https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

AGGREGATION WITHIN LIVES*

By Larry S. Temkin

I. Introduction

In recent years, many philosophers have discussed problems of aggregation. Much of this discussion has focused on problems of additively aggregating across lives in the assessment of outcomes. In this essay, I shall suggest that problems of additive aggregation arise within lives, as well as between lives. More specifically, I shall suggest that in certain cases anti-additive aggregationist principles apply within lives, as well as between lives, and hence that we should reject a widely accepted conception of individual self-interest. As I shall note at the end of this essay, ultimately, the implications of this position are significant and far-reaching, though I cannot adequately pursue them here.

The essay is divided into seven further sections. In Section II, I offer a general account of aggregation. In Section III, I present two examples of problems of additive aggregation across lives: Derek Parfit's Repugnant Conclusion,[1] and my Lollipops for Life Case. In Section IV, I suggest that many people may have misdiagnosed the source and scope of anti-additive aggregationist considerations, due to the influence of John Rawls's and Robert Nozick's claims about the separateness of persons. Accordingly, I suggest that many people, including Rawls and Nozick, leave unchallenged Henry Sidgwick's conception of individual self-interest, which incorporates an additive aggregationist approach for assessing the value of individual lives. In Section V, I follow Parfit in suggesting that some who have stressed the importance of the separateness of persons may have conflated the issue of compensation with the issue of moral balancing. While compensation is possible within lives, but not between lives, that does not settle the question of whether moral balancing is possible between lives, nor does it settle the question of whether an additive aggregationist approach is plausible within

* Much of this essay is taken from a long manuscript I am working on tentatively titled *Rethinking the Good, Moral Ideals, and the Nature of Practical Reasoning*, in particular, chapter four of the current draft, "On the Separateness of Individuals, Compensation, and Aggregation within Lives." Parts of Section II also appear in my essay "A 'New' Principle of Aggregation," in Ernest Sosa and Enrique Villanueva, eds., *Philosophical Issues,* vol. 15, *Normativity* (Boston and Oxford: Blackwell Publishing, 2005), 218–34. I am grateful to Ruth Chang, Tom Hurka, Shelly Kagan, Frances Kamm, Jeff McMahan, and Derek Parfit for useful comments.

[1] Parfit's Repugnant Conclusion is presented and discussed in part four of Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984), 381–90.

doi:10.1017/S0265052509090013
© 2008 Social Philosophy & Policy Foundation. Printed in the USA.

lives. In Section VI, I offer numerous considerations in support of the claim that in certain cases an additive aggregationist approach is deeply implausible for determining the overall value of a life. In Section VII, I suggest that an analogue of Parfit's Repugnant Conclusion, the Single Life Repugnant Conclusion, is also genuinely repugnant, offering further support for rejecting a simple additive aggregationist approach for assessing individual lives. I also respond to a number of sophisticated reasons offered by J. M. E. McTaggart for why we should not trust our intuitions about such cases. Finally, in Section VIII, I conclude with a summary of the essay, and a brief indication of further work that needs to be done.

This essay is only one piece of a much larger research project.[2] But it is a piece that raises some important issues that have been largely overlooked.

I should note a methodological point before proceeding. I follow Sidgwick in thinking that an important task for moral philosophy, though certainly not the only one, is to seek "a coherent, systematic, and non–ad hoc method of accommodating and explaining both our pretheoretical . . . [intuitive moral] judgments and our firm considered ones. Moreover, the method sought should offer a plausible and principled way of assessing—and, where appropriate, revising—people's . . . [intuitive moral] judgments."[3] On this view, people's pretheoretical intuitive moral judgments, and firm considered ones, are important data to be considered in developing one's moral theory. They are among the first words, as it were, in the enterprise of moral theorizing, though certainly not the last. This helps explain the relatively prominent role this essay gives to what people believe regarding certain moral claims. I should add that when I claim that "many or most people believe" such and such, this is usually a generalization from intuitive responses I have received to various cases from hundreds, and often thousands, of people over the years. Often the respondents were "naïve" relatives, friends, or undergraduates, in which case their responses were usually untutored pretheoretical moral intuitions. But often the respondents were professional moral philosophers, in which case their responses were more likely to reflect considered moral judgments. For my purposes here, I don't need to distinguish between pretheoretical and considered intuitive moral judgments. Though I agree that the latter should generally be accorded more weight than the former, for the cases discussed in this essay there was no significant or systematic difference between the views of "untutored naïve" respondents and those of professional moral philosophers.

[2] Temkin, *Rethinking the Good*. See note *.

[3] The slightly amended quotation is from my book *Inequality* (New York: Oxford University Press, 1993), 6. Sidgwick's methodological approach is developed in Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907).

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

II. Aggregation: A General Account

Aggregation is the process or method of combining different parts into a whole. In ethics, there are principles of aggregation for assessing individual lives. Basically, these tell us how the different elements or components of individual lives combine to affect the goodness, comparative quality, or desirability of those lives. Likewise, there are principles of aggregation for assessing outcomes. Basically, these tell us how the different elements or components of outcomes combine to affect the goodness, comparative quality, or desirability of those outcomes.

It is often *assumed* that the relevant components to be combined in assessing the goodness or desirability of outcomes are the individual lives of sentient beings, or perhaps the goodness or desirability of the individual sentient lives within those outcomes. But more inclusive conceptions of the relevant components are possible. For example, some people might believe that the continued existence of certain nations, groups, customs, or cultures contributes to the goodness of the outcomes of which they are a part, over and above the impact their continued existence has on the individual sentient lives in those outcomes. Similarly, some people believe that the existence of natural wonders, like the Grand Canyon, Mount Everest, or Victoria Falls, or of unspoiled wilderness or jungles, or of different species of flora and fauna, contributes to the value of the earth over and above the impact their existence has on the lives of sentient beings. Likewise, some perfectionists believe that there is value in the existence of great works of art, music, poetry, or architecture, beyond the extent to which such works affect the lives of sentient individuals. For now, I take no stand on what kinds of factors might be relevant to the goodness of outcomes, or how they might contribute to such goodness. But a principle of aggregation for assessing outcomes must give *due* weight to *each* of the factors that the full theory of the good deems relevant to the goodness of outcomes—combining them in *whatever* way is ultimately dictated by the full theory of the good.

In philosophy, aggregation is often associated with simple summing. This is partly the legacy of classical utilitarianism, which is often associated with an aggregative approach to ethics for assessing both individual lives and outcomes. As classical utilitarianism is frequently interpreted, the goodness or desirability of a person's life is just the sum total—that is, a simple additive function—of the positive and negative experiences, or utility, contained within that life. Likewise, the goodness or desirability of an outcome is just the sum total—that is, a simple additive function—of the positive and negative experiences, or utility, contained in each sentient life within that outcome.

Although aggregation is often associated with simple summing, *additive* aggregation is but one of many ways of combining the different elements relevant to the goodness of lives or outcomes, to generate an

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

assessment of the comparative desirability of those lives or outcomes. Indeed, as I shall use the notion, a principle of aggregation will be *any* function from the relevant elements—whatever those turn out to be—to the assessment of the goodness or comparative desirability of the item being assessed, say, a life or an outcome. Thus, for example, one might have principles of aggregation that focused on averages, like average utilitarianism, or on weighted totals, like prioritarianism,[4] or on the highest or best achievements, like some forms of perfectionism, or on the welfare levels of the worst off, like maximin. Thus, as I use the notion, a principle of aggregation may accommodate distributive considerations in assessing individual lives or outcomes. How *much* good obtains in lives or outcomes may or may not matter as much as *how* that good is *distributed* within lives or outcomes.

III. Problems of Additive Aggregation

For some time now, people have recognized and discussed what have been called "problems of aggregation." In fact, however, the problems in question are usually problems of *additive* aggregation. Such problems, which reflect a powerful *anti-additive* aggregationist position, have been widely discussed as objections to total utilitarianism—where, roughly, total utilitarianism is the view that the right thing to do is to maximize the *sum total* of goodness (utility) obtaining in the world. Let me note two examples opposing additive aggregation.

