioritarianism.

A different kind of non-lifetime prioritarianism is *attribute-based* prioritarianism. Assume that, in a given outcome set, *M* types of individual attributes are described during each period. (For example, if each individual's health and consumption is described in each period, then *M* = 2.). Use the symbol $a^m(x; i; t)$ to mean the numerical level of attribute *m* that individual *i* has in outcome *x* during period *t*. Then each outcome corresponds to a grand vector of *NTM* attributes (*N* individuals, *T* periods, *M* attributes per period). Attribute-based

[35] Let us say that a "segment" is a group of one or more consecutive periods of a given outcome, paired with a person. (We could generalize further to allow for segments with non-consecutive periods, for example the segment consisting of being individual *i* in outcome *x* in period 1 plus being individual *i* in outcome *x* in period 4; but it is intuitively very odd to assign sublifetime well-being to such segments, and the kinds of sublifetime prioritarianism suggested by Parfit, McKerlie, and other critics of lifetime prioritarianism do not involve such segments.)

Consider each life-history (*x*; *i*). If outcomes have *T* periods, this history is associated with *T* segments that have one period, *T*-1 that have two periods, and so forth – with *T*(*T*+1)/2 in total. **F** is either the grand set of all these, for all life-histories in **H**; or some subset. (A particular version of sublifetime prioritarianism might only need to assign sublifetime utilities to certain segments.) If *s* and *s** are two segments in **F**, then the sublifetime well-being associated with them is represented by the utility functions in **V** (just as **U** represents the lifetime well-being associated with the life-histories in **H**.) In other words, the sublifetime well-being realized by *s* is at least as large as the sublifetime well-being realized by *s** iff, for all *v*(.) in **V**, $v(s) \geq v(s^*)$. The elements of *v*(.) might also represent the differences in sublifetime well-being between the elements of **F**, or comparisons to zero (where the subject is not alive during a period).

prioritarianism ranks outcomes by applying the continuous prioritarian SWF to these attribute vectors. In other words, it says: outcome *x* is morally at least as good as outcome *y* iff

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} g(a^{m}(x;i;t)) \ge \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} g(a^{m}(y;i;t))$$.[36]

The following table provides a simple example which illustrates sublifetime, attribute based, and lifetime prioritarianism

	Outcome x			Outcome y		
	Period 1	Period 2	*Lifetime*	Period 1	Period 2	*Lifetime*
Joe						
Attribute *a*	100	10		81	100	
Attribute *b*	49	100		64	1	
Sublifetime	4900*c*	1000*c*	5900*c*	5184*c*	100*c*	584*c*
Sue						
Attribute *a*	36	49		81	49	
Attribute *b*	25	25		36	9	
Sublifetime	900*c*	1225*c*	2125*c*	2916	441*c*	3357*c*

Attribute score: 53.16228	Attribute score: 53
Sublifetime score 166.6228√c	Sublifetime score:157√c
Lifetime score: 122.9092√c	Lifetime score: 130.6308√c

In this example, there are two individuals, two time periods, and two attributes in each period. The attribute-based approach, with the square root as the *g*-function, is to assign an outcome a score equaling the sum of the square roots of all the attribute levels in the outcome. The simplest sublifetime approach, again with the square root as the transformation function, is to assign an outcome a score equaling the sum of the square roots of each individual's sublifetime utility in each period. The lifetime approach, once more with the square root as the transformation function, is to assign the outcome a score equaling the sum of the square roots of each individual's lifetime utility.

The table illustrates these approaches. It is assumed that sublifetime utility, for purposes of the sublifetime approach, is the product of the attribute levels; and that lifetime utility is the sum of these same sublifetime utilities.

Note that the attribute and sublifetime approaches here both rank *x* over *y*, while the lifetime approach ranks *y* over *x*.

Before presenting the basic case for the lifetime approach, several preliminary points are worth underscoring. To begin, sublifetime utilities play a different role within sublifetime prioritarianism than they do within lifetime prioritarianism. As I discussed in the previous section, the utility assigned to a life-history by each lifetime utility function in **U** *might* be

[36] This approach presupposes some methodology for measuring the level of each attribute. By contrast, lifetime and sublifetime prioritarianism do not require that attributes be assigned numerical levels. The characterization of outcomes *may* express attributes numerically, but it also may describe some or all attributes in qualitative terms (for example, by describing an individual's health attribute in terms of whether she is healthy or, if not, what diseases she has). And lifetime or sublifetime utility functions can take as their arguments life-histories, or periods of life histories, in which some or all attributes are characterized qualitatively.

expressible as a function of the subject's sublifetime utility in each period (most simply, as the sum of sublifetime utility in each period). However, nothing in the concept of lifetime well-being or in the extended-preference account for assigning lifetime utilities requires that lifetime utility be expressible in this manner. Moreover, even where lifetime utility *is* expressible as a function of sublifetime utility, lifetime prioritarianism uses individuals' lifetime utilities – not their sublifetime utilities – as the direct arguments for the prioritarian SWF.

By contrast, sublifetime prioritarianism is *committed* to the existence of sublifetime utility numbers. These numbers are "fed" directly into the sublifetime formula, i.e., outcome *x* is at least as good as outcome *y* iff, for all sublifetime utility functions *v*(.) in **V**,

$$\sum_{i=1}^{N}\sum_{s=1}^{s_i(x)} g(v(x;i;s)) \geq \sum_{i=1}^{N}\sum_{s=1}^{s_i(y)} g(v(y;i;s))$$

A related point is that the sublifetime prioritarian might deploy a range of different methods for constructing the sublifetime utility functions in **V**. Assume (as I have argued) that an extended preference account is the most attractive methodology for arriving at lifetime utilities. The sublifetime prioritarian might: (1) construct the set **V** by looking to spectators' extended preferences over temporal segments rather than whole life histories (if the notion of an extended preference regarding a temporal segment is indeed coherent); (2) stipulate that spectators' preferences over whole life histories must satisfy conditions sufficient to ensure that the lifetime utility functions in **U** *are* expressible as a function of subjects' sublifetime utilities, and use **U** to construct **V**; or (3) in some other manner. My critique of sublifetime prioritarianism will be entirely agnostic on these issues.

A final, subtle point is that we can construct non-lifetime approaches which are extensionally equivalent to lifetime prioritarianism. Take any arbitrary account of lifetime well-being *W*, which produces some set **U** of lifetime utilities in each choice situation. Then we can "gerrymander" a version of sublifetime prioritarianism which yields the very same ranking of outcomes, in each choice situation, as a lifetime prioritarian SWF using **U**. In other words, there is some strictly increasing, concave *g*(.) function such that the ordering of outcomes achieved by this "gerrymandered" version of sublifetime prioritarianism is exactly the same as the ordering achieved by the lifetime rule: *x* is at least as good as *y* iff, for all *u*(.) in **U**, $\sum_{i=1}^{N} g(u_i(x) \geq \sum_{i=1}^{N} g(u_i(y))$.[37] Moreover (with certain assumptions), we can "gerrymander" a

[37] To gerrymander sublifetime prioritarianism, note that some philosophers of well-being suggest that a individual's well-being during some period of time need not be solely a function of contemporaneous facts or attributes. Consider the simplest version of sublifetime prioritarianism, where each segment is a single period, and (*x*; *i*; *t*) denotes period *t* of individual *i*'s life in outcome *x*. Let us allow *v*(*x*; *i*; *t*) to depend upon all of individual *i*'s attributes in *x*, not just her attributes during period *t*. For a given set of lifetime utility functions **U** and an outcome set with *T* periods, arbitrarily pick two increasing and strictly concave functions *g*(.) and *f*(.). For a given *u*(.), define a corresponding *v*(.) as follows: for a given individual *i* and outcome *x*, her sublifetime utility in every period is the very same value, namely $v(x;i;t) = f^{-1}(g(u(x;i))/T$. Define **V** as the set of all *v*(.). Then the following

version of attribute-based prioritarianism which yields the very same ranking of outcomes, in each choice situation, as a lifetime prioritarian SWF using **U**.[38]

This observation, in turn, is relevant to the question whether non-lifetime approaches necessarily violate the lifetime Pareto, Pigou-Dalton, and separability-across-persons axioms. These are the key substantive axioms satisfied by lifetime prioritarianism.[39]

The "lifetime" Pareto indifference principle is just the Pareto indifference principle that has been discussed throughout the book, starting in chapter 1. I add the adjective "lifetime" to underscore that this principle has been framed in terms of lifetime well-being: in terms of the well-being associated with life-histories. (Lifetime) Pareto indifference says: if (*x*; *i*) is equally good as (*y*; *i*), for every individual *i*, then *x* and *y* are equally morally good. In other words, if each individual's lifetime well-being is the same in two outcomes, the outcomes are morally indifferent. The Pareto superiority principle discussed throughout the book is also a "lifetime" principle, similarly framed in terms of the well-being associated with whole life-histories. It

sublifetime-prioritarian rule produces the very same ordering of outcomes as the lifetime approach using *g*(.): *x* is at least as good as *y* iff, for all *v*(.) belonging to **V**, $\sum_{i=1}^{N}\sum_{i=1}^{T} f(v(x;i;t) \geq \sum_{i=1}^{N}\sum_{i=1}^{T} f(v(y;i;t))$

[38] Assume that the utility functions in **U** are atomistic and, further, insensitive to background facts; and that **U** is unique up to a positive ratio transformation. Arbitrarily pick *g*(.) which is Atkinsonian and thus invariant to a ratio transformation. (In other words, $g(u(x;i)) = \frac{1}{1-\gamma}u(x;i)^{1-\gamma}$.) Arbitrarily pick some strictly increasing, strictly concave *f*(.). If there are *T* periods and *M* attributes per period, arbitrarily pick some *u**(.) in **U** and assign *every* attribute of individual *i* in outcome *x* in each period the very same attribute level, namely $a^m(x;i;t) = f^{-1}(g(u^*(x;i))/MT)$. Refer to this single level as *a*(*x*; *i*). Attribute-based prioritarianism, using the *f*(.) function, says: *x* is at least as good as *y* iff $\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} f(a^m(x;i;t)) \geq \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} f(a^m(y;i;t))$. This becomes $\sum_{i=1}^{N} MT[g(u^*(x;i))/MT] \geq \sum_{i=1}^{N} MT[g(u^*(y;i))/MT]$, or $\sum_{i=1}^{N} g(u^*(x;i)) \geq \sum_{i=1}^{N} g(u^*(y;i))$. Because *g*(.) is invariant to a ratio transformation and **U** is unique up to a ratio transformation, this is in turn the same as the lifetime prioritarian rule, $\sum_{i=1}^{N} g(u(x;i)) \geq \sum_{i=1}^{N} g(u(y;i))$ for all *u*(.) in **U**.

[39] Lifetime prioritarianism, here, means the use of a continuous prioritarian SWF. That approach also, of course, satisfies the continuity axiom – more precisely, a continuity axiom in terms of lifetime utilities. Although I embrace that axiom, it seems to me to flow less directly from the basic idea of ranking outcomes as a function of individuals' claims, plus the valencing of those claims in terms of lifetime well-being, than the four substantive axioms mentioned in the text. (Lifetime leximin or the lifetime application of the prioritarian SWF with an absolute threshold are, it seems to me, consistent with that basic idea.)

I am not sure whether it is possible for a non-lifetime continuous prioritarian approach to satisfy the four substantive axioms but violate continuity in terms of lifetime utilities. Assuming that this *is* possible, I would level the same critique against this approach as I would against a non-lifetime approach which is extensionally equivalent to lifetime continuous prioritarianism (see below), namely that this is a problematic way to think about the ranking of outcomes.

says, if each individual's lifetime well-being in *x* is greater than or equal to her lifetime well-being in *y*, with lifetime well-being strictly greater for at least one individual, then *x* is a morally better outcome. Finally, the Pigou-Dalton principle, discussed in Chapter 4, was framed in terms of transfers of lifetime well-being, from an individual at a higher level of lifetime well-being to an individual at a lower level of lifetime well-being. And the separability-across-persons axiom, also discussed in Chapter 4, involved individuals who were "unaffected" in the sense of having equal levels of lifetime well-being in the outcomes being ranked.