The first is Derek Parfit's Repugnant Conclusion: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living."[5]

The second is what I call the Lollipops for Life Case. Suppose there were two alternative universes. In one, countless people would each receive many licks of different lollipops over the course of their lives. Unfortunately, there would also be one innocent person who suffered unbearable agony for fifty years, before eventually dying a slow, lonely, torturous

[4] Prioritarianism is roughly the view that one wants each person to fare as well as possible, but the worse off someone is, in absolute terms, the greater priority is given to improving his or her position. I first introduced and discussed this view, which I then called "extended humanitarianism," in my 1983 Princeton University Ph.D. dissertation, "Inequality." Derek Parfit later discussed this view, which he called the Priority View, in his essay "Equality or Priority?" which he first delivered as the Lindley Lecture at the University of Kansas in 1991, and which was copyrighted by the University of Kansas Philosophy Department in 1995. Parfit's essay has been reprinted in Matthew Clayton and Andrew Williams, eds., *The Ideal of Equality* (London: MacMillan Press Ltd., and New York: St. Martin's Press, 2000), 81–125. I first introduced the term "prioritarianism" and its cognates in my essay "Equality, Priority, and the Levelling Down Objection," which also appeared in *The Ideal of Equality,* 126–61.

[5] Parfit, *Reasons and Persons,* 388.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

death. In the second universe, each of the countless people would receive *one* less lick of a lollipop over the course of their lives, but the innocent person would be spared the agony and instead live a full, rich life. Which universe would be better?

Most people firmly believe that the Repugnant Conclusion is, indeed, repugnant. They believe that an outcome, A, with at least ten billion people, all with a very high quality of life, would be better than an outcome, Z, with a larger population all of whom have lives that are barely worth living, no matter *how many* people live in Z. Similarly, most firmly believe that no matter *how many* people would get the *one* extra lollipop lick, the universe without the extra licks would be better than the universe with them, if the latter, but not the former, involved an innocent person suffering fifty years of unbearable agony followed by a slow, torturous death.

Notoriously, total utilitarians are committed to rejecting these claims. However counterintuitive it may seem, they are committed to the simple additive aggregationist view that if only there are *enough* people in Z, Z would be better than A, and if only *enough* people get an extra lick, the universe with the extra lollipop licks would be better than the one without them.

Should we regard such examples as serious objections to total utilitarianism? Total utilitarians, of course, think not. They deny that the so-called "problem of additive aggregation" is in fact a problem. Insisting that more of the good is better than less of the good, they offer a number of sophisticated explanations for why we have the intuitions we do about such cases, arguing, in essence, that our intuitions were not developed to adequately respond to such wild science-fiction cases involving unimaginably large numbers of relatively small amounts.[6] Correspondingly, they contend that we do not intuitively grasp how *much* total good is dispersed among Z's innumerable masses, or how *much* total pleasure would be enjoyed by countless people each getting one extra lick of a lollipop.

For the total utilitarian, then, no matter how small the amount of good may be in a life that is barely worth living, or how small the amount of pleasure may be from one lick of a lollipop, if only there are *enough* such lives, or licks, *eventually* the total amount of good or pleasure will outweigh, and then be better than, *any* finite amount of good or pain that might be balanced off against it. Here, our understanding supposedly leads us to recognize truths that our imagination fails to appreciate; both

[6] This strategy for dealing with such cases was made famous by the total utilitarian Richard M. Hare. See, for example, his "Ethical Theory and Utilitarianism," in Hywel D. Lewis, ed., *Contemporary British Philosophy, IV* (London: Allen and Unwin Press, 1976), 113–31. I should add that it is not only strict total utilitarians who make such moves. Fred Feldman made similar moves and claims in correspondence he sent me about these issues. Likewise, John Broome has made such claims in response to my arguments; see Broome, *Weighing Lives* (Oxford and New York: Oxford University Press, 2004).

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

the Repugnant Conclusion and the Lollipops for Life Case can be rejected as objections to utilitarianism.

The utilitarian's position is coherent and admirably consistent. But it reminds one of Emerson's contention that "a foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines."[7] Few people are willing to "bite the bullet" and accept the total utilitarian's response for the Repugnant Conclusion and the Lollipops for Life Case. I believe they are right not to do so.

The simple additive aggregationist approach of total utilitarianism has been rejected by philosophers representing a wide range of philosophical positions. This includes consequentialists like Parfit and James Griffin, deontologists like Frances Kamm, contractualists like John Rawls and Thomas Scanlon, libertarian rights theorists like Robert Nozick, and pluralists like myself.[8] Unfortunately, it is much easier to see that additive aggregation has deeply implausible and unacceptable implications, than to develop a plausible and coherent alternative that does not itself have deeply implausible and unacceptable implications. This is disturbing and important, but I have argued for this elsewhere, and shall not take up these issues here.[9]

IV. Misdiagnosing the Source and Scope of Anti-Additive Aggregation

Despite widespread agreement that we should reject simple additive aggregation, I think the basis for this view has been partially misdiagnosed. Correspondingly, I think the full scope of anti-additive aggregationist principles has not always been recognized. This is, I think, largely due to the influence of three philosophical giants: Henry Sidgwick, John Rawls, and Robert Nozick. To explain, I shall first present, in some detail, a brilliant argument of Sidgwick's in support of utilitarianism. I shall then

[7] The quote is attributed to Ralph Waldo Emerson, from his *Essays: First Series: Self-Reliance.* It is reprinted in John Bartlett, *Familiar Quotations,* 10th ed., revised and enlarged by Nathan Haskell Dole (Boston: Little, Brown, 1919).

[8] See Parfit, *Reasons and Persons;* James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Oxford University Press, 1986); Frances Kamm, *Intricate Ethics* (New York: Oxford University Press, 2007); Kamm, "Health and Equity" (unpublished manuscript); John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971); T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998); and Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974). I have argued against additive aggregation in many places, including my essay "A 'New' Principle of Aggregation," but most extensively in my manuscript *Rethinking the Good.*

[9] For some of the disturbing implications, including problems of iteration, and threats to the transitivity of all-things-considered better than, see my "A 'New' Principle of Aggregation"; my "A Continuum Argument for Intransitivity," *Philosophy and Public Affairs* 25, no. 3 (1996): 175–210; and my "Rethinking the Good, Moral Ideals, and the Nature of Practical Reasoning," in Jonathan Dancy, ed., *Reading Parfit* (Oxford: Basil Blackwell, 1997), 290–344.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

note, and comment on, the response Sidgwick's argument prompted in Rawls, which in turn sparked Nozick.

In Book III of *The Methods of Ethics,*[10] Sidgwick provides the basis of an ingenious argument for utilitarianism. Sidgwick begins with a "rational intuition" about prudence, or rational self-love, which he describes as an "equal and impartial concern for all parts of our conscious life."[11] Of this intuition, which is, in essence, premise one of his argument, Sidgwick writes:

> We might express it concisely by saying 'that Hereafter *as such* is to be regarded neither less nor more than Now.' . . . 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 than to that of another. The form in which it practically presents itself to most men is 'that a smaller present good is not to be preferred to a greater future good' (allowing for differences of certainty). . . . [T]he principle . . . is equally applicable to any . . . interpretation of 'one's own good,' in which good is conceived as a mathematical whole, of which the integrant parts are realized in different parts or moments of a lifetime.[12]

Sidgwick next adds a "rational intuition" about "Universal Good," which serves as premise two of his argument. He claims that:

> We have formed the notion of Universal Good by comparison and integration of the goods of all individual human—or sentient—existences. And here again, just as in the former case, by considering the relation of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any one individual is of no more importance from the point of view . . . of the Universe, than the good of any other. . . . And it is evident to me that as a rational being I am bound to aim at good generally . . . not merely at a particular part of it.[13]

Sidgwick concludes:

> From these two rational intuitions we may deduce . . . the maxim of Benevolence in an abstract form: viz. that each one is morally bound to regard the good of any other individual as much as his own,

[10] See note 3 above.
[11] Sidgwick, *The Method of Ethics,* 124n1.
[12] Ibid., 381.
[13] Ibid., 382.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

> except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.[14]

This argument provides the basis for classical utilitarianism only given certain assumptions. One assumption is that when rational self-love, or, as most people would now call it, *rational self-interest,* conceives "one's own good" as a "mathematical whole, of which the integrant parts are realized in different parts or moments of a lifetime," it is conceiving an individual's good as comprised of the amounts of good contained in the different moments of the individual's life *added together*. Moreover, the view "that a smaller present good is not to be preferred to a greater future good" has to be interpreted as a special instance of the general claim that rational self-interest is concerned with the individual's life going as well as possible, and that this involves the life having as *much* good as possible. That is, rational self-interest must be understood as an *optimizing* view that involves *additive maximization,* according to which rationality aims at the *best* achievable life, where this is the one with the greatest *sum total* of good contained within the different moments of that life. Similarly, then, Sidgwick believes that the Universal Good is comprised of the good of each sentient individual, added together, where the best outcome will be the one with the greatest sum total of individual goods. Correspondingly, since Sidgwick believes that rational beings are "bound to aim at good generally, [and] . . . not merely at a particular part of it," rational beings must be utilitarians, and act so as to maximize the good.

Now, strictly speaking, most of the logical work of the argument is done by the second premise—Sidgwick's rational intuition about Universal Good. But the first premise—his rational intuition about prudence or self-love—"softens up" the reader, so as to make the second premise seem more plausible. Intuitively, Sidgwick is reasoning as follows. Rationality requires that I be impartial between the different moments of my life. Thus, I should be willing to accept less in one moment, so as to receive even more in another, as that will make my life go better overall. Likewise, in accordance with "the self-evident principle that the good of any one individual is of no more importance . . . than the good of any other," rationality requires that I be impartial between lives. Correspondingly, Sidgwick suggests, I should be willing to promote less good in one life, so as to promote even more good in another, as that will make the outcome, or "Universal Good," go better overall.

Rawls recognizes this basis for classical utilitarianism. He suggests that, on utilitarianism, "the principle of choice for an association of men is interpreted as an extension of the principle of choice for one man."[15] He then observes that "the striking feature of the utilitarian view of justice is

[14] Ibid.
[15] Rawls, *A Theory of Justice,* 24.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

that it does not matter, except indirectly, how [the] sum of satisfactions is distributed among individuals, any more than it matters, except indirectly, how one man distributes his satisfactions over time. The correct distribution in either case is that which yields the maximum fulfillment."[16] Rawls's response to utilitarianism, and the Sidgwickian line underlying it, is both famous and, to most nonutilitarians, crushing: "Utilitarianism does not take seriously the distinction between persons."[17]

Interestingly, Rawls does not challenge Sidgwick's characterization of the demands of rational self-interest. He allows that perhaps it is rational in my own case to pursue trade-offs between the different *moments of my life* so as to maximize the total amount of good that *I* experience over the course of my life. But he denies that in pursuing the good of society—Sidgwick's Universal Good—we should similarly permit or require trade-offs between different *people* so as to maximize the total amount of good in *society.* Rawls recognizes the striking difference between someone choosing for *herself,* to have less at one moment, so that she might have even more at other moments over the course of her life, and someone imposing a loss on someone *else,* so that *other* persons might have more, collectively. Setting aside important questions about autonomy and rights that arise in comparing these cases, it appears that, in the former case, the one who loses is also the gainer; she is the *beneficiary* of the loss. Thus, in the former case, there isn't *really* a loser, just a winner. In the latter case, the one who loses really *is* a loser. Her situation has helped promote the greater good, but all the benefits of her loss accrue to *others.* In essence, her welfare has been *sacrificed* for the greater good.

Acknowledging his debt to Rawls,[18] Robert Nozick echoes these themes. He writes:

> Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm: we go to the dentist to avoid worse suffering later; we do some unpleasant work for its results. . . . In each case, some cost is borne for the sake of the greater overall good. Why not, *similarly,* hold that some persons have to bear some costs that benefit other persons more, for the sake of the overall social good? But there is no *social entity* with a good that undergoes some sacrifice for its own good. There are only . . . different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Nothing more. What happens is that something is done to him for the sake of others. . . . To use a person in this way does not

[16] Ibid., 26.
[17] Ibid., 27.
[18] See the following note.

> sufficiently respect and take account of the fact that he is a separate person,[19] that his is the only life he has. *He* does not get some overbalancing good from his sacrifice. . . .[20]

Nozick employs such claims in defense of his notion of moral side constraints, and in particular his view that there is "a libertarian side constraint that prohibits aggression against another."[21] Rawls, of course, was developing a theory of social justice; his goal was to determine the correct principles and institutions of a just society. But many people who reject the overall theories of Nozick and Rawls, nevertheless accept their particular claims about the separateness of persons, and see them as providing compelling reasons to reject the simple additive aggregationist approach of classical utilitarianism.

Once doubts about additive aggregation across lives have been raised, there is ample room for other considerations to enter into the assessment of outcomes, besides simply the sum total of the amount of goodness that obtains in those outcomes. So, for example, there is room for *distributive* considerations to enter: what matters is not merely how much good obtains, but also, in part, how the good is distributed.

Accordingly, most people accept anti-additive aggregationist principles like the following ones:

(1) *Trade-offs between quality and number are sometimes undesirable even when vast numbers are at stake:* If the quality of one kind of benefit is "sufficiently" low, and the quality of another kind of benefit is "sufficiently" high, then an outcome in which a relatively small number of people received the higher quality benefit would be better than one in which virtually any number of (otherwise) similarly situated people received the lower quality benefit.[22]

(2) *The Minimize Great Additional Burdens View:* In general, if additional burdens are dispersed among different people, it is better for a given total burden to be dispersed among a vastly larger number of people, so that the additional burden any single person has to bear within her life is "relatively small," than for a smaller total burden to fall on just a few, such that their additional burden is substantial.[23]

[19] At this point Nozick has a note (note 5) citing Rawls on the separateness of persons—specifically, sections 5, 6, and 30 of *A Theory of Justice.*

[20] Nozick, *Anarchy, State, and Utopia,* 32–33 (italics in the original).

[21] Ibid., 33.

[22] I discuss this principle, and numerous other anti-additive aggregationist principles, in much more detail in chapters 2 and 3 of my manuscript *Rethinking the Good.*

[23] This principle is presented and discussed in my "A 'New' Principle of Aggregation," 219–23.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

Examples where people intuitively reason in accordance with anti-additive aggregationist principles like the foregoing ones abound. Most people's reactions to the Repugnant Conclusion and the Lollipops for Life Case noted earlier exhibit such reasoning. Similarly, in accordance with such reasoning, virtually all agree that, other things equal, it would be worse if fifty people suffered from AIDS, quadriplegia, severe psychosis, or being deaf, dumb, and blind, than if virtually *any* number of people suffered from a minor nosebleed, a slight cold, a sprained finger, or a short mild headache.

Importantly, it isn't that people believe that nosebleeds, colds, sprains, and headaches don't *matter.* They do. They are conditions of ill health, albeit mild ones, and, other things equal, they make a claim on our sympathy and resources, although not a particularly strong one. But we *don't* think that the claims of those suffering from such minor afflictions *add up* in the way they would need to in order to outweigh the overwhelmingly powerful claims of those suffering from AIDS, quadriplegia, severe psychosis, or being deaf, dumb, and blind.

Here, the Rawlsian and Nozickian point seems relevant and illuminating. *Given the separateness of persons,* it is natural to suppose that *of course* one should prefer tiny burdens for many rather than huge burdens for one, or a huge benefit for one rather than tiny benefits for many, precisely *because* the burdens or benefits of the many don't *add up across lives* in the way they would need to in order to outweigh the significant impact on the single individual life that the huge benefit or burden would have. Thus, anti-additive aggregationist principles may seem like natural and plausible extensions of the now familiar and widely accepted point about the separateness of persons.