The reader should be reminder that these principles are *relative* to an account of lifetime well-being. Which particular ranking of outcomes is required by the lifetime Pareto, Pigou-Dalton, and separability-across-persons axioms depends on which account of well-being is being used.

The possibility of a non-lifetime approach which is extensionally equivalent to lifetime prioritarianism shows that – for any given account of lifetime well-being -- sublifetime approaches do *not* necessarily violate the lifetime Pareto, Pigou-Dalton, and separability-across-persons axioms. However, there obviously can be non-lifetime approaches which *do* violate one or more of these axioms,[40] as the following tables illustrate.

These tables illustrate the potential conflict between simple sublifetime prioritarianism and lifetime Pareto indifference, Pareto superiority, Pigou-Dalton, and separability-across-persons.[41] (Similar examples could be constructed for attribute based prioritarianism or other variants of sublifetime prioritarianism.). Assume that the prioritarian SWF being used by the sublifetime approach is some Atkinsonian SWF. In all cases, the numbers in the tables are individuals' sublifetime utilities. Lifetime utility, assumed to be unique up to a ratio transformation, is the sum of sublifetime utility. These same sublifetime utility values are used by the sublifetime approach.

Outcome *x* Outcome *y*

[40] There can be non-lifetime approaches which satisfy some but not all of these axioms. For example, assume that each *u*(.) in **U** takes the form of summing sublifetime utilities. In other words, $u(x;i)=\sum_{t=1}^{T} w(\mathbf{a}_i^t(x),\mathbf{a}_{imp}^t(x))$. Pick an increasing, concave function *h*(.). For each such *u*(.), specify a corresponding function $v(\mathbf{a}_i^t,\mathbf{a}_{imp}^t)=h^{-1}(w(\mathbf{a}_i^t,\mathbf{a}_{imp}^t))$. Note that $u(x;i)=\sum_{t=1}^{T} h(v(\mathbf{a}_i^t,\mathbf{a}_{imp}^t))$. Consider simple sublifetime prioritarianism, where each segment is a single period, and (*x*; *i*; *t*) denotes period *t* of individual *i*'s life in outcome *x*. Set $v(x;i;t)=v(\mathbf{a}_i^t,\mathbf{a}_{imp}^t)$, with *v*(.) as just defined. Designate by **V** the set of all such *v*(.), each corresponding to one *u*(.). Consider the ordering of outcomes achieved by the sublifetime rule, *x* is at least as good as *y* iff, for all *v*(.) in **V**, $\sum_{i=1}^{N}\sum_{t=1}^{T} h(v(x;i;t)) \geq \sum_{i=1}^{N}\sum_{t=1}^{T} h(v(y;i;t))$. It is not hard to show that this satisfies lifetime Pareto indifference, Pareto superiority, and separability-across-persons but does not satisfy lifetime Pigou-Dalton.

[41] Again, simple sublifetime prioritarianism maps each outcome onto a vector or vectors of *TN* sublifetime utilities, one for each of the *N* individuals in each of the *T* periods, and ranks vectors by summing an increasing, concave function of these sublifetime utilities.

	Period 1	Period 2	*Lifetime utility*	Period 1	Period 2	*Lifetime utility*
Jim	90*c*	10*c*	100 *c*	50*c*	50*c*	100*c*
Sue	10*c*	90*c*	100*c*	50*c*	50*c*	100*c*

This illustrates the conflict with *lifetime Pareto indifference*. That axiom requires that the outcomes be ranked as equally good, but the simple sublifetime approach using any Atkinsonian SWF (one with any value of $\gamma > 0$) will count *y* as the better outcome, because it equalizes sublifetime utilities.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	*Lifetime utility*	Period 1	Period 2	*Lifetime utility*
Jim	90*c*	10*c*	100 *c*	(50-ε)*c*	(50-ε)*c*	(100-2ε)*c*
Sue	10*c*	90*c*	100*c*	(50-ε)*c*	(50-ε)*c*	(100-2ε)*c*

The above table illustrates the conflict with *lifetime Pareto superiority*. The sublifetime approach, using any Atkinsonian SWF, will count *y* as better than *x* for ε sufficiently small, because *y* equalizes sublifetime utilities and loses only a little (ε). But lifetime Pareto superiority requires that *x* be preferred.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	*Lifetime utility*	Period 1	Period 2	*Lifetime utility*
Jim	50*c*	90*c*	140 *c*	30*c*	90*c*	120*c*
Sue	50*c*	10*c*	60*c*	70*c*	10*c*	80*c*

The above table illustrates the conflict with *lifetime Pigou-Dalton*. The sublifetime approach, using any Atkinsonian SWF, will count outcome *x* as better than *y*, because the individuals' sublifetime utilities are equalized in period 1. But the lifetime Pigou-Dalton principle requires that outcome *y* be ranked better.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	*Lifetime utility*	Period 1	Period 2	*Lifetime utility*
Jim	50*c*	50*c*	100 *c*	50*c*	60*c*	110*c*
Sue	50*c*	50*c*	100*c*	50*c*	40*c*	90*c*
Fred	90*c*	10*c*	100*c*	50*c*	50*c*	100*c*

Sum of square root of sublifetime utilities: 40.93338 42.35479

	Outcome *x**			Outcome *y**		
	Period 1	Period 2	*Lifetime utility*	Period 1	Period 2	*Lifetime utility*
Jim	50*c*	50*c*	100 *c*	50*c*	60*c*	110*c*
Sue	50*c*	50*c*	100*c*	50*c*	40*c*	90*c*
Fred	6*c*	6*c*	6*c*	6*c*	6*c*	6*c*

Sum of square root of sublifetime utilities: 33.85815 33.11164

The above two tables illustrate the conflict with *lifetime separability across persons*. Note that Fred has the same lifetime utility in *x* as in *y*, and in *x** as in *y**. Moreover, Jim has the same lifetime utility in *x* as in *x**, and in *y* as in *y**. The same is true of Sue. Lifetime separability-across-persons thus requires that *x* be ranked at least as good as *y* iff *x** is ranked at least as good as *y**. Note, however, that the sublifetime approach using a square root function ranks *y* over *x* but *x** over *y**.

Now for the basic case in favor of lifetime prioritarianism. Take any account *W* – whatever it may be – which the reader believes to be the most attractive account of lifetime well-being. Now consider, first, a non-lifetime prioritarian approach which violates one or more of the axioms of lifetime Pareto indifference, lifetime Pareto superiority, lifetime Pigou-Dalton, or lifetime separability-across-persons, in terms of *W*.

Assume, for example, that the non-lifetime approach violates lifetime Pareto indifference. Then there will be some outcome set, where *x* is (lifetime) Pareto indifferent to outcome *y*, and yet the non-lifetime approach prefers *x* to *y*. Such a preference flies in the face of the idea of ranking outcomes by comparing individuals' claims across outcomes, plus the continuity of personal identity across a human lifetime. Take any arbitrary member *i* in the population. That *person* is equally well off in both outcomes. It may well be the case that individual *i*'s sublifetime well-being during some temporal segment in outcome *x* is greater than his sublifetime well-being during some temporal segment in outcome *y*; but, if so, he has been *compensated* in *y* for that sublifetime divergence, in the sense that his attributes in *y* over his entire lifetime are such that his *lifetime well-being* is exactly the same in both outcomes. (For example, Jim's sublifetime well-being in period 1 may be greater in *x* than in *y*; but if *x* and *y* are lifetime Pareto indifferent, then there is something about Jim's life in *y* which *equilibrates* this period 1 difference in his sublifetime well-being. Perhaps his sublifetime well-being is greater in *y* in period 2.) Similarly, it may well be the case that the level of some attribute of individual *i* is greater in outcome *x* than in *y*; but, if so, he has been *compensated* in *y* for that attribute divergence, in the sense that his attributes in *y* over his entire lifetime are such that his *lifetime well-being* is exactly the same in both outcomes. (For example, the quality of Jim's health at one time may be greater in outcome *x* than in outcome *y*; but perhaps his consumption at that time is greater in outcome *y*; or perhaps the quality of Jim's health at some other time is greater in outcome *y* than in outcome *x*.)

Because individual *i* is equally well off in both outcomes, in what sense does *he* have a claim in favor of either outcome? How can *he* complain if one outcome rather than the other obtains? Because this is true for all individuals – because the outcomes are lifetime Pareto indifferent – no one can complain if one outcome rather than the other obtains. An attractive version of prioritarianism should never rank one outcome over the other in such a case.

What work, exactly, is the continuity of personal identity over a human lifetime doing in this argument? The work is this. Each person in the population is a distinct locus of moral concern. This is the Rawlsian idea of the "separateness of persons." The concept of ranking outcomes by balancing individuals' claims across outcomes tries to give fuller content to this idea. It assigns each individual a claim in favor of one or another outcome. Most naturally, this should be a single claim, in the sense that for any pair of outcomes, an individual either has a claim in favor of *x*, a claim in favor of *y*, no claim either way, or an incomparable claim. The idea of the "separateness of persons," intuitively, is that moral deliberation should be sensitive both to the fact that (1) each person is an entity which is distinct from every other person and to the fact that (2) each person is a locus of interests that are *integrable*, in the sense that we can arrive at a single, all-things-considered ranking of whether outcomes and actions are better or worse *for him*. We *could* assign each person a multiplicity of claims between outcomes, each corresponding to a different temporal segment, or a different attribute; but to do so would be in serious tension with this idea of integrability. Wouldn't ranking outcomes by assigning each individual multiple claims, corresponding to different temporal segments or attributes, be to see each such *part* as a locus of moral concern, rather than each person?

In suggesting that this idea of integrability is a natural part of the claim-across-outcome view, I am in part appealing to intuitions. But for many readers, I hope, those intuitions will be no less powerful than the intuitions undergirding the claim-across-outcome view itself or, more fundamentally, the idea that morality should be sensitive to the separateness of persons. Each person is seen to be such that *he* has his own, unitary, perspective on how outcomes should be ranked; morality, in turn, is supposed to accommodate these different perspectives in an impartial manner.

Now, for the claim-across-outcome view to mesh with welfarism, an individual's (single) claim must be valenced in terms of his well-being. And because each individual person corresponds to a single human being, from human birth to human death, it seems natural to valence that claim in terms of the individual's lifetime well-being. Imagine, instead, that we valence his claim in terms of his sublifetime well-being during some stipulated segment. (For example, we say that each individual has a claim in favor of *x* over *y* if his sublifetime well-being during the first period is greater in *x* than *y*.). Or imagine that we valence his claim in terms of the level of some particular attribute. (For example, we say that each individual has a claim in favor of *x* over *y* if his health in period 2 is greater in *x* than *y*.) Wouldn't doing so be arbitrary? What would justify our choosing one rather than another temporal segment, or one rather than another attribute, as the "currency" for individual claims? Moreover, why should the direction and strength of an individual's moral claim be determined by the properties of one of his temporal parts, or by one of his attributes, to the exclusion of others?

In short, the claim-across-outcome view, plus the continuity of personal identity, suggest that an individual should have a unitary claim regarding any pair of outcomes, valenced in terms of his lifetime well-being. This implies that the ranking of outcomes should satisfy lifetime Pareto indifference. A very similar line of argument shows why the ranking of outcomes should satisfy lifetime Pareto superiority, lifetime Pigou-Dalton, and lifetime separability across persons.

The argument, thus far, has focused on non-lifetime approaches that violate at least one of the key axioms of lifetime Pareto indifference, lifetime Pareto superiority, lifetime Pigou-Dalton, and lifetime separability-across-persons. What about a non-lifetime approach that *satisfies* all these axioms – for example, a non-lifetime approach which is extensionally equivalent to lifetime prioritarianism?

The two key premises I have been invoking (the claim-across-outcome view, combined with the continuity of personal identity across a human lifetime), suggest that a non-lifetime approach provides a roundabout and unhelpful way to *think* about the moral ranking of outcomes, even if the approach does satisfy lifetime Pareto indifference, Pareto superiority, Pigou Dalton, and separability across persons. If these two key premises are correct, then -- in order to *justify* any such non-lifetime approach --- we would need to appeal to the account W of lifetime well-being and a sense of how a given change in lifetime well-being translates into a moral claim. We would say: "In ranking any pair of outcomes x and *y*, this non-lifetime approach appropriately balances the gains in lifetime well-being of the persons who are better off in *x*, and the losses in lifetime well-being of the persons who are worse off in *x*." But why not rank outcomes using a formula that more directly and immediately reflects this justification – namely, the lifetime formula, which sees the ranking of outcome as being a direct function of changes in individuals' lifetime utilities?