For many, Rawls's and Nozick's point about the separateness of persons—which has obvious Kantian roots—has crystallized the importance of anti-additive aggregationist principles in assessing the goodness of outcomes. Unfortunately, however, I believe that Rawls's and Nozick's focus on the separateness of persons, together with the fact that they left the Sidgwickian conception of rational self-interest largely unchallenged, helped obscure both the basis and the scope of anti-additive aggregationist principles. In fact, I believe that while the separateness of persons is *relevant* to anti-additive aggregationist reasoning, the latter is not dependent on the former. More particularly, I believe that anti-additive aggregationist principles are *not* simply extensions of the familiar point about the separateness of persons, and do not *rely* on that point for their plausibility. Specifically, I believe that anti-additive aggregationist principles are *also* relevant for assessing the goodness of individual lives. That is, on reflection, I believe that anti-additive aggregationist principles apply *within,* as well as between, lives. If this is right, then we should reject the Sidgwickian conception of rational self-interest that many, including both Rawls and Nozick, seem willing

to grant.[24] The view here need not involve rejecting the optimizing element of Sidgwick's conception, according to which one aims at the *best* achievable human life (though one might want to reject such a view on independent grounds);[25] rather, it involves rejecting the additive assumption according to which the best life is the one with the greatest *sum total* of goods spread throughout the life. In Sections VI and VII, I shall offer various examples and considerations in support of this position. First, however, let me briefly comment on the topic of compensation.

V. Compensation Versus Moral Balancing

In part three of *Reasons and Persons,* Derek Parfit claims that advocates of the point about the separateness of persons often conflate the issue of compensation with the issue of moral balancing.[26] I think he is right. When Nozick points out that, when someone is used for the benefit of others, "*He* does not get some overbalancing good from his sacrifice,"[27] this is undeniable as a point about *compensation.* The one who is burdened is *not* compensated by the fact that others may benefit even more than he has been burdened by his "sacrifice." But, as Parfit rightly notes, despite the rhetorical tone of Nozick's remarks, this fact about compensation does not *settle* the question of whether the moral balance of benefits versus burdens is sufficient to make the outcome better, or possibly even to warrant "sacrificing" the one for the others.

Consider the following example. Suppose that in one situation someone was extremely well off at level 10,000, and a million people had lives barely worth living at level 1. Suppose, further, that the people's levels were determined purely by luck, and that, in fact, the fortunate one only deserved to be at level 100, while the unfortunate million each deserved to be at level 20,000. Next, suppose that a second situation resulted from the first one being transformed into one where the best-off person received an uncompensated burden of 1, lowering him to level 9,999, while, thanks to the invention of some new technology, everyone else received benefits of 19,999, raising them to level 20,000. One *might* claim that the second situation was worse than the first, because it involved someone who had received an uncompensated burden. But this would be either absurd or

[24] It is possible that Rawls and Nozick merely grant Sidgwick his conception of rational self-interest for the sake of argument. Still, I think it is fair to claim that Sidgwick was expressing the conception of individual self-interest that has dominated Western thought since Plato. A notable challenge to this conception is presented by Derek Parfit in part two of *Reasons and Persons.*

[25] Interesting arguments against always optimizing, and in favor of the rationality of satisficing, can be found in Michael Slote, *Beyond Optimizing* (Cambridge, MA, and London: Harvard University Press, 1989).

[26] See especially Parfit, *Reasons and Persons,* 336–39.

[27] Nozick, *Anarchy, State, and Utopia,* 32–33 (italics in the original).

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

deeply misleading. One might plausibly claim, perhaps, that the second situation would be worse than the first in *one* respect, involving, as it does, an uncompensated burden for someone. But, clearly, the second situation would be better than the first *all things considered.* The second situation would be vastly superior to the first in terms of utility, maximin, absolute justice, comparative justice, and priority; and, depending on the details of the case, it might even be vastly better in terms of perfection and freedom. Surely, our concern about uncompensated burdens does not "trump" or have lexical priority over our concern about every other ideal combined!

The preceding is obvious but important. In assessing situations, one needs to give due weight to *all* of the factors that are relevant and significant for the overall goodness of outcomes. This will involve much more than questions about compensation. It will involve a careful balancing of the competing benefits, burdens, and interests of different individuals, as well as a careful balancing of each of the moral ideals relevant to the goodness of outcomes.

While in some cases compensation is *relevant* to the desirability of outcomes involving different distributions of benefits and burdens, facts about compensation *itself* do not solely determine the desirability of such outcomes. Rather, what matters is the net effect that the distribution of benefits and burdens has on *all* of the factors and ideals relevant to assessing outcomes. Thus, while in some cases compensation makes the difference between the desirability and undesirability of distributions of benefits and burdens, in other cases it does not, as distributions of benefits and burdens will be desirable, or undesirable, whether they befall the same people or different ones, and hence whether or not they involve compensation. Moreover, importantly, in some cases distributions of burdens and benefits will be *more* desirable if they do *not* involve compensation than if they do. This might be so if the outcomes involving uncompensated burdens were sufficiently better than the outcomes involving compensated burdens regarding such ideals as freedom, perfection, maximin, absolute justice, comparative justice, and equality.[28]

VI. Compensation, Prudential Balancing, and Additivity

As I have noted, Rawls and Nozick appeal to the separateness of persons to challenge Sidgwick's argument for the utilitarian conclusion that the best outcome is the one with the greatest sum total of individual goods. But they do not challenge his assumption that the best achievable life is the one with the greatest sum total of good contained within the different moments of that life. This may be because compensation is

[28] A much lengthier discussion of the issue of compensation versus moral balancing is contained in chapter 4 of my manuscript *Rethinking the Good.*

possible within lives, but not between lives, and they may assume that additive aggregation stands and falls with the possibility of compensation. But this, I believe, is a great mistake. As I suggested above, moral balancing across lives is possible, even if compensation is not, and it is the full set of relevant ideals for assessing the goodness of outcomes, rather than the mere lack of possible compensation, that determines when, and why, additive aggregation should be accepted, or rejected, in assessing outcomes. Likewise, the prudential balancing of different moments of a life will turn on the *full* set of relevant ideals for assessing the goodness of a life, and the correct method of combining the relevant ideals so as to aggregate across the moments of a life is *not* simply determined by the fact that compensation within a life is possible. Specifically, while the possibility of compensation may *permit* an additive approach, and may make such an approach seem more plausible within lives than between lives, it neither entails nor requires such an approach. Indeed, as we shall see, notwithstanding their differences regarding the possibility of compensation, aggregation within lives parallels aggregation across lives, in that there may be many circumstances where the "correct" prudential balancing of different moments of a life involves rejecting additive aggregation. Given this, we should reject the conception of individual self-interest advanced by Sidgwick—what he called prudence, or rational self-love—and left unchallenged by Rawls and Nozick.

Many objections can be raised to a simple additive approach for aggregating the different moments of a life to determine the life's overall goodness. Most of these are familiar, and can be stated briefly. For simplicity, I shall assume that there can be a level of precision in measuring the quality of people's lives, and of different portions of those lives, which is not, in fact, possible. Making this assumption makes it easier to illuminate the shortcomings of the view I am attacking, but does not prejudice the case against it, since the additive approach should not be any less plausible if such precision *were* possible.

Consider a scale of well-being where someone with an average quality of life is at level 500, someone with an extremely high quality of life is at level 1,000, and someone whose life is barely worth living is at level 1. Now suppose one could live for fifty years at level 1,000, or for a hundred years at level 500. On a simple additive approach to measuring the goodness of lives, these two lives would be exactly equally as good; both would contain a total of 50,000 units of well-being. But it is dubious that these two lives are equally good. Perhaps, as long as one is living a "sufficiently" long life (where, let us suppose, fifty years counts as such), quality counts more than duration, so that for any given total of well-being to be experienced, it is more important that the years one lives be at a higher level, than that one merely lives longer at a lower level. Or perhaps, as long as one is living a "sufficiently" high-quality life (where, let us suppose, level 500 counts as such), duration counts more than

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

quality, so that for any given total of well-being to be experienced, it is more important that one lives a longer life at a lower level than a shorter life at a higher level. Either of these views, or more context specific variations of them, might be correct. In either case, this would undermine a simple additive approach to aggregating the different moments of a life.

Or consider two different ways that one's life might sum to a total of 50,000 units of goodness over the course of a fifty-year life. In the first, which I'll call the *drab life*,[29] one would consistently experience a mildly pleasurable existence, of approximately three units of good per day, such that one was equally well-off throughout one's life, and totaled 1,000 units of well-being per year. In the second, which I'll call the *wildly fluctuating life*, one's life would be filled with many highs, but also many lows. During the highs, one would experience great achievements and rich relationships, and would be as well off as is humanly possible, but the lows would involve great setbacks, pain, and suffering, so much so that during those periods one's life would be well below the level at which life ceases to be worth living. The positive periods might total well into the thousands of units of goodness per day, but the negative moments would likewise total to a large number of units of badness, or negative units of goodness, as it were, and would cancel out all but 50,000 units of goodness over the course of one's life. On a simple additive approach to measuring the goodness of lives, these two lives would be equally good. But that seems deeply implausible. I can imagine significant arguments for regarding the second alternative as much better, or, depending on how the details of the case are fleshed out, significant arguments for regarding it as much worse, but what I don't find plausible is the additive approach's claim that as long as there is the same sum total of goodness minus badness in the two lives, they would be equally as good. The differences between the two lives are extremely significant, and their impact on the overall quality of the lives is not to be captured by simply adding up the goodness and badness within each life day by day, or moment by moment.

Here is another way to put the point. As indicated, depending on the details of the case, I think it likely that one of the two lives would be *much* better than the other. Correspondingly, I don't believe that a *slight* increase or decrease in the total benefits or burdens in the two lives would affect their relative rankings. In other words, I don't believe that it would make a difference to the ranking of the two lives whether, for two days out of fifty years, the drab life had reached level 4, so that the sum total of goodness in that life was actually 50,002, or whether, for two days out of fifty years, the drab life had only reached level 2, so that the sum total of goodness in that life was actually only 49,998. Likewise, I don't believe that it would make a difference to the ranking of the two lives whether,

[29] Readers familiar with Parfit's work will recognize that I borrow the expression "drab life" from Parfit, though I characterize such a life a bit differently than he does.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

for two days out of fifty years, the wildly fluctuating life had reached level 2,076, rather than 2,074, so that the sum total of goodness in that life was actually 50,002, or whether, for two days out of fifty years, the wildly fluctuating life had "only" reached level 2,072, so that the sum total of goodness in that life was actually only 49,998. It seems to me clear that the differences between the two lives are so stark, and so significant, that a mere difference of a couple of extra units of good or bad during two days of a fifty-year life span couldn't possibly alter the relative ranking of the two lives.[30] If the drab life would be better than the wildly fluctuating one in the case where the sum total of goodness happened to be 50,002, then it would be better if its sum total happened to be 49,998. Similarly, if the drab life would be worse than the wildly fluctuating one in the case where the sum total of goodness happened to be 49,998, then it would be worse if its sum total happened to be 50,002. Likewise for the wildly fluctuating life.

There are many factors that seem relevant to the overall goodness of a life, besides merely the sum total of goodness or badness that obtains within that life. One such factor might be referred to as the "pattern," "direction," or "shape" of a life.[31] To illustrate this, consider two different scenarios one's life might follow. In the first, one's life starts out poorly and steadily improves, so that at the end of one's life one is very well off. For simplicity, let us just divide the life into five twenty-year segments, and say that as the life progresses one's levels would be 100, 300, 500, 700, and 900, respectively. In the second scenario, the pattern is reversed. One's life starts out very well and steadily worsens, so that one's life ends poorly. Specifically, as one's life progresses, one's levels would be 900, 700, 500, 300, and 100, respectively. On the simple additive approach, the two lives are equally good, and if, in fact, the second life ended at level 120 instead of level 100, then that life would be better than the first, all things considered. I know of few people who would actually choose a life plan for themselves, or for a loved one, in accordance with this position.

To be sure, here, as elsewhere, advocates of the additive approach can argue that we are being misled by our intuitions in such cases. They might claim that the *experience* of living a certain kind of lifestyle, *after*

[30] I am not claiming here that no two lives that differ starkly in certain respects could be equally good, all things considered. Rather, my view is that for the particular lives in question the difference between them is stark in ways that have significant implications for the two lives' overall goodness, such that it is implausible to believe they might, in fact, be equally good, or even almost equally good.

[31] After I had written a draft of this essay, someone pointed out to me that David Velleman also made this point, and other similar ones, in his essay "Well-Being and Time," *Pacific Philosophical Quarterly* 72 (1991): 48–77. This does not surprise me. I believe that many people have seen this point, as well as some of the other points noted in this section. Unfortunately, however, the full implications of incorporating these points into one's theory of the goodness of lives has not always been recognized. See note 9 for some of these implications, and for references to essays where I explore these implications more fully.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

having previously lived a better lifestyle, is actually worse than the experience of living the same kind of lifestyle with nothing better to compare it to. Similarly, they might note that one will *appreciate* living a good lifestyle more if one has previously lived a poor lifestyle, than if one has never known what it is like to be badly off. Correspondingly, it is suggested that while people are *told* that in the first and second lives people's levels are, respectively, 100, 300, 500, 700 and 900, and 900, 700, 500, 300, and 100, *in fact* they are *intuitively* reacting *as if* the levels are really something like 100, 350, 600, 800, and 1,000, and 900, 650, 400, 200, and 50, respectively. And, *of course,* the argument concludes, if we are thinking about the lives that way, we will regard the first as much better than the second, since it will have a much greater sum of well-being.

I do not deny that there are kernels of psychological truth underlying such a position. Ultimately, however, it amounts either to the view that people's judgments about such alternatives are unreflective, or to the view that even on reflection people cannot accurately assess the alternatives presented. I find each of these views both deprecating and dubious. When I think about such cases, I do not have to think that the people are dwelling on the past, anticipating the future, or otherwise experiencing the present through the prisms of past and future. I can, instead, consider someone who basically lives life as it comes, appreciating where he currently is in life, whether good or ill, simply for what it is. When I do this, I can imagine people's lives actually unfolding in the way described above. One life starts off poorly, and ends very well. The other starts out very well, and ends poorly. But both lives have the same total amount of good and bad occurring within them. Thinking about these alternatives as clearly and carefully as I can, I stand by my earlier claims. It seems evident that the former life is better than the latter. While each life has the same total amount of well-being spread among its moments, each life has a pattern or direction that crucially matters.

The claim that the shape of a life can matter resonates with a basic insight common to the thinking of certain holists, gestalt psychologists, and advocates of organic unity: namely, that sometimes the whole can be greater or less than the sum of its parts. While this insight is notoriously difficult to pin down, it seems neither mystifying nor implausible if it is "merely" understood as the rejection of an additive aggregationist approach to understanding the goodness of certain "wholes." On reflection, it seems clear that this insight applies within lives as well as between lives.

In sum, of two lives with the same total amount of goodness and badness, as measured additively moment by moment, one of which began poorly, steadily improved, and ended well, and the other of which began well, steadily declined, and ended poorly, the first seems better than the second. It is the life I would wish for myself, and for anyone I cared about. Moreover, I am confident that most people, even after careful deliberation, would agree.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

Next, let me suggest that an analogue of the anti-additive principle of aggregation, which I presented above (in Section IV) as the Minimize Great Additional Burdens View, applies within lives as well as between lives. To illustrate this, consider the following example.

Suppose one is faced with a choice of living one of two lives. In the first, one will get to live at the exceptionally high level of 910 for a hundred years; in the second, one will get to live at the slightly higher level of 930 for ninety-nine years, but the cost of doing so is a year of *exceptionally* bad pain and suffering, where one's life is *much* lower than the level at which life ceases to be worth living, at a level of −900. As constructed, the simple additive approach would claim that the second life is better, since there would be 91,170 total units of good spread throughout the moments of one's life (here, the additive approach involves subtracting the total amount of bad from the total amount of good), as opposed to 91,000. But I reject this ranking as deeply implausible, and I believe that most people would strongly agree with me.