Thus the basic case for lifetime prioritarianism. The line of argument I have presented is very much in accord with the work of Thomas Nagel, from whom I derive the idea of ranking outcomes as a function of individual claims. Nagel writes: "[T]he 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 …." And he buttresses the lifetime perspective with reference to the possibility of compensation:

> By itself the possibility of intrapersonal compensation neither supports nor undermines egalitarian theories. It implies only that *if* an egalitarian theory is accepted, it should apply only across lives rather than within them. It is a reason for taking individual human lives, rather than individual experiences, as the units over which any distributive principle operates.

More generally, lifetime framings of the Pareto principles and lifetime approaches to employing an SWF are the standard approaches in welfare economics. [42] Inequality metrics are often used in a sublifetime manner (for example, measuring the inequality of annual income); but a substantial body of work does attempt to estimate the inequality of lifetime income or other characteristics of whole human lifetimes.

[42] In the case where outcomes are specified in terms of multiple attributes, economists invariably understand the Pareto principle in terms of an individual's well-being as a function of her attributes, not on an attribute-by-attribute basis. And (except for the small literature on multi attribute equality) the SWF literature handles such outcomes by applying the SWF to individuals' utilities as a function of their attributes, or perhaps to a single normalized attribute, and does not use the vector of attributes as the input for the SWF. Similarly, in the case of multi-period outcomes, the Pareto principle is typically understood in terms of an individual's lifetime well-being; and the SWF is typically (although not invariably) applied to vectors of individuals' lifetime utilities, not their sublifetime utilities.

Finally, philosophers of distributive justice quite often adopt the view that its concern is how individuals fare across their entire lifetimes (whether individuals' lifetime attainments are measured in well-being or some other currency). To give two prominent examples: John Rawls argues that distributive justice requires a fair allocation of lifetime shares of primary goods, and Ronald Dworkin argues that it requires each person to receive a fair lifetime share of resources.

Does the Deflationary Cast of Personal Identity Argue against Lifetime Prioritarianism?

In *Reasons and Persons*, Derek Parfit suggests that his account of personal identity undermines accounts of distributive justice that focus upon fair distribution between persons. In particular, he argues, a proper understanding of personal identity changes the "scope" of distributive justice, so that it now focuses on fair distribution between temporal segments of persons. At the same time, his account of personal identity reduces the "weight" of distributive justice; morality becomes less concerned with fair distribution and moves closer to utilitarianism.

I have already described Parfit's account of personhood and personal identity, and taken pains to explain that it is *consistent* with the premise that personal identity continues over a normal human lifetime. And Parfit admits as much. So why would the account change the scope and/or weight of distributive justice?

Parfit's arguments concerning distributive justice rest on the *deflationary* character of his account of personhood and personal identity. He characterizes his account as a "Reductionist" account – a term he uses with great frequency. Unlike the Cartesian ego/soul view, Parfit's account says that personhood, and personal identity, is reducible to psychological and physical states and connections. "We are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events. Our existence just involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events." There is no "deep further fact" of personhood and personal identity above and beyond these psychological and physical facts.

Parfit's discussion of the relation between "Reductionism" and distributive justice is dense and somewhat meandering. Rather than recapitulating his discussion and responding point by point, I will structure the analysis by considering, first, whether Parfit's account of personhood and personal identity argues for a shift from lifetime prioritarianism to *momentary* prioritarianism; second, whether it argues for a shift to a different kind of sublifetime prioritarianism; and, third, whether it argues for a shift to utilitarianism. I will argue that none of these three shifts is warranted.

(1) *Revising the Across-Outcome Conception of Fairness: Seeing Momentary Time Slices s the Units of Moral Concern.*

One possibility is that Reductionism about personal identity might justify a revisionary version of the claim-across-outcome view: specifically, one that makes each human person at each moment, rather than each whole human person, the locus of moral concern and the holder of a claim for or against outcomes. This would lead us to a particular kind of sublifetime prioritarianism: namely, *momentary* sublifetime prioritarianism, where the arguments for the SWF are not lifetime utilities, but *momentary* utilities, measuring the momentary well-being of each person in the population at each moment. (Momentary sublifetime prioritarianism requires that outcomes be divided into a series of moments, rather than longer periods; what I earlier described as the "simple" sublifetime rule is then employed to rank these outcomes.)

Indeed, Parfit suggests that Reductionism may well have exactly this upshot. Reductionism may make it impossible for a person who is made worse off at one moment to be compensated by benefits at later moments, because any such compensation involves a "deep further fact" of personal identity denied by Reductionism. If so, the upshot will be momentary prioritarianism. Parfit writes:

> In becoming Reductionists, we cease to believe that personal identity involves the deep further fact. [A plausible] argument claims that what compensation presupposes is not personal identity on any view, but personal identity on the Non-Reductionist View. Compensation presupposes the deep further fact. Psychological continuity, in the absence of this fact, cannot make possible compensation over time.
>
> If this is not possible, what will our distributive principles tell us to do? They will roughly coincide with *Negative Utilitarianism*: the view which gives priority to the relief of suffering. Nagel talks of the *unit* over which a distributive principle operates. If this is the whole of a person's life, as is assumed by Rawls and many others, a Principle of Equality will tell us to try to help those people who are worst off. If the unit is the state of any person at a particular time, a Principle of Equality will tell us to try to make better, not the lives of the people who are worst off, but the worst states that people are in.

In short, if "only the deep further fact [of personal identity that would be provided by a soul/Cartesian ego] makes possible compensation over different parts of a life," then absent such a fact "the units [for distributive justice] shrink to peoples' states at particular times."

For short, I will refer to Parfit's claim that Reductionism renders compensation over time impossible as the "no-compensation" claim. In *Reasons and Persons*, Parfit tentatively defends the no-compensation claim and the suggestion that this claim, in turn, argues for momentary prioritarianism. In a subsequent set of comments, responding to critics of the book, he makes this argument more wholeheartedly.

> If there cannot be compensation over time, as my argument suggests, we should change the scope of distributive principles. According to one such principle, we should give priority to helping the people who are worse off. On my argument, we should give priority not to those who are worse off in their lives as a whole, but to those who are worse off at particular times.

I suggested earlier that there is a conceptual link between compensation and lifetime well-being. Consider a case in which individual *i* is worse off in outcome *x* than *y*, with respect to some attribute or with respect to sublifetime well-being at some time. Individual *i* is *compensated* for this shortfall if his lifetime well-being in *x* is greater than or equal to that in *y*.

A surplus in some other attribute(s), or in sublifetime well-being at some other time(s), is sufficient to counterbalance the shortfall.

Parfit's no-compensation claim, thus, is that Reductionism has the following upshot: If Jim at moment *t* is worse off in outcome *x* than outcome *y*, then his lifetime well-being in *x* is not equal to (or greater than) his lifetime well-being in *y*, even if Jim at other moments has greater momentary well-being in *x*. The reader might initially wonder how this is even logically coherent. Imagine that Jim at moment *t* is worse off in *x* than *y*, and at moment *t** is worse off in *y* than *x*. If Reductionism renders compensation over time impossible, and thus means that Jim's lifetime well-being in *x* is not greater than or equal to his lifetime well-being in *y*, and that Jim's lifetime well-being in *y* is not greater than or equal to his lifetime well-being in *x*, aren't we left with a logical contradiction? The answer is no – once we allow for the possibility that the well-being ranking of life-histories might be a *quasiordering* rather than a complete ordering, and thus that two life histories might be *incomparable* with respect to well-being.

In short, I interpret Parfit's no-compensation claim to be the following: *By virtue of Reductionism, if there is some moment when individual i is worse off in outcome x than y, and some other moment when individual i is worse off in outcome y than x, then the life histories (x; i) and (y; i) are incomparable.* The no-compensation claim, if true, means that Reductionism injects massive incomparability into the ranking of life-histories. And massive incomparability in the well-being ranking of life-histories would, in turn, undermine lifetime prioritarianism.

I of course take the position that the ranking of life-histories is a quasi-ordering, and thus allow for *some* incomparability in ranking life-histories. But why on earth believe that the no-compensation claim is true? Parfit suggests that, if we are Non-Reductionists, we do not believe that a person can be compensated by benefits to some other person psychologically and physically linked to him. This suggestion involves a kind of atomism about the sources of well-being; but if the premise of atomism is granted, Parfit's suggestion is correct. He illustrates it by discussing a case in which a person's brain is divided in half and put into two bodies, yielding two persons, Righty and Lefty. Let us further suppose (which seems plausible) that, on a Non-Reductionist view, the original person will continue as Righty or Lefty, by virtue of some deep further fact above and beyond psychological and physical facts.

> Assume that we believe the Non-Reductionist View, and that we suppose that I shall be Righty. Before the division, I had more than my fair share of many resources, living for many years in luxury. After the division, I and Lefty will each get less than a fair share. This is claimed to be justified, in my case, because it will have the result that in my life as a whole I shall receive a fair share. My lesser share now was fully compensated in advance by my greater share before the division.
>
> Could we plausibly claim the same about Lefty? Does psychological compensation make possible compensation, even in the absence of personal identity? It is defensible to answer:
>
> > No. Lefty never enjoyed a larger share. *He* did not enjoy these years of luxury. It is irrelevant that he can quasi-remember *your* enjoyment of this luxury. It is irrelevant that he is physically and

> > psychologically continuous with someone who had more than his fair share, at a time when Lefty did not exist. It would now be unfair to give Lefty less than a fair share. In the absence of personal identity, psychological continuity cannot make compensation possible.

> Suppose next that we come to believe the Reductionist View. We had claimed, defensibly, that only the deep further fact makes compensation possible. We had claimed that, as the case of Lefty shows, physical and psychological continuity cannot by themselves make compensation possible. We now believe that there is no deep further fact, and that personal identity just consists in these two kinds of continuity. Since we could defensibly claim that only this further fact makes compensation possible, and there is no such fact, we can defensibly conclude that there cannot be compensation over time. We can claim that a benefit at one time cannot provide compensation for a burden at another time, even when both come within the same life. There can only be simultaneous compensation, as when the pain of exposing my face to a freezing wind is fully compensated by the sight of the sublime view from the mountain I have climbed.

But this argument for the no-compensation claim involves a huge non sequitur. What it says is the following: (1) Given Non-Reductionism and atomism: If Jim's attributes in outcomes *x* and *y* at moment *t* are such that he is worse off in outcome *x* than *y* at that moment, then the fact that some person *other than Jim* is better off in outcome *x* than *y* at other moments does not suffice to make *Jim's* lifetime well-being in *x* greater than or equal to *y*, even if this other person happens to have psychological or physical links to Jim. *Therefore*: (2) Given *Reductionism* and atomism: If Jim's attributes in outcomes *x* and *y* at moment *t* are such that he is worse off in outcome *x* than y at that moment, then the fact that *Jim* is better off at other moments does not suffice to make Jim's lifetime well-being in *x* greater than or equal to *y*, since Jim remains the same person at different moments only in virtue of psychological and/or physical links.

The leap from proposition (1), which is true, to proposition (2), the no-compensation claim, is a massive non sequitur, because what it ignores is the fact that *our understanding of the determinants of lifetime well-being would shift along with our understanding of personal identity*. If we are Non-Reductionists, a given person (Jim) consists in a series of momentary time slices of a human being, linked by a deep further fact of personal identity. Given this view (plus atomism), Jim's lifetime well-being in some outcome is a function of the momentary attributes of the set of human time slices that exist at different moments but are linked by this deep further fact, and that, collectively, we refer to as "Jim." If we are Reductionists, a given person consists in a series of momentary time slices, knitted by various physical and/or psychological connections. Our understanding of the sources of lifetime well-being shifts accordingly. Given *this* view (plus atomism), Jim's lifetime well-being in some outcome is a function of the momentary attributes of a set of human time slices that exist at different moments, *and that have the right sort of physical and/or psychological links,* and that, collectively, we refer to as "Jim."