Here, it seems clear that one should apply reasoning similar to that underlying the Minimize Great Additional Burdens View. Specifically, in this case, where additional burdens are to be distributed among the different periods of a life, it is preferable for a given total burden to be dispersed among many periods of that life, so that the additional burden the person has to bear at any given time is "relatively small," as opposed to having a smaller total burden fall on just one period, or just a few periods, of the life, such that the additional burden the person has to bear during that period, or those few periods, is substantial.

Note that I do not need to claim that such reasoning will always be plausible. There are different cases, and perhaps in some cases one might plausibly reject the reasoning in question. But in cases like the example noted, anti-additive reasoning of the sort underlying the Minimize Great Additional Burdens View seems both applicable and compelling. The first life can be seen as involving a total burden of 1,980 relative to the second, since there is a burden of 20 for each of the ninety-nine years of being at level 910 rather than at level 930, but there is no burden involved in being at level 910 rather than at level −900 for the remaining year. Similarly, the second life can be seen as involving a total burden of 1,810, since there is a burden of that amount involved in being at level −900 rather than at level 910 for a year, but there is no burden involved in being at level 930 rather than at level 910 for the other ninety-nine years. Thus, the *total amount* of burden imposed is greater in the first life than in the second, but in the first life the burden is spread out so that no period is significantly affected adversely, while in the second life the burden all falls on one period, with disastrous consequences for that period. Moreover, importantly, there is a level of pain and suffering in the second life that I believe significantly impacts the overall shape and quality of the life. This is not so in the first life. Surely, it is better to live a life *all* of which is at an

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

exceptionally high level, than to live a life that is slightly better throughout most of one's life, but which includes a *full year* of pain and suffering which is so terrible that, throughout that year, one's life is not even close to being worth living.

A similar anti-additive aggregationist principle applies regarding the distribution of benefits, rather than burdens, within a life. Consider a choice between living a life at the exceptionally high level of 920 for a hundred years, or living a life at the slightly less exceptionally high level of 910 for ninety-nine years in order to reach the truly blissful level of 1,700 for one year. On the simple additive approach, the first life is better, since there would be 92,000 total units of good spread throughout one's life, as opposed to 91,790. But once again I reject this ranking, and I believe most others would as well. The life that includes a year of bliss seems clearly superior to its alternative. It seems to me that having a full year of true bliss within a life that would otherwise be devoid of anything remotely so elevated, would be much more significant, and have a much greater impact on the overall quality of one's life, than would having a lot of small gains throughout many periods of one's life, none of which, at any time, would substantially alter, impact, or distinguish the quality of the different periods of one's life, relative to what they would otherwise be without such gains. And this is so even if the *sum total* of value in one's life would be a bit greater in the latter life. Here, as elsewhere, what matters is not merely the sum total of good within the life, but the way the good is "packaged" or distributed.

I suggest, then, that both within lives and between lives, there will often be reason to consolidate benefits or disperse burdens, as this can sometimes make a significant difference to the overall quality of the lives in question. In general, relatively small benefits *don't add up across people* to have the same impact on the overall goodness of an outcome that a really major benefit would have if it came to a single individual in that outcome. Likewise, in general, relatively small benefits *don't add up across time* to have the same impact on the overall goodness of a life that a really major benefit would have if it came to an individual at a single moment, or period, of her life. Analogous claims apply regarding relatively small burdens versus major burdens. Small burdens don't add up across people or time in the way they would need to in order to offset major burdens that accrue to a single individual, or to an individual at a single moment or period of her life. Thus, in some cases, both the better outcome and the better life will be one with a lesser sum total of benefits, or a greater sum total of burdens, but a better *distribution* of those benefits or burdens.

On reflection, then, the normative significance of benefits and burdens does not depend merely on their sum total. Correspondingly, any theory that hopes to adequately reflect the actual normative significance of benefits and burdens will need to incorporate anti-additive aggregationist principles in certain cases and contexts, both within and between lives.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

That is, both the full theory of the goodness of outcomes and the full theory of the goodness of lives will need to incorporate anti-additive aggregationist elements.

VII. The Single Life Repugnant Conclusion

I have presented various considerations in support of the view that, in certain cases, we need to adopt anti-additive aggregationist principles both within and between lives. In this section, I want to present one final consideration of this sort and discuss it in some detail.

In Section III, I briefly discussed Derek Parfit's Repugnant Conclusion. Let us next consider an analogue of Parfit's Repugnant Conclusion, involving a single life. Paraphrasing Parfit, the analogue might be put as follows:

> *The Single Life Repugnant Conclusion:* For any possible life, no matter how long or how high the quality of that life might be, there must be some much longer imaginable life whose existence, if other things are equal, would be better, even though each period of that life would be barely worth living.

Just as most people view Parfit's Repugnant Conclusion as genuinely repugnant, and see it as illustrating the need to reject a simple additive aggregationist approach for assessing the goodness of outcomes, I believe that most people would view the Single Life Repugnant Conclusion as genuinely repugnant, and see it as illustrating the need to reject a simple additive aggregationist approach for assessing the goodness of individual lives. I believe they would be right to do so.

Imagine the highest form of human life—filled with honor, dignity, knowledge, power, the creation and appreciation of great artistic achievements, good looks, great wealth, fine dining, deep friendships, outstanding health, enduring love, and stupendous sex! Embellish the description as much as you want to make the life seem as desirable as humanly possible. Next, imagine that you, or a loved one, might live such a life for a hundred years, before dying peacefully and painlessly in your sleep. Alternatively, you or your loved one might live for many eons in a state of mostly unconscious suspended animation, but once every decade, perhaps, you might rise to a state of consciousness, during which time you would briefly experience a minute, say, of the slightest noticeable pleasure, perhaps akin to the pleasure of having your hair gently stroked. Surely, most people would choose the former life over the latter for themselves or a loved one, and this is so even if there would be *so* many eons in the latter life, that the total sum of goodness within that life would slightly (or even greatly!) outweigh the vast, but still finite, *total sum* of goodness within the former life. Though the Single Life Repugnant Conclusion might be harder to conceptualize than Parfit's many-person

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

version, few people have any doubt that it is, indeed, repugnant, and that it should be rejected for the same kind of reason that Parfit's original version should be. Sum totals of benefits and burdens are *not* all that matter within lives, any more than between lives.

Interestingly, a version of the Single Life Repugnant Conclusion was first discussed by J. M. E. McTaggart, many years before Parfit presented his own Repugnant Conclusion. But McTaggart denies that the repugnance "in this case would be right."[32] Though lengthy, McTaggart's discussion is brilliant and worth presenting in detail. He writes the following:

> [T]he values of states in time vary, caeteris paribus, with the time they occupy.
>
> From this principle it follows that any value, which has only a finite intensity, and which only lasts a finite time, may be surpassed by a value of much less intensity which lasts for a longer time. Take a life which in respect of knowledge, virtue, love, pleasure, and intensity of consciousness, was unmixedly good, and possessed any finite degree of goodness you choose. Suppose this life prolonged—for a million years if you like—without its value in any way diminishing. Take a second life which had very little consciousness, and had a very little excess of pleasure over pain, and which was incapable of virtue or love. The value in each hour of its existence, though very small, would be good and not bad. And there would be some finite period of time in which its value would be greater than that of the first life, and another period in which it would be a million times greater.
>
> This conclusion would, I believe, be repugnant to certain moralists. But . . . I can see no reason for supposing that repugnance in this case would be right. So far as I can see, it rests on a conviction that quality is something which is inherently and immeasurably more important than quantity. And this seems to me neither self-evident nor capable of demonstration.
>
> Nor is there any reason to doubt our conclusion because it is highly probable that many people, if offered the million years of brilliant life, followed by annihilation or by a state whose value should be neither good nor evil, but zero, would prefer it to any length of an oyster-like life which had a slight excess of good. For it must be remembered that men's choice in such cases is very much affected by their imagination. Now it is much easier to imagine the difference between the two sorts of life which we have considered, than it is to imagine the difference between an enormously long time and another time which is enormously longer. And, again, we are generally affected more than is reasonable by the present or the near

[32] See the following note.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

> future in comparison with the far future. And a change which will only happen after the end of a million years is in a very far future.[33]

Here, McTaggart clearly recognizes the intuitively repugnant implications that additive aggregation can have for lives of sufficient length; he perceptively sees much of what is at stake in the matter; and he offers several sophisticated explanations for why we should not trust our intuitions about such cases. However, while McTaggart's remarks are deeply insightful, ultimately I find his deflationary account of our intuitions about such cases unconvincing. Let me now suggest why.