It is a truism that the lifetime well-being of a particular person is largely determined by the attributes, at different times, of *that particular person*. (Atomism replaces "largely" with "exclusively.") Non-Reductionists and Reductionists disagree, however, about the criteria of individuation. They disagree about when human beings at various times are parts of *the very same person*. Parfit's argument for the no-compensation claim seems to imagine that the Reductionist, in crafting an understanding of a person's lifetime well-being, will be a sort of non-Reductionist *manqué*. The Reductionist will try to use non-Reductionist individuation criteria in picking out the temporal attributes that are relevant to the lifetime well-being of any particular person and (because there are no such criteria) will conclude that compensation over time is impossible. But it is absurd to suppose that the Reductionist will think about lifetime well-being this way. Rather, she will say: temporal slices of a human person at various times are parts of the same particular person if they bear stipulated physical and/or psychological links to each other (not the deep further fact); and it is the attributes of the thus-linked time slices that determine the lifetime well-being of that particular person.

In short, the no-compensation claim is false, and thus *this* argument for momentary prioritarianism is a non-starter. Perhaps a better argument available? One of the main themes in *Reasons and Persons* is that Reductionism about personal identity undercuts the normative relevance of personal identity as such.

> Personal identity is not what matters. What fundamentally matters is Relation R [i.e., psychological connectedness and/or psychological continuity], with any cause. This relation is what matters even when, as in a case where one person is R-related to two other people, Relation R does not provide personal identity.

Consider again the case in which Jim's brain is divided into two halves, Lefty and Righty. Because personal identity, on Parfit's account, requires *non-branching* psychological continuity (so as to avoid intransitive identity), that account says that Jim ceases to exist when his brain is divided; neither Lefty nor Righty are the same person as Jim. But Jim would rationally care almost as much about Lefty and Righty as if Jim *had* survived. Jim is rationally concerned about the attributes of human beings with whom he has strong psychological links, *even if Jim is not the same person as those human beings*. Thus it is Jim's psychological links to other human beings, not whether he is the very same person, that should determine his rational preferences and choice.

This observation, in turn, might be seen to undercut the extended preference account of lifetime well-being that I have offered. Imagine that the *N* members of the population are normal human beings, without division. Assume that Parfit's account of personal identity is true. Jim is one of the normal human beings, with attributes at various times, from his birth until his death. Jim at age 8 is the very same person as Jim at 90 because the two humans are psychologically continuous, with the right sort of cause. Consider, now, Jim at a given time (say, when he is 50), functioning as "spectator." I have argued that spectators must be temporally neutral. But, if "personal identity is not what matters," wouldn't Jim be temporally biased? Although Jim is psychologically *continuous* with himself at all times, Jim at age 50 has stronger psychological *connections* with more proximate time slices of himself. He shares more memories, desires, character traits, etc. with these time slices. So wouldn't Jim at 50, functioning as spectator, and ranking his own life-histories, care more about his attributes at age 45 or age 55, as opposed to his attributes at ages 80 or 8? Wouldn't a given spectator's ranking of his own life histories incorporate a discount factor for degrees of psychological connectedness? And if this is true for a spectator's ranking of his own life histories, wouldn't this also be true for his ranking of other life-histories? Moreover, as I explained earlier, temporal bias on the part of spectators could well induce large-scale incomparability in the ranking of life-histories.

Call this the "temporal bias" argument for momentary prioritarianism: Reductionism about personal identity engenders temporal bias on the part of spectators, which induces large-scale incomparability in the ranking of life-histories, undermining lifetime prioritarianism.

This argument, too, is unpersuasive. As I explained earlier, the extended preference account does not say that temporal neutrality is rationally required in all contexts. In ordinary life, it may well be rationally permissible for individuals to have various sorts of temporal biases – and a deflationary view of personal identity may well reinforce the case for the permissibility of some such biases. My claim is only that preferences in the particular idealized choice situation that I have constructed, for purposes of constructing lifetime well-being, must be temporally neutral. I don't see why Reductionism about personal identity would make it impossible or irrational for spectators to be temporally neutral. Here's an analogy: even though each person is distinct from every other (on any account of personal identity, reductionist or not), it is presumably possible for persons to be impartial between their own interests and those of others; and a normative account of some sort might identify choice scenarios, actual or hypothetical, in which such impartiality is required Similarly, even though each person at a given time is more closely psychologically connected to some of his past and future time slices than others, it is presumably possible for persons to be impartial between all of their time slices; and a normative account of some sort (in this case an account of lifetime well-being) might identify choice scenarios, actual or hypothetical, in which such temporal neutrality is required.

There is a third potential argument for momentary prioritarianism, which is suggested by some of Parfit's remarks. [43] Call this the "unity" argument. Ceteris paribus, we should specify the claim-across-outcome view so that the holders of claims are *internally unified*. In defending the view that *L* is the proper locus of moral concern (where *L* could be a person, a time-slice of a person, or something else), we appeal both to the fact that each *L* is distinct from every other, *and* to the fact that each *L*'s concerns are *integrable* – that each *L* is a *single* entity, with a single set of interests and concerns. Thus considerations of internal unity are an important factor in adjudicating between different candidates for *L*.

[43] His discussion at pp. 332-34, 336-39, might be understood to suggest that Reductionism changes the scope of distributive justice because the unity between different parts of a human life becomes a matter of degree, with shorter time slices being more unified -- rather than unity being an all-or-nothing matter grounded in a "deep further fact" of identity.

Momentary time slices of persons are much more tightly unified, psychologically and physically, than whole persons. In a normal human life, there *is* an overlapping chain of direct psychological connections between any two time-slices; but, still, a human person at one time may have a different (perhaps quite different) set of beliefs, preferences, affects, and so forth than the very same human person at another time. By contrast, because momentary time slices do not endure over time, a time slice's mental (or physical) attributes do not change.

While the "unity" argument does indeed provide a ceteris paribus argument in favor of momentary prioritarianism, there are decisive countervailing factors against seeing momentary time slices rather than whole persons as the loci of moral concern and the holders of claims. A key basis for crafting moral norms in light of the separateness of *persons* is that the resultant norms can be seen as norms *for* the community of persons: the norms can now be justified *to* each person, who can be expected to act in compliance with the norms. No such basis can be given for taking *L* to be momentary time-slices of persons. Time slices are not capable of temporally extended normative deliberation – it is not possible to justify a norm *to* a time-slice – nor are they capable of acting (or, at least, performing more than a single action). Moreover – and perhaps more dramatically -- important aspects of human well-being are ascribable only to whole human persons, or temporally extended segments of persons, rather than to momentary time slices. This includes items such as accomplishment, relationships with friends or family, civic life, or the pursuit of knowledge.[44] Momentary prioritarianism would mean that these aspects of human well-being do not figure in moral reasoning at all. As David Brink observes:

> It is difficult to regard person-slices as agents. … [To begin] it is not clear that person-slices do have interests. Whether the entity me-now can have interests depends upon which theory of welfare is correct.
>
> If pleasure is a simple, qualitative sensation or mental state and a hedonistic theory of welfare is correct, then perhaps me-now has interests. Person-slices may contain qualitative mental states, such as pleasures …..
>
> [However] hedonism seems an implausible theory of welfare, because a large part of a person's good seems to consist in his *being a certain sort of person* – that is, a person with a certain sort of character who exercises certain capacities and develops certain kinds of personal and social relationships. … [T]his implies that it is temporally extended beings… rather than person-slices who are the bearers of interest.

[44] For exactly these reasons, it is also hard to see what would justify a moral view that sees momentary time-slices of human beings as the units of moral concern, rather than a moral view that seems momentary time-slices of human beings or non-human animals as the units of moral concern. While there are major normative differences between human *persons* and non-human animals, these differences evaporate when we compare time-slices of humans with time-slices of non-human animals.

The proponent of momentary prioritarianism could bite this particular bullet, arguing for an *inclusive* momentary prioritarianism, which takes the momentary utilities of animals as well as human beings as the inputs for the prioritarian formula. However, because momentary prioritarianism in *either* its inclusive or person-centered form (1) is not justifiable to a community of entities that can engage in action and normative reasoning, and (2) ends up giving zero moral weight to long-term aspects of human well-being, I believe that either sort of momentary prioritarianism is on balance less attractive than lifetime person-centered prioritarianism.

> Not only is it doubtful whether person-slices have interests, it is also questionable whether having interests is sufficient for having reasons for action. A person-slice will not persist long enough to perform actions or receive the benefits of actions. If so, then person-slices cannot have reasons for action even if it is possible for them to have interests.

A final point is that momentary prioritarianism is a poor candidate to be a moral choice evaluation procedure, because it demands a high degree of specificity in the temporal description of outcomes. Consequentialist choice-evaluation procedures control decision costs by permitting outcomes to be described in simplified ways. One way to do so is to characterize outcomes as having a relatively small number of periods. But momentary prioritarianism demands that outcomes be divided into moments, each with its own array of individual attributes.

(2) *Shifting from Persons to Stages as the Loci of Moral Concern*

Consider the normal human being, Jim, at ages 5,7, 15, 18, 35, 40, 75, and 85. Jim at each age is the very same person as Jim at every other age, because all the ages are psychologically *continuous* with each other. But, colloquially, we might say that the first two ages are parts of Jim's childhood "self"; the second two, parts of his adolescent "self"; the third two, parts of his middle-aged "self"; the last two, parts of his older "self." We might sharpen this analysis by observing that there are tight psychological connections between the first two ages in this series, the second two, the third two, and the fourth two, but not between any other two ages (e.g., between Jim at 7 and 35, Jim at 75 and 35, etc). And we might make "selves," defined in terms of such connections, rather than persons, the units of moral concerns.

Because the term "self" is also used to refer to whole persons, I will instead refer to a psychologically unified, temporally extended portion of a person's life as a "person-stage." Parfit, himself, does not clearly argue in favor of making person-stages a morally or normatively relevant concept - although some of his remarks seem to point in this direction. Other scholars have claimed more decisively that Parfit's account of personal identity warrants a reorientation of morality or other norms oriented around person-stages. It is therefore worth considering this kind of revision of the claim-across-outcome approach – one that would argue for a sublifetime prioritarianism in which the arguments for the SWF are the utilities of person-stages.

At the outset, I should reiterate that the question on the table in this chapter (and book) is how to specify welfarism for the case of a population of normal human beings. Consider, by contrast, the case of Sue, who at age 45 suffers a traumatic episode, causing her to forget all of what went before, and altering her personality quite radically. Parfit's account of personal identity says that Sue up to age 45 and Sue after 45 are two different persons, corresponding to different non-overlapping segments of the life of the human being, Sue. This is, indeed, fairly plausible – and it is therefore also plausible to see Sue-before-45 and Sue-after-45 as each holding a claim across outcomes, rather than assigning a single claim to Sue.

But how the prioritarian should handle Sue is not a topic I will further pursue. The question on the table is *not* whether a whole person might correspond to some fraction of a whole human life, if the human suffers an abnormal, radical, psychological rupture. The question, rather, is whether we should endorse person-stage prioritarianism, understanding that the "stages" are merely fractions of *persons*. Each time-slice of a normal human life is sufficiently directly connected to adjacent time-slices, so that (on Parfit's account) each human remains one and the same person from birth to death. In this case, a person-stage corresponds to a fraction of a human being's life *and thus a fraction of that person's life*.

With this clarification behind us, how exactly should person-stage prioritarianism be fleshed out? One possibility is a *consecutive* approach. This would divide a person's life into a series of consecutive, non-overlapping person-stages. If an outcome set has *T* periods, a given human person, *i*, would have $s_i(x)$ stages in outcome *x*: one stage in periods 1 through $i_2(x)$-1, where "$i_2(x)$" denotes the first period of stage 2 in outcome *x*; a second stage in periods $i_2(x)$ through period $i_3(x)$-1; a third stage in periods $i_3(x)$ through $i_4(x)$ -1;…; and a final stage in periods $i_{s_i(x)}(x)-1$ through period *T*.