Consider first his contention that it is "much easier to imagine the difference between the two sorts of life which we have considered, than it is to imagine the difference between an enormously long time and another time which is enormously longer." I readily grant that we may have no clear intuitive grasp on numbers as large as a million, and so it might be argued that our intuitions about such cases do not accurately reflect the difference between a million years and some indefinitely much longer period of time. But though true, it isn't evident how much weight this observation bears.

First, it isn't as if we would have any trouble choosing between a million years in heaven or hell, and some indefinitely much longer period of time in heaven or hell. We would clearly, and rightly, prefer the longer time in heaven, and the million years in hell. Here, our judgment readily guides us to "correct" for our intuitive shortcomings. But in the choice between the brilliant life and the oyster-like life, our judgment does not similarly lead us to "correct" for our intuitive shortcomings; rather, for some of us at least, it leads us to reject additive aggregation and trust our intuitions about such cases.

Second, it isn't as if we can pretend to have an accurate intuitive grasp of the *real* difference between the brilliant life and the oyster-like life. Rather, we just have the firm conviction that the quality of the former life would be *much* higher than that of the latter. Similarly, we can have the firm conviction that the indefinitely long life would be *much* longer than the life of a million years. Moreover, while our judgment might lead us to think that the size of the gap in quality between the different lives must, for finite beings, be naturally limited, the gap in number of years could be as large as one likes; thus, our judgment naturally leads us to recognize that the difference in duration between the two lives might be many times greater in magnitude than the difference in their quality. Yet even as our judgment recognizes this truth, it does not cause us—or at least not me,

[33] I am grateful to Thomas Hurka, who first brought McTaggart's claims to my attention, and then subsequently sent me these passages. They are from J. M. E. McTaggart, *The Nature of Existence,* vol. 2 (Cambridge: Cambridge University Press, 1921), 452–53 (note omitted). (The main passage comprises sections 869–70 of ibid., book VII, chap. 67, "The Total Value in the Universe.")

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

anyway—to revise our judgment about which life would be better. Rather, it causes us to question the additive aggregationist approach for comparing such alternatives, thereby confirming our intuitions, rather than dismissing them as distorting.

Third, while the Single Life Repugnant Conclusion may seem *especially* repugnant in the case McTaggart imagines, where we are to forsake a million years of flourishing for an indefinitely long period of oyster-like life, as indicated above, I believe that most people would make the same judgment if the choice were between a mere hundred years and an indefinitely long period of oyster-like life. Consider seriously the difference between the two lives. As discussed above, the first would involve a hundred years as rich and rewarding as any human life that has ever been lived, filled with great achievements of the highest order, deep and lasting loves, rewarding friendships, and so on. The second would involve eons of a life barely worth living; such a life would be utterly devoid of the achievements and relationships that contribute to human flourishing, and would contain only the lowest and mildest forms of conscious pleasure. Surely, most people would choose the former life for themselves, or for a loved one. But in this case one can hardly explain away the choice as resulting from our inability to "imagine the difference between an enormously long time and another time which is enormously longer." We certainly can imagine that the difference between a hundred years and many eons might be of such a magnitude that, in a sense, the *total amount* of utility would be greater in the longer oyster-like life than in the shorter flourishing life. But what we cannot believe, on reflection, is that the *goodness* of the different lives is a simple additive function of the total amount of utility in those lives. While there would be *more* utility in the oyster-like life with each passing year, we do not believe that the different oyster-like pleasures spread out over many eons *add up* in the way they would need to in order to make the oyster-like life *better* than the flourishing life.

The truth is that we do not believe that there would be a *significant* difference between the goodness of an oyster-like life that lasted a hundred years and one that lasted a thousand years, or ten thousand years. And these are numbers that we certainly can, I believe, intuitively grasp. Moreover, importantly, we do not see the difference between such lives as being on a trajectory that, over time, would eventually amount to a great difference. Correspondingly, we judge that there would not be a significant difference between the goodness of an oyster-like life that lasted a hundred years, and one that lasted a million years, or many eons. I suggest, then, that our view about such cases rests not on our intuitive inability to grasp large numbers, as McTaggart suggests, but on our reflective *judgment* that merely adding more years to an oyster-like life does not significantly increase the goodness, or desirability, of such a life.

To be sure, advocates of additive aggregation may continue to insist that our intuitions about such cases are not to be trusted, even for cases involving intuitively graspable numbers. For example, they might insist that the difference between oyster-like lives of a hundred, a thousand, and ten thousand years *is* on a trajectory that would, eventually, amount to a great difference, but that the slope of the trajectory is so slight that we do not intuitively notice it or perceive its long-term implications. But though it is possible that such a view is correct, it has the air of an untestable article of faith that advocates of additive aggregation are compelled to invoke to explain away our intuitions, and I doubt that many will find it sufficiently compelling to alter their judgments about such cases.

Consider next McTaggart's suggestion that our intuitions about such cases are not to be trusted because "we are generally affected more than is reasonable by the present or the near future in comparison with the far future. And a change which will only happen after the end of a million years is in a very far future." McTaggart is certainly right in his observation, but I seriously doubt that our "bias" toward the near future accounts for our attitude about such cases. Suppose, for example, that people were offered the following choice. In each of two outcomes, people would, starting today, live eons and eons, followed by annihilation. In the first outcome, they would live oyster-like lives the entire time. In the second outcome, they would spend virtually the whole time in "a state whose value [w]ould be neither good nor evil, but zero," perhaps in a state of suspended animation, but in their final hundred years they would live as rich and rewarding a life as is humanly possible. Confronted with such a choice, I have no doubt that most people would clearly prefer the latter, for themselves or a loved one. But if this is right, then the explanation of that choice would have nothing to do with our bias toward the present, or the near future, which in this case would presumably favor the first outcome over the second, since by hypothesis the first outcome would be better for us in the present, and for eons to come.

I submit, then, that McTaggart's attempt to explain away the preference most people would have for an extended flourishing life over a much longer oyster-like life in terms of our imaginative limitations regarding differences between large numbers and our bias toward the present and the near future is unconvincing. However, as should be evident, I believe he is very close to being on target when he suggests that the repugnance "certain moralists" would feel about his view "rests on a conviction that quality is something which is inherently and immeasurably more important than quantity." But the view is not simply that quality is *always* "inherently and immeasurably more important" than number or duration. There are plenty of instances where McTaggart's opponents would permit, and even insist on, certain trade-offs between quality and number or duration. (This is a point I shall return to briefly below.) Rather, what

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

his opponents believe (and his opponents surely include almost everyone, as McTaggart himself recognizes) is that when the differences in quality are *sufficiently* large, as they are between the richest possible human life and an oyster-like life, then as long as the higher-quality life persists for a "sufficient" period of time, it would be better, or more valuable, than any duration of the lower-quality life. Of course, as McTaggart rightly observes, this view is neither "self-evident" nor capable of "demonstration." But despite the rhetorical flourish with which McTaggart makes that observation, his opponents needn't think otherwise. Neither self-evidence nor demonstrability is necessary for truth, much less for reasonable belief or acceptance.