A serious difficulty with this approach is characterizing the degree of attenuation of psychological connectedness that yields a new stage without yielding a new person. Individual *i* is one stage during periods 1 through $i_2(x)$-1. Something happens, psychologically, at that point to produce a new stage, who comes into being in period $i_2(x)$ and exists through period $i_3(x)$-1. But whatever happens prior to the beginning of period $i_2(x)$ is not so dramatic to sever personhood; individual *i* remains one and the same person from the first through the last period. Is there really some intermediate range of change in memories, desires, character traits, and so forth, that births a stage without birthing a person? As far as I'm aware, no proponent of the stage approach has explained what such a moderately transformative psychological alteration would consist in.

A different approach avoids this difficulty. (Call this "overlapping stage" prioritarianism.) On Parfit's account, of course, a certain degree of connectedness is necessary (via an overlapping chain of connections) to create continuity and, thus, personal identity. Parfit, again, terms this degree of connectedness "strong connectedness." So consider using this concept to define a person-stage, as follows: For any person *i*, in a given outcome *x*: if there is some group of consecutive periods, such that *i* in each period in the group is strongly connected with *i* in every other period, and this group is "maximal" (adding an additional period at either end would mean that *i* in each period is not connected with *i* in every other period), then this group of periods is one of *i*'s stages in *x*.

Person-stages, in this sense, can overlap. For example, Jim can have 1 stage that runs from Jim at birth through Jim at age 10; a second stage that runs from Jim at age 3 through Jim at age 15, a third stage that runs from Jim at age 6 through Jim at age 15; and so forth. In this case, Jim at age 7 would be part of three stages.

Overlapping stage prioritarianism uses a set **V** of sublifetime utility functions, where each *v*(.) maps an outcome *x* onto a grand vector, including an entry for the sublifetime utility of each stage of each person in *x*. Outcomes are then ranked as a function of these vectors.

The "unity" argument I presented earlier can be used to make a case for "overlapping stage" prioritarianism, as against lifetime prioritarianism: stages are more tightly internally unified than persons.[45] Moreover, "overlapping stage" prioritarianism does *not* suffer the critical deficits of momentary prioritarianism. Stages, by contrast with momentary time-slices, *are* full-blown moral agents, and *can* realize the full range of aspects of human well-being (long-term goal fulfillment, relationships, etc.).

However, in positing that a single human being at a single point in time can "house" a plurality of person-stages, overlapping-stage prioritarianism involves a kind of "four-dimensionalism." These entities are "spread out" over space-time, rather than being wholly present at any one time. I lack space to discuss four-dimensionalism here, but it should be noted that four-dimensionalist accounts of *persons* have been subjected to serious objection, and parallel objections might well be leveled against overlapping person-stages.

Another objection is that overlapping stage prioritarianism places a weird moral premium on events that occur to less psychologically unified persons. Imagine that the person, Oscar, has a highly unified mental life. Oscar has but a single stage. Now consider Sarah, who has 5 person-stages, all overlapping at Sarah aged 30. If Sarah at 30 suffers a painful episode, this affects the utility of 5 distinct person-stages; if Oscar does, this affects the utility of only 1 person-stage. So, ceteris paribus, there is greater moral reason to stop Sarah's pain than Oscar's. This is highly counterintuitive.

(3) *Third possibility: Abandoning Fairness*

Parfit suggests that Reductionism about personal identity will not only reorient distributive justice around distribution between temporal segments of persons (momentary time slices or perhaps stages) rather than distribution between whole persons. It will also reduce the relative *weight* of considerations of distributive justice, as compared to overall well-being.

> [W]hatever their scope, we should give less weight to distributive principles. These principles are often held to be founded on the separateness, or non-identity, of different persons. This fact is less deep on the Reductionist View, since identity is less deep. It does not involve the further fact in which we are inclined to believe. Since the fact on which they are founded is less deep, it is more plausible to give less weight to distributive principles. If we cease to believe that persons are separately existing entities, and come to believe that the unity of a life involves no more than the various relations between the experiences in this life, it becomes more plausible to be more concerned about the quality of experiences, and less concerned about whose experiences they are. This gives some support to the Utilitarian View, making it more plausible than it would have been if the Non-Reductionist View had been true.

[45] By contrast, the no-compensation and temporal bias arguments for momentary prioritarianism – if these were valid – would undercut overlapping stage prioritarianism as much as they would lifetime prioritarianism.

Call this the deflation argument. The argument, if valid, cuts against lifetime prioritarianism in favor of utilitarianism. In coming to see that personhood and personal identity is reducible to psychological and physical facts, we feel less pressure to structure normative deliberation so as to respect the separateness of persons. And because a similar deflation could be accomplished for any other candidate locus of moral concern (such as a momentary time slice or a person-stage), we end up with utilitarianism rather than some non-lifetime version of prioritarianism.

What the deflation argument overlooks is that *all* moral concepts may well supervene on psychological and physical facts –indeed, on physical facts, since psychological facts themselves likely supervene on physical facts. Personhood is not unusual in this regard. Consider two possible worlds which are identical in their physical facts. If psychological facts do indeed supervene on physical facts, then – on Parfit's account of personal identity – the two worlds are identical *qua* facts about persons (how many persons are, how long they endure, etc.). There is no "deep further fact" which would yield a different population of persons in one world than another. *But the same is true* about every other morally relevant concept: for example, death, injury, pain, or well-being. The two worlds are also identical in terms of the lifespan of human beings, the ages at which they die, the injuries they suffer, how much pain they incur, and what their well-being is. Does the absence of a "deep further fact" about death, injury, pain, or well-being mean that these, too, are not morally important concepts? If the realization that personhood supervenes on physical and psychological facts undercuts the moral force of the separateness of persons and of distributive principles, why wouldn't the realization that human pain and well-being also supervene on such facts undercut the force of overall well-being and shift us towards amoralism – the conclusion that nothing has moral significance?

To put the point slightly differently, Parfit conflates the *metaphysical* deflation of the concept of personhood (showing how it can be analyzed in terms of metaphysically more basic facts, in particular physical and psychological facts), with the *normative* deflation of the concept of personhood. There is no logical reason to infer the second sort of deflation from the first; and once we see that all moral concepts are subject to a similar metaphysical deflation, we should strongly resist the inference.

Intuitions about Equalization

I have argued that lifetime prioritarianism has a firm foundation in a claim-across-outcome conception of the moral ranking of outcomes, and (pace Parfit) is not undercut by a sophisticated understanding of personal identity. But "reflective equilibrium" involves bringing into coherence one's theoretical views with intuitions about concrete cases. It is possible, therefore, that such intuitions might prompt us to abandon lifetime prioritarianism, notwithstanding its theoretical warrant.

Indeed, various scholars have pointed to intuitions regarding the equalization of attributes or sublifetime well-being, and have argued that such intuitions cut against whole-lifetime approaches to distributive justice. Here, I divide the intuitions into three categories and consider whether they indeed undermine lifetime prioritarianism. I consider, first, intuitions regarding the equalization of attributes or sublifetime well-being within the life of a single person; second, intuitions regarding the *synchronization* of attribute levels or sublifetime well-being between persons; and, third, intuitions regarding the equalization of attributes or sublifetime well-being between different persons.

A general point to keep in mind is that an individual's lifetime well-being may be a quite complicated function of her attributes during various time periods. Nothing requires that the lifetime utility function be atomistic or (if atomistic) that it fulfill one or another separability or additivity condition with respect to attributes or sublifetime well-being. I will use this insight to help respond to the various challenges to lifetime prioritarianism now to be described.

Attribute-based prioritarianism not only requires some methodology for assigning numerical levels to attributes. To be at all plausible, higher numerical levels of the attributes must be "better." More precisely, my examples will generally assume that attributes are measured so that both sublifetime and lifetime well-being are increasing in the attributes.

Intrapersonal Equalization of Attributes or Sublifetime Well-Being

In a number of articles, Dennis McKerlie has claimed that lifetime prioritarianism is too simplified. He has argued, repeatedly, that some degree of priority should be given to individuals at lower levels of sublifetime well-being. McKerlie has also suggested that some degree of priority be given to individuals at lower attribute levels. McKerlie does not express these claims using mathematical formalism, but much of his discussion – translated into the language of SWFs – seems to suggest that a prioritarian SWF should be applied in an attribute-based or sublifetime manner, or in a manner that hybridizes these approaches with lifetime prioritarianism.

McKerlie eschews reliance on claims about personal identity, and instead appeals to various intuitions: not only intuitions concerning *interpersonal* equalization, to be discussed below, but also intuitions concerning *intrapersonal* equalization, which are the focus of this section.

For the sake of clarity, let us distinguish between intrapersonal equalization of *attributes* and intrapersonal equalization of *sublifetime well-being*. Intrapersonal equalization of *attributes* occurs when an individual is at higher level of some attribute during period *t* than period *t**, and we reduce his attribute level during *t* and increase it during *t** by the same amount, still leaving him at an attribute level in *t** which is lower than or equal to that in *t*. Intrapersonal equalization of *sublifetime well-being* occurs when an individual is at a higher level of sublifetime well-being during period *t* than *t**, and we reduce his sublifetime well-being level during *t* and increase it during *t** by the same amount, still leaving him at a sublifetime well-being level in *t** which is lower than or equal to that in *t*. In other words, intrapersonal equalization of attributes is an intrapersonal, intertemporal, Pigou-Dalton transfer in some attribute; intrapersonal equalization of sublifetime well-being is an intrapersonal, intertemporal Pigou-Dalton transfer in sublifetime well-being.

In his various examples of intrapersonal equalization, McKerlie suggests that both sorts of intrapersonal equalization are intuitively desirable. For example, he approves the intrapersonal equalization of pain, a kind of hedonic attribute.

> [T]he idea behind using the priority view in interpersonal cases also applies to intrapersonal cases. The basic claim is that a benefit is especially important if it is received by someone who is badly off. Nothing limits the application of this idea to cases involving different people. We should be able to say, thinking about a single person, that a benefit will be more important if it is experienced when that person is badly off. A person might think that it is more important to relieve pain when he is suffering intensely than to bring about a larger reduction in milder suffering at some other point in his life.

Discussing the treatment of the aged, McKerlie argues that a life-history in which the individual has a high sublifetime well-being level when young, and a low sublifetime well-being level when older, is worse than a life-history in which there is the same total amount of sublifetime well-being but it is distributed more equally.

Attribute-based or sublifetime prioritarianism can indeed reach a result consistent with intuitions concerning intrapersonal equalization, at least to some extent, as shown in the following tables.[46]

In all these tables, sublifetime utility is the multiplicative product of attribute levels. The attribute-based approach sums a square root of attribute levels; the sublifetime approach is the simple version, and sums the square root of sublifetime utilities. In all the tables, only the levels of the *a* attribute are changed by the move from outcome *x* to *y*.

	Outcome *x*		Outcome *y*	
	Period 1	Period 2	Period 1	Period 2
Joe				
Attribute *a*	50	100	75	75
Attribute *b*	16	16	16	16
Sublifetime	800	1600	1200	1200
Attribute score	25.07107		25.32051	
Sublifetime score	68.28427		69.28203	

In the above table, outcome *y* equalizes the *a* attribute. Because the *b* attribute is at the same level in both periods, this is a case in which equalization in the *a* attribute also equalizes sublifetime utility. Both the attribute-based and sublifetime approaches prefer outcome *y*.

[46] I say "can" because depends on form of sublifetime approach and measurement of attributes.

	Outcome *x*		Outcome *y*	
	Period 1	Period 2	Period 1	Period 2
Joe				
Attribute *a*	50	100	75	75
Attribute *b*	20	40	20	40
Sublifetime	1000	4000	1500	3000
Attribute score	27.86776		28.1172	
Sublifetime score	94.86833		93.50209	

In the above case, outcome *y* equalizes the *a* attribute. However, because attribute *b* is at a different level in period 2 than 1, outcome *y* does not equalize sublifetime utility (i.e., is not a Pigou-Dalton transfer in sublifetime utility) and the sublifetime approach prefers outcome *x*. Thus this is a case in which the sublifetime approach fails to prefer attribute equalization.

	Outcome *x*		Outcome *y*	
	Period 1	Period 2	Period 1	Period 2
Joe				
Attribute *a*	50	200	75	150
Attribute *b*	40	20	40	20
Sublifetime	2000	4000	3000	3000
Attribute score	32.00989		31.70439	
Sublifetime score	107.9669		109.5445	

In the above case, outcome *y* equalizes sublifetime utility. However, because attribute *b* is at a different level in period 2 than 1, outcome *y* does not equalize the *a* attribute (i.e., is not a Pigou-Dalton transfer in the *a* attribute) and the attribute-based approach prefers outcome *x*. This is a case in which the attribute-based approach fails to prefer equalization of sublifetime utility

However, such cases can also be handled by combining lifetime prioritarianism with an appropriately nuanced understanding of the structure of the lifetime utility function. Indeed, as I will now show, the lifetime utility function need not be particularly complicated to favor intrapersonal equalization. As already explained, my working assumption in the next chapter (on estimation), consistent with the approach generally taken by SWF scholars, will be that each lifetime utility function *u*(.) in **U** is atomistic and additive in sublifetime utility. In other words, it takes the form $u(x;i)=\sum_{t=1}^{T} v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$. As I will now show, even if *u*(.) adopts this fairly simple form – let alone a more complicated one – intrapersonal equalization of attributes or sublifetime well-being may well be favored by the lifetime prioritarian.

Let us discuss, first, intrapersonal equalization of *attributes*. Consider, to begin, life-histories that are characterized solely in terms of an individual's consumption in each period. *If* sublifetime utility is a linear function of consumption, then the lifetime prioritarian will have no preference for equalizing an individual's consumption. But sublifetime utility is presumably *not* a linear function of consumption. Economists typically make the very plausible assumption that consumption has diminishing marginal utility: that a given increment of consumption makes a smaller contribution to well-being when the consumer who realizes that increment is at a higher consumption level, than when she is at a lower consumption level.

In the multi-period context, diminishing marginal utility of consumption means that the sublifetime utility function is a *concave transformation* of individual consumption, rather than a linear function. And that, in turn, means, that the lifetime prioritarian using a lifetime utility function of the form $u(x;i)=\sum_{t=1}^{T}v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$ will prefer to equalize an individual's consumption.

As discussed earlier, this idea generalizes to attributes other than consumption, and to outcomes that are characterized in terms of multiple individual attributes. Where the lifetime utility function has the form just stated, the sublifetime utility function *v*(.) might well be non-additive in any given metric of individual attributes. And this in turn shows why the lifetime prioritarian might favor attribute equalization, as the following tables illustrate. To be clear, my claim is not that a sublifetime utility function which is non-additive in a given attribute metric will *necessarily* favor attribute equalization,. Rather, the claim is that non-additivity *can* favor attribute equalization, and that some plausible non-additive forms (such as sublifetime utility functions involving a concave transformation of attribute levels) will.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	Lifetime	Period 1	Period 2	Lifetime
Joe						
Attribute *a*	50	100		75	75	
Attribute *b*	16	16		16	16	
Sublifetime	66	116	182	91	91	182

In the above table, sublifetime utility is the sum of the two attribute levels. Outcome *y* equalizes the *a* attribute as well as equalizing sublifetime utility, but Joe's lifetime utility levels are the same in both outcomes and the lifetime prioritarian ranks them as equally good.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	Lifetime	Period 1	Period 2	Lifetime
Joe						
Attribute *a*	50	100		75	75	
Attribute *b*	16	16		16	16	
Sublifetime	800	1600	2400	1200	1200	2400

This is the same table as used earlier, to show how attribute and sublifetime prioritarianism can prefer equalization. In this table, sublifetime utility is the *product* of the *a* and *b* attribute, and lifetime utility is the sum of sublifetime utility. Outcome *y* equalizes both in terms of attributes and in terms of sublifetime utility. However, Joe's lifetime utility is the same in both cases, and thus any lifetime prioritarian SWF is indifferent. This table illustrates that shifting to a non-additive sublifetime utility function does not *necessarily* mean that lifetime prioritarianism favors attribute equalization.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	Lifetime	Period 1	Period 2	Lifetime
Joe						
Attribute *a*	50	100		75	75	
Attribute *b*	16	16		16	16	
Sublifetime	27.184	32	59.184	30.001	30.001	60.002

This table shows how making sublifetime utility a function of a concave transformation of an attribute can render the lifetime prioritarian favorable to attribute equalization. Now, sublifetime utility is the product of the *logarithm* of the attribute and the *b* attribute. (The logarithm function is concave.) Note that Joe's lifetime utility is now higher in *y*, and thus the lifetime prioritarian favors *y*.

	Outcome *x*			Outcome *y*		
	Period 1	Period 2	Lifetime	Period 1	Period 2	Lifetime
Joe						
Attribute *a*	50	100		75	75	
Attribute *b*	20	40		20	40	
Sublifetime	21.699	42	63.699	21.875	41.875	63.750

This table illustrates a different kind of non-additive sublifetime utility function, equaling the logarithm of the *a* attribute plus the *b* attribute. If sublifetime utility were instead calculated as the product of the two attributes, lifetime utility would be lower in *y* and thus equalizing the *a* attribute would not be preferred. If sublifetime utility were calculated as the logarithm of the *a* attribute times the *b* attribute, lifetime utility would also be lower in *y*, because the *b* attribute is not at the same level in both periods. (These calculations are not shown.) *However*, because of how sublifetime utility is here calculated, Joe's lifetime utility is higher in *y* and thus the lifetime prioritarian favors *y*.

Turn, now, to intuitions regarding intrapersonal equalization of *sublifetime well-being*. It is not apparent, at the outset, how such intuitions can be handled with a lifetime utility function of the form $u(x;i)=\sum_{t=1}^{T} v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$. Such a function makes lifetime utility *additive in sublifetime utility*. Given this sort of additivity, won't lifetime prioritarianism necessarily be indifferent to intrapersonal transfers of sublifetime well-being from high to low periods?

Not necessarily. In presenting cases involving equalization of sublifetime well-being, the philosophical literature rarely pays attention to how, exactly, sublifetime well-being is measured. Assume that *u*(*x*; *i*) does indeed have the form $\sum_{t=1}^{T} v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$. Then *v*(.) is a number which is assigned to a subject's attributes during each period, and which has a certain functional role: the sum of such numbers equals the subject's lifetime utility (which in turn has the functional role of representing spectators' preferences over life-history lotteries).

If $u(.)$ has the form $\sum_{t=1}^{T} v(\mathbf{a}_i(x), \mathbf{a}_{imp}(x))$, then lifetime prioritarianism will *not* favor intrapersonal equalization of an individual's *v*(.) values. But it may favor intrapersonal equalization of sublifetime well-being in some other sense. Imagine that we have some alternative grasp of how to measure sublifetime well-being, other than as a determinant of lifetime utility. For example, we might have intuitions about the level of sublifetime well-being realized by an individual in various periods; about differences in sublifetime well-being between periods; and about whether a given package of attributes is better or worse than being dead during the period. These intuitions might be captured by some function *v**(.); and *v*(.) could be a concave rather than linear function of *v*(.). If so, lifetime prioritarianism would recommend equalization with respect to sublifetime well-being as measured by *v**(.).

I have, up to this point, discussed why intrapersonal equalization of attributes or sublifetime well-being may be favored by a lifetime utility function which has the relatively simple form of being additive in sublifetime utility. It should be noted that shifting to a more complicated form supports a variety of further possibilities for preferring equalization.[47]

The reader might protest that I have simply discussed a variety of ways in which lifetime prioritarianism *could* handle examples involving intrapersonal equalization of attributes or sublifetime well-being. I have not shown that all the examples are most plausibly handled in this manner, rather than by attribute-based or sublifetime prioritarianism.

Systematically surveying different cases of intrapersonal equalization; different possible forms of lifetime utility functions; and different possible approaches to measuring attributes and sublifetime well-being is beyond the scope of this section. It *is* certainly true that a sufficiently powerful case of intrapersonal equalization could push us away from lifetime prioritarianism. But it must be remembered that lifetime prioritarianism has a strong basis in general principles of fairness, while attribute-based or sublifetime prioritarianism do not. For such a case to have this effect, it would need to have the following features: (1) the intrapersonal transfer of attributes or sublifetime well-being is such that no plausible lifetime utility function, combined with lifetime prioritarianism, would approve it; and (2) the intuition in favor of the intrapersonal transfer is very powerful, sufficiently so to overwhelm the theoretical merits of the lifetime view. I am not aware of an example with these features.

<u>Interpersonal synchronization of Attributes or Sublifetime Well-Being</u>

[47] For example, imagine that lifetime utility is the multiplicative product of sublifetime utility in each period, rather than the sum. Then equalizing sublifetime utility increases lifetime utility. For example, if an individual's sublifetime utility levels in the three periods in outcome *x* are, respectively, 60, 85, and 95, and we equalize these to 80 in each period in outcome *y*, then lifetime utility, calculated multiplicatively, is 484,500 in *x* and 512,000 in *y*.

In early work on lifetime egalitarianism, McKerlie identified cases in which, intuitively, we prefer to synchronize individuals' attributes. Although McKerlie has since abandoned reliance on such cases, they are worth considering. He provides the following example:

> [Imagine a society] that contains great inequality, with happier lives attached to certain social positions. But at a fixed time people change places and switch from a superior position to an inferior one or vice versa. One example would be a feudal society in which peasants and nobles exchange roles every ten years. The result is that people's lives as wholes are equally happy. Nevertheless during a given time period the society contains great inequality … If equality between complete lives were all that mattered, an egalitarian could not object to it. But I think that many egalitarians would find it objectionable.

Larry Temkin provides a very similar example: "A caste system involving systematic and substantial biases towards, and differential treatment of, the members of different castes might be objectionable on egalitarian grounds *even if* the demographic composition of the castes periodically changed so that each person was a member of each caste and the *overall* quality of each life was equivalent." Temkin also offers a somewhat different case, in which Job1 has an absolutely wonderful life for the first 40 years of his life, Job2 an absolutely terrible life, and then the two switch places.

> Job1's life has been filled with all the blessing life can bestow. His herds and crops flourish. He and his family are healthy and wealthy. He has the love and respect of all who know him. In addition, his plans are realized, his desires fulfilled, and has complete inner peace. Job2, on the other hand, has led a wretched life. His health is miserable, his countenance disfigured. He has lost his loved ones. He is a penniless beggar who sleeps fitfully in the streets, and whose efforts and desires are constantly frustrated.

A simpler and more modern example of the same sort would be this. Each individual earns a low income in a certain number of years of his life, a high income in the remaining years. Having these sequences be synchronized, so that individuals' annual incomes are completely equal each year, would seem to be better than alternative sequencings – even though the inequality of lifetime income is zero regardless.[48]

	Outcome x					*Outcome y*				
	Period 1	Period 2	Period 3	Period 4		Period 1	Period 2	Period 3	Period 4	
Joe	Peasant	Noble	Peasant	Noble		Peasant	Noble	Peasant	Noble	
Sue	Noble	Peasant	Noble	Peasant		Peasant	Noble	Peasant	Noble	
					Lifetime Income					*Lifetime*
Joe	$20K	$100K	$20K	$100K	$240,000	$20K	$100K	$20K	$100K	$240,000
Sue	$100K	$20K	$100K	$20K	$240,000	$20K	$100K	$20K	$100K	$240,000

[48] The intuitive pull towards reducing the degree of periodic income inequality even if lifetime income is held constant may help explain why much of the literature on income inequality focuses on annual income. However it should be observed that many scholars in this field seem to view lifetime income inequality measurement as the gold standard, and annual income inequality measures as adopted for reasons of data availability.

Translated into the "outcome" setup employed in this book, these cases can be seen to involve the following sort of ranking of outcomes. Consider two outcomes x and y, each of which has T periods. There is a single sequence of individual attribute bundles, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T$. Each individual, in x, receives some permutation of this sequence (in other words, individual i has bundle $\mathbf{a}_1$ in some period, bundle $\mathbf{a}_2$ in some other period, etc; and the same is true for individual j, k, and every other individual). Every individual, in y, also receives some permutation of this sequence. However, in x the individual sequences are synchronized: in each period, every individual has the same bundle of attributes. In outcome y the sequences are not synchronized. We have the intuition that outcome x is morally better than outcome y. How shall such a preference – for short, a preference for interpersonal synchronization -- be accounted for? [49]

Note that lifetime prioritarianism cannot account for a preference for interpersonal synchronization if the lifetime utility function is atomistic, and thus is insensitive to the attributes of other individuals. (This includes the simplified lifetime utility function $u(x;i) = \sum_{t=1}^{T} v(\mathbf{a}_i(x), \mathbf{a}_{imp}(x))$, which is not only atomistic but separable and additive in sublifetime utility; it also includes all other types of atomistic lifetime utility functions.) Note, further, that attribute-based prioritarianism, or sublifetime prioritarianism with an atomistic sublifetime utility function, also fail to explain a preference for interpersonal synchronization.

These observations might prompt us to adopt some SWF which is distinct from all of these, and which is structured so as to be sensitive to the distribution of attributes during each time period. Iwao Hirose has argued along these lines. The natural way to structure such an SWF would be as follows: For each period, measure the inequality of sublifetime well-being during that period. Make the overall ranking of each outcome a function of both overall well-being, and the inequality of sublifetime well-being during each period.

However, such an approach overlooks an important feature about our intuitions regarding interpersonal synchronization – namely, that the attributes in these cases tend to be *salient* and, further, to be linked in some way to social status. This is certainly true of Temkin's caste example and McKerlie's noble/peasant example, as well as my example of synchronizing incomes. It is also true of Temkin's Job1/Job2 example.[50]

Conversely, if we imagine synchronizing attributes that are not salient or, if salient, not socially meaningful, intuitions in favor of synchronizing are weaker or disappear. Imagine that

[49] Note that synchronization, as here defined, means synchronization in attributes – but if the utility function measuring sublifetime well-being is atomistic, attribute synchronization will also be a permutation of sublifetime utility levels that synchronizes them.

[50] One of each Job's misfortunes during the period he is badly off is to be a "penniless beggar who sleeps fitfully in the streets" – an obvious determinant of status. One of each Job's good fortunes during the period he is well off is to be "wealthy."

everyone is the same with respect to his visible attributes, but suffers a period of unhappiness during some stretch of his life – which he manages to hide from others in his society. Is it better if everyone has these private episodes of distress at the same time? It seems not. Or imagine that individuals have high or low levels of attributes (perhaps salient) during short periods of time– sufficiently short that individuals at high levels of the attributes do not thereby gain an elevated social status during the period. Does synchronizing these sorts of attributes improve matters? Again, it seems not, as Klemens Kappel observes:

> In a small society all the male members survive by hard labour in the mines. They work day and night shifts. Thus, while the day shift suffers, the other part of the work force relaxes, and vice versa. On the whole, however, everybody has an equally good life. …. There would be many reasons to improve working conditions in this case, but the persistent simultaneous inequality is not obviously a candidate among them.

McKerlie himself observes:

> [A preference for simultaneous equality] will seem implausible if the time periods during which the inequality is measured are too short. If two people will see a dentist tomorrow, it would tell them to schedule simultaneous appointments so that there will be equality in suffering at that time. Are there serious egalitarian reasons for preferring two 10:30 appointments to an appointment at 10 and an appointment at 11?

These observations cut against the sort of proposal tendered by Hirose. And they cut in favor of handling synchronization cases by combining lifetime prioritarianism with some kind of non-atomistic lifetime utility function. Interpersonal synchronization of certain attributes can affect the well-being of those bearing the attributes. This is most clearly the case with salient, socially meaningful attributes (although the well-being impact of synchronization is not necessarily limited to such attributes). Being an impoverished peasant in a society with landed gentry is worse for the peasant's well-being, than being an impoverished peasant in a society where everyone else is an impoverished peasant too. Similarly, having a periodic income of $20,000, during a period when others have a substantially higher income, is worse for the low-earner's well-being than having a periodic income of $20,000 when everyone else has the same income. (By contrast, the well-being effect of going to the dentist doesn't depend on whether others are going to the dentist at the same time.)

A lifetime utility function which is non-atomistic can pick up these sorts of effects. (For example, an individual's sublifetime utility might be a function both of his consumption, and of whether his consumption is low or high in the overall distribution of consumption.) Lifetime prioritarianism with a lifetime utility function which is atomistic and additive in sublifetime utility is simpler to implement – but if the policy analyst believes that synchronization effects and similar impacts on well-being which are thereby overlooked are substantial, she can certainly use a more complicated, non-atomistic function.

Interpersonal Equalization of Attributes or Sublifetime Well-Being

Interpersonal equalization of an *attribute,* in the sense I will discuss here, occurs when someone is at some level of some attribute during period *t*, and some other person is at a lower level of that attribute during period *t** (either the same period or a different one), and we reduce the first person's attribute level during *t* by some amount, and increase the second person's attribute level during *t** by the same amount, leaving the second person's attribute level still no greater than the first person's. Similarly, *interpersonal* equalization of *sublifetime well-being* occurs when someone is at some level of sublifetime well-being during period *t*, and some other person is at a lower level of sublifetime well-being during period *t** (either the same period or a different one), and we reduce the first person's sublifetime well-being level during *t* by some amount, and increase the second person's sublifetime well-being during *t** by the same amount, leaving the second person's sublifetime well-being level still no greater than the first person's.

In other words, interpersonal equalization of attributes involves an interpersonal Pigou-Dalton transfer in some attribute, and interpersonal equalization of sublifetime well-being involves an interpersonal Pigou-Dalton transfer in sublifetime well-being.

At least in the simple case of outcomes characterized in terms of a single attribute, interpersonal equalization of attributes or sublifetime well-being is straightforwardly approved by attribute-based and sublifetime prioritarianism, and also straightforwardly approved by lifetime prioritarianism where the transferee is at a lower level of lifetime well-being.[51]

	Outcome *x*						Outcome *y*					
	1	2	3	4	5	*Lifetime*	1	2	3	4	5	*Lifetime*
Joe	10	10	10	10	10	50	50	10	10	10	10	90
Sue	90	90	90	90	90	450	50	90	90	90	90	410

The numbers here represent the level of a single attribute and of sublifetime utility. Lifetime well-being is the sum of sublifetime utility. There is a transfer of 40 units of the attribute from Sue to Joe in period 1. Because this is a Pigou-Dalton transfer in attributes, an attribute-based prioritarian using any transformation function will favor it. Because this is a Pigou-Dalton transfer in sublifetime well-being, a simple sublifetime prioritarian using any transformation function will favor it. Finally, because the transferee, Joe, is worse off in terms of lifetime well-being and lifetime well-being is the sum of sublifetime utility, the transfer is a Pigou-Dalton transfer in lifetime terms and any lifetime prioritarian will approve it.

Outcome *y* would also be approved by all three approaches if the numbers displayed attribute levels and sublifetime utility were a concave function of the attribute, rather than a linear function (calculations not shown).

[51] More precisely, interpersonal equalization of *attributes* is straightforwardly approved by all three approaches if sublifetime utility is either linear or concave in the metric used to quantify attributes and lifetime well-being is the sum of sublifetime utility. With this same lifetime utility function, interpersonal equalization of *sublifetime* well-being is approved by all three approaches if sublifetime utility is linear in the attribute. If sublifetime well-being is (plausibly) concave in the attribute, attribute-based prioritarians need not favor sublifetime well-being equalization. Lifetime prioritarians will still straightforwardly do so, if the transferee is at a lower level of lifetime well-being.

As in the case of intrapersonal equalization, discussed in a prior subsection, complications can arise with multiattribute outcomes and transfers that equalize one attribute. So as not to make the analysis overly complicated, the examples in this section involve single-attribute outcomes. The challenge to lifetime prioritarianism in rationalizing attribute or sublifetime well-being transfers where the transferee is at a *higher* level of lifetime well-being is clearly displayed by the single-attribute case.

What about the case where the transferee is *better off* in lifetime terms? Dennis McKerlie, Derek Parfit, and Klemens Kappel have all argued that we have an intuition in favor of interpersonal equalization in certain such cases – namely, cases where the transferee is currently in physical pain. McKerlie, Parfit and Kappel describe the following sort of case. Jim, let us imagine, is at a high level of life-time well-being; Sally, at a lower level. But Jim is in terrible pain right now; Sally is in mild pain. Both are in the emergency room, and we have one vial of pain reliever, which we can use to relieve Jim's pain or Sally's. Intuitively, we should provide the pain medicine to Jim.

Depending on what we imagine about the nexus between the pain medicine, Jim's and Sally's pain levels, and their sublifetime well-being, this case can be seen to involve an interpersonal equalization in a kind of attribute (a hedonic attribute, how the individuals feel); an interpersonal equalization in sublifetime well-being; or both.

Status Quo (no pain relief given to either)

	Current Period	Other periods	Lifetime Well-Being
Jim	Terrible Pain	…..	High
Sally	Mild Pain	…..	Low

Outcome x: Giving Pain Relief to Sally

	Current Period	Other periods	Lifetime Well-Being
Jim	Terrible Pain	……	High
Sally	No Pain	……	Low + slightly better

Outcome y: Giving Pain Relief to Jim

	Current Period	Other periods	Lifetime Well-Being
Jim	Moderate Pain	…..	High + slightly better
Sally	Mild Pain	…..	Low

Depending on the attribute and sublifetime well-being numbers we assign to the states of terrible, mild, moderate, and no pain, outcome *y* could be an equalization relative to *x* in terms of pain, sublifetime well-being, or both.

McKerlie, Parfit, and Kappel each claim that our intuitions in this kind of case argue against lifetime prioritarianism, and in favor of some kind of sublifetime or attribute-based prioritarianism. McKerlie has also suggested that the case generalizes beyond pain. Whenever someone can be seen as suffering a hardship right now, we will feel an intuitive "tug" in favor of relieving his hardship, as opposed to helping someone who is not suffering the hardship, even if the first person is better off in lifetime terms. As McKerlie explains:

> Consider the much-discussed conflict between the interests of Afro-Americans in inner city ghettos and the interest of Asian-Americans who own stores in the same neighborhoods The store immigrants might be recent immigrants who suffered greatly in their countries of origin, experiencing the deep poverty of less-developed countries. Now they are modestly well off, and they can expect even better lives for their children. If we think about lifetimes, the complete life of such an Asian-American might well be worse than

> the complete life of an unemployed Afro-American single mother. Nevertheless, the special concern with poverty applies to the Afro-American who *is* living in poverty, not to the Asian-American who is not. It supplies one reason to support a policy that would help Afro-Americans, even if [it] might be possible to help the Asian-Americans more.

Do the sorts of cases just described really cut against lifetime prioritarianism? It is certainly true that attribute-based and sublifetime prioritarianism rationalize interpersonal equalization, even where the transferee is better off in lifetime terms. It is also true that lifetime prioritarianism will not favor such equalization (in attributes or sublifetime well-being), *if* the lifetime utility function takes the form $u(x;i)=\sum_{t=1}^{T} v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$ *and* the sublifetime utility function is additive-in-attributes.

	Outcome *x*						Outcome *y*					
	1	2	3	4	5	*Lifetime*	1	2	3	4	5	*Lifetime*
Joe	10	90	90	90	90	460	50	90	90	90	90	500
Sue	90	10	10	10	10	130	50	10	10	10	10	90

This is a variation on the table above. Joe, who remains worse off than Sue in period 1, is now better off in all the other periods. As above, the numbers represent sublifetime utility levels as well as attribute levels. As above, outcome *y* transfers 40 units of the attribute from Sue to Joe in period 1. Because Joe is now better off than Sue in lifetime terms, *and* because sublifetime utility is linear in the attribute, this transfer is a Pigou-Dalton disequalizing transfer in lifetime well-being, and any lifetime prioritarian disapproves it. By contrast, attribute-based and simple sublifetime prioritarians favor *y*.

On the other hand, the key point I stressed in my discussion of *intra*personal equalization – that the sublifetime utility function is very plausibly non-additive in attributes – is just as relevant in discussing *inter*personal equalization. Using such a sublifetime utility function can prove surprisingly powerful in prompting the lifetime prioritarian to favor interpersonal *attribute* transfers even where the transferee is at a higher lifetime well-being level.

			Outcome x	
		Current year	Other 99 years	Lifetime Utility
Joe:	Income	$10,000	$90,000	
	Sublifetime utility	4	4.954	494.470
Sue:	Income	$90,000	$10,000	
	Sublifetime utility	4.954	4	400.954

			Outcome y		
		Current year	Other 99 years	Lifetime Utility	Change from *x*
Joe:	Income	$50,000	$90,000		
	Sublifetime utility	4.699	4.954	495.169	0.699
Sue:	Income	$50,000	$10,000		

Sublifetime utility	4.699	4		400.699	-.255

Joe and Sue each live for 100 years. In outcome *x*, Joe has the higher income ($90,000) in all years except the current year, where his income is lower ($10,000). Sue has the lower income in all years except the current year. Sublifetime utility is now calculated using the logarithm of income. Although Joe is better off in lifetime terms than Sue, equalizing current income (outcome *y*) yields a larger improvement in Joe's lifetime utility than the loss in Sue's. (This is because the logarithm function is concave and because lifetime utility sums up sublifetime utility.)

Thus a lifetime prioritarian may well approve *y*. Indeed, it can be shown that inequality aversion of the Atkinson SWF must be increased to γ larger than 3 [find exact value] before the lifetime prioritarian favors *x*.

Behind the numbers, the critical point is this: If the sublifetime utility function is non-additive in attributes, attribute equalization can[52] produce a greater increase in the sublifetime utility of the transferee than the loss in sublifetime utility of the transferor, and thus a greater increase in the *lifetime* well-being of the transferee than the loss in lifetime well-being of the transferor. This in turn means that a lifetime prioritarian SWF which is not too inequality averse will approve the transfer even though the transferee is at a higher level of lifetime well-being.

Interpersonal transfers in *sublifetime well-being* are a bit trickier. In discussing *intrapersonal* equalization of sublifetime well-being, I noted that this might be supported even with a lifetime utility function of the form $u(x;i)=\sum_{t=1}^{T} v(\mathbf{a}_i(x),\mathbf{a}_{imp}(x))$ if sublifetime well-being is measured using some metric other than *v*(.). The same is true with regards to interpersonal equalization of sublifetime well-being levels in favor of a transferee at a higher lifetime well-being level.

In short, the lifetime prioritarian *may well* support some degree of interpersonal attribute or sublifetime well-being equalization where the transferee is better off in lifetime terms. Whether she does so depends on the form of the lifetime and sublifetime utility functions and the way in which attributes and sublifetime well-being are measured.

It might be objected that this strategy for handling the sorts of cases described by McKerlie, Parfit and Kappel is inadequate. First, it might be observed, our strongest intuitions in favor of equalization, notwithstanding the fact that the transferee is better off in lifetime terms, involve *hardship*. To put the point more formally, there is a kind of threshold effect with respect to equalization in any given attribute: to be in a condition of hardship or poverty with respect to some attribute is to be below a threshold level with respect to that attribute. Consider an

[52] I say "can" because whether the lifetime prioritarian will approve attribute equalization depends on the particular form of the sublifetime utility function. For example, if the function were *convex* in the attribute, the lifetime prioritarian would not approve equalization. However, in the one-attribute case, sublifetime utility functions which are convex rather than concave or linear in natural metrics of attributes are unusual.

individual's hedonic state -- the attribute at issue in the sort of pain case described by Kappel, Parfit, and McKerlie. In that case, one individual is at a very low hedonic level in some period (he's in great pain) and we intuitively want to provide him pain relief, whatever his lifetime well-being. But imagine, now a case in which both individuals are at a reasonably good hedonic level in some period. Jim is not in pain; he is just bored. Sally is in a good mood. Jim has a higher lifetime well-being level than Sally. We can provide some entertainment to Jim, which will increase Jim's hedonic state by a given amount, or to Sally, which will increase her hedonic state by the same amount. In this case – which differs only from the pain case in that Jim and Sally are now above rather than below a hedonic threshold –the reader may well intuit that Sally should get the entertainment, or at least will not intuit that Jim should. And this threshold effect may well generalize to other attributes (nutrition, shelter, income, health intervention).

However, the threshold effect just described does not necessarily count against lifetime prioritarianism. It can be handled, at least in principle, via a lifetime utility function which incorporates a non-additive sublifetime utility function of a particular kind: namely, one with a threshold.[53]

A stronger objection is that intuitions in favor of equalization, in hardship cases, are *insensitive* to the lifetime well-being levels of transferor and transferee. If Jim is in terrible pain right now and Sally is in mild pain right now, then – intuitively – we should provide a unit of pain relief to Jim rather than Sally quite independent of their lifetime well-being levels. The lifetime prioritarian cannot account for this. However nuanced the lifetime utility function may be – regardless of whether the sublifetime utility function is non-additive, whether it incorporates a threshold, and so forth – whether the lifetime prioritarian approves a given interpersonal transfer of attributes or sublifetime well-being *does* surely depend upon the lifetime well-being levels of transferor and transferee.

This is indeed an important difficulty for lifetime prioritarianism. To reiterate: we intuitively feel a tug in favor of relieving someone's current hardship (pain, low income, poor health), as against helping someone who is currently better off, *quite independent of the lifetime well-being levels of the two individuals.* For short, let us call the apparent normative reason to relieve short-term hardship, the strength of which is independent of the lifetime well-being of the beneficiary, a LWI ("lifetime well-being independent") reason to relieve hardship. Thus an important objection to lifetime prioritarianism, suggested by McKerlie's, Kappel's, and Parfit's examples, is that lifetime prioritarianism cannot account for the existence of LWI reasons to relieve hardship.

[53] Moreover, as I have already suggested, the idea that there is a threshold in the function from certain attributes to well-being has independent plausibility. (Of course, this sort of threshold will have implications for intrapersonal as well as interpersonal equalization.) It should also be noted that threshold effects cut against attribute- and sublifetime prioritarianism.

In response, the lifetime prioritarian can point out – to begin – that LWI reasons to relieve hardship, if they exist, are not grounded in *fairness*. If Jim and Sally are normal human beings, who each remains the same person for his or entire life, and Jim is better off than Sally in lifetime terms, and the pain relief would produce no greater change in Jim's lifetime well-being than in Sally's, then Jim has a *weaker* claim to the relief than Sally. To give Jim the relief would be *unfair*. Because fairness is best specified by lifetime prioritarianism, and because LWI reasons are not based on fairness, it is not surprising that the lifetime prioritarian cannot rationalize such intuitions.

The person-centered welfarist might therefore conclude that there is no normative reason at all to provide Jim the relief. She might see this as a case where an attractive theoretical framework (person-centered welfarism, the claim-across-outcome view of the moral ranking of outcomes, whole persons as the loci of moral concern) pushes her to a point of reflective equilibrium in which certain intuitions need to be rejected.

Here, it is worth noting that intuitions in favor of LWI reasons to relieve hardship are not universally shared. In the field of health policy, a substantial amount of survey work has been undertaken to determine the public's judgments regarding the allocation of scarce treatment for various diseases – in particular, whether treatment should be allocated based on the disease's severity, or based on indicators of lifetime well-being (for example, the patient's age; presumably a longer life yields more lifetime well-being, ceteris paribus). Some respondents prefer treatment based on severity, but other prefer to channel treatments to individuals at lower expected lifetime well-being levels. [CHECK]

Alternatively, the person-centered welfarist might take the position that there is a *non-moral* reason to provide Jim the relief. On this manner of thinking, person-centered welfarism and fairness are coextensive; person-centered welfarism is best specified by lifetime prioritarianism. Morality gives rise to moral reasons, but there are also various non-moral reasons, such as reasons to promote animal well-being and reasons to reduce suffering, both animal suffering *and the suffering or other hardship of human beings independent of their claim to such relief as a matter of fairness.* Jaime Mayerfield has argued strongly that there is a reason independent of prioritarianism to alleviate suffering:

> [W]hereas the intrinsic property view emphasizes the *intrinsic awfulness of suffering*, the priority view emphasize the *harm that suffering does to persons*. … The intrinsic property view doesn't particularly care *who* is hurt; the identity of the victim is not an issue. In this it resembles utilitarianism, which has been criticized on the grounds that it treats persons as mere vessels of happiness and suffering. For the intrinsic property view, it is the evilness of suffering that counts, not the harm done to a particular persons. … By contrast, the priority view directs its attention, not to the intrinsic evilness of suffering, but to the person affected by it. It says that the urgency of helping this person increases the worse off he or she is. … It sets itself apart from both utilitarianism and the intrinsic property view in asserting the moral separateness of persons.

Mayerfield also observes: “My own view is that the suffering of non-human animals carries no less moral weight than the suffering of humans, and that consequently the duty to relieve suffering applies with equal force to both.”

But why not take a third position? Why not say that person-centered welfarism encompasses not only fairness but also certain non-fairness considerations – namely, LWI reasons to relieve hardship – which, like fairness, are *moral* considerations?[54]

The difficulty, here, is reconciling LWI reasons to relieve human hardship with the Pareto principles. Person-centered welfarism says, of course, that the moral ranking of outcomes must satisfy the Pareto principles. As I have argued at length in this Chapter, the ranking of outcomes in terms of fairness needs to satisfy the Pareto principles in whole-lifetime terms. But an LWI reason to relieve hardship is recalcitrant in terms of the lifetime Pareto principles. Imagine a case in which Jim suffers short-term hardship in outcome *x* but not *y*. Other attributes, however, compensate him for the hardship: his lifetime well-being in *x* is equal to or even greater than his lifetime well-being in *y*. A moral reason to alleviate Jim’s hardship, if seen as strong enough to override fairness considerations and thus to rank *y* as all-things-considered morally better than *x*, would violate lifetime Pareto indifference or superiority.

Creatively, we might imagine a hybrid version of person-centered welfarism that conjoins the lifetime Pareto principles (to track fairness) and short-term Pareto principles (to track the LWI reason to alleviate hardship). But, for reasons discussed in the margin, this strategy for accommodating that reason within a Paretian moral view also fails.[55]

To summarize a complicated discussion: The lifetime prioritarian may well approve interpersonal equalization of attributes or sublifetime well-being, even where the transferee is at a higher level of lifetime well-being. She will do so if – by virtue of the form of the lifetime utility function and the way in which sublifetime well-being is measured – the increase in lifetime well-being of the transferee is sufficiently greater than the loss in lifetime well-being of

[54] Mayerfield, himself, emphatically classifies the duty to relieve suffering as a moral one.

[55] The conjoined version of Pareto indifference says: if each person is equally well off in outcome *x* as outcome *y*, both in lifetime terms and in each short period, the two outcomes are equally morally good. The conjoined version of Pareto superiority says: if everyone is at least as well off in outcome *x* as outcome *y*, both in lifetime terms and during each short period, and at least one person is strictly better off during a whole lifetime or a short period, outcome *x* is better.

The LWI reason to alleviate hardship (if such exists) has a threshold structure: there is a reason independent of Jim’s long-term well-being to bring him up to a certain level of attributes, not to increase his short-term well-being indefinitely. (For example, there may be an LWI reason to stop his pain, but not to stop his boredom). To see why the threshold structure of the reason to alleviate hardship runs afoul of the conjoined specification of the Pareto principles now under discussion, consider a case in which Jim is above a threshold with respect to some attribute during some short period. Imagine increasing his well-being during that period, and making a compensating change at some other point so that his lifetime well-being is equal. Then the conjoined version of Pareto superiority would say that this package of changes is a moral improvement – but fairness does not argue for such a package, and *neither does* the LWI reason to alleviate hardship.

the transferor, *and* this difference swamps the difference in their lifetime well-being levels. What the lifetime prioritarian will *not* approve is equalization *independent of* the lifetime well-being levels of transferee and transferor.   LWI reasons to relieve hardship would recommend such equalization – but such reasons cannot be seen as moral reasons, at least within the framework of a moral view which focuses on human well-being and does so by accepting the Pareto principles. Such reasons might be seen as illusory, lacking normative force entirely – after all, they are not grounded in the fair treatment of persons – or, alternatively, as non-moral reasons, akin to the way person-centered welfarism conceptualizes animal well-being.[56]

[56] The fact that a Paretian moral view must reject LWI reasons to relieve hardship might be seen as a yet another reason to reject the Pareto principles. Perhaps so: although given the plausible case that short-term human suffering has no greater normative weight than short-term animal suffering, I don't think this implication of person-centered welfarism is a strong count against it.