When we compare a flourishing human life of a hundred years with an indefinitely long oyster-like life, we seem to have an example of what James Griffin described as "discontinuity between values," where "enough of A outranks any amount of B."[34] As Griffin puts it, such cases involve "the suspension of addition; . . . we have a positive value that, no matter how often a certain amount is added to itself, cannot become greater than another positive value, and cannot, not because with piling up we get diminishing value or even disvalue . . . but because they are the sort of value that, even remaining constant, cannot add up to some other value."[35] Of course, Griffin is not denying a basic truth of mathematics, that any positive real number, no matter how small, would, if added to itself enough times, eventually be larger than any other positive real number, no matter how large. What he is saying is that the normative significance of certain kinds of things is such that no number of instances of things of that sort has the normative significance of certain other kinds of things.

Consider the following analogy. While a B paper has genuine value, no number of B papers has the kind of value, as a corpus, that some number of A papers might have. Thus, a teacher might well be justified in giving an A to a student who handed in three A papers, but a B to a student who handed in a plethora of B papers. After all, no matter how many papers the latter student handed in, each of which had a genuine value, it remained true that he was doing consistently B work. Clearly, then, the value of a body of work is not reached by simply adding up the individual values of each piece of that work. Similarly, the value of an outcome or a life is not reached by simply adding up the individual values of each element of that outcome or life.

On reflection, then, the overall goodness of a life is not just an additive function of the goodness of each moment of that life, considered separately from the others. Thus, a simple additive aggregationist approach should be rejected within lives as well as between lives.

[34] Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance*, 85.
[35] Ibid.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

VIII. Conclusion

Let me conclude this essay with a summary, and then a brief comment about some issues I did not address and important work that remains to be done.

In Section II, I offered a general account of aggregation, which is the process or method of combining different parts into a whole. I noted that, in ethics, aggregation is often associated with the additive aggregationist approach of classical utilitarianism, but that, as I use the notion, a principle of aggregation is *any* function from a set of relevant elements—whatever those turn out to be—to the assessment of the goodness or comparative desirability of a set of items being assessed, say, lives or outcomes. Among other things, then, a principle of aggregation may accommodate distributive considerations in assessing individual lives or outcomes, to reflect the fact that *how much* good obtains in lives or outcomes may or may not matter as much as how that good is *distributed* within lives or outcomes.

In Section III, I presented two examples illustrating some of the problems of additive aggregation: Parfit's Repugnant Conclusion and my Lollipops for Life Case. Both examples reflect a powerful anti-additive aggregationist position and challenge the plausibility of employing an additive aggregationist approach in the assessment of outcomes. Such examples have been widely discussed as an objection to total utilitarianism. I briefly noted that utilitarians offer a deflationary response to such examples, and indicated that I found the total utilitarian's position admirably consistent but unpersuasive. I also noted that anti-additive aggregationist reasoning has been advocated by philosophers representing a wide range of philosophical positions, including consequentialism, deontology, contractualism, libertarianism, and pluralism.

In Section IV, I noted that Rawls's and Nozick's emphasis on the separateness of persons helped illuminate the inadequacy of classical utilitarianism and, more particularly, helped pave the way for people to readily see the shortcomings of a simple additive aggregationist approach for assessing the value of outcomes. However, I suggested that their discussions may have also obscured the source and scope of anti-additive aggregationist reasoning, by conveying the impression that such reasoning ultimately rested on the foundation of the separateness of persons. More particularly, I suggested that Rawls's and Nozick's emphasis on the significance of the separateness of persons, combined with their failure to significantly challenge the Sidgwickian conception of rational self-interest, may have lent implicit support to the view that additive aggregation is appropriate within lives, despite its being inappropriate between lives. Or, to put the point more cautiously, Rawls's and Nozick's focus on the separateness of persons may help explain why most theorists have not thought to challenge an additive aggregationist approach for assessing

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

the goodness of individual lives, despite the widely recognized shortcomings of such an approach for assessing outcomes.

In Section V, I followed Parfit in suggesting that some who have stressed the importance of the separateness of persons may have conflated the issue of compensation with the issue of moral balancing. Like Parfit, I believe that while the separateness of persons makes it true that compensation is only possible within lives—so that benefits to one cannot ever be *compensation* for another's burdens—this fact is only *relevant* to whether the moral balance of benefits and burdens is desirable; it does not *settle* that question. In particular, I argued that in assessing the relative desirability of two outcomes, one must take account of much more than the brute fact of whether anyone in the different outcomes suffers uncompensated burdens. One must, per force, take account of the total impact that the benefits and burdens have on *all* of the ideals relevant to assessing the desirability of outcomes, including, perhaps, utility, equality, priority, maximin, perfection, freedom, comparative justice, and absolute justice. Similarly, I argued, while the possibility of compensation allows trade-offs between benefits and burdens within a life, there can be many factors relevant to assessing the overall goodness or desirability of a life, besides merely the sum total of the benefits and burdens in that life. Correspondingly, in assessing the desirability of a life one must take account of the total impact that the distribution of benefits and burdens has on *all* of the relevant factors for assessing lives, and not simply add up the benefits and burdens.

In Section VI, I presented a number of examples that directly challenge a simple additive aggregationist approach for assessing the goodness of individual lives. This is the approach underlying Sidgwick's conception of individual self-interest, according to which the best life is the one with the greatest sum total of goods spread throughout the life, where the value of each moment can be determined separately from the others.

I readily grant that facts about personal identity allow for compensation, and therefore license certain trade-offs of benefits for burdens, within lives, that would not, due to the separateness of persons, be permissible between lives. Nevertheless, I argued that within lives, as well as between lives, what matters is not just how much total good obtains, but how that good is distributed. For example, I suggested that a consistently drab life would have a significantly different value than a wildly fluctuating life, even if the sum total of benefits minus burdens was the same in the two lives. Similarly, I suggested that the overall goodness of a life might be affected by its shape, pattern, or direction, beyond the ways in which these might alter the total *sum* of the life's goodness. Thus, for example, I suggested that a life of steady improvement that started out poorly and ended well, could be better than a life of steady decline that started well and ended poorly, even if the sum total of goodness in the two lives were the same, as measured moment by moment. In addition, I argued that a

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

consistently high-quality life would be better than one that was only somewhat better most of the time, but included a year of exceptionally bad pain and suffering, even if the total sum of goodness minus badness was slightly higher in the second life. Similarly, I argued that a consistently high-quality life would be worse than one that was only somewhat worse most of the time, but truly blissful for a full year, even if the sum total of goodness was slightly higher in the first life.

Finally, in Section VII, I presented an analogue of Parfit's Repugnant Conclusion for the case of a single life, and contended that just as most people regard the Repugnant Conclusion as genuinely repugnant, so, too, most people would regard the Single Life Repugnant Conclusion as genuinely repugnant. For example, most people would regard a flourishing human life enduring for a hundred years as more desirable than an oyster-like life of very long duration, even if they admitted that *in a sense* there would be a greater total sum of goodness spread throughout the latter life as compared to the former one. I argued that in holding such a view, most people judge that the goodness that obtains in each moment of an oyster-like life does not add up in the way it would need to in order to make such a life more desirable, overall, than a flourishing human life of significant duration.

Interestingly, the first person to present a version of the Single Life Repugnant Conclusion, J. M. E. McTaggart, recognized that most people would regard it as repugnant, but denied that their view would be correct. But while McTaggart offered interesting and sophisticated considerations for dismissing our intuitions about the case, I found his considerations unpersuasive, and argued that they should be rejected.

Together, this essay's arguments suggest that we should reject Sidwick's conception of individual self-interest. More specifically, they suggest that most people would accept that in certain cases anti-additive aggregationist principles apply within lives, as well as between lives.

Unfortunately, because of the role played by the notions of compensation and the separateness of persons in contemporary thought, the inadequacy of additive aggregation for assessing lives has not, perhaps, been as widely recognized as its inadequacy for assessing outcomes. Nevertheless, I suspect that most people, on reflection, will regard this essay's conclusions as unsurprising. But, unsurprising or not, such results are important, and they raise a host of significant issues that need to be addressed.

One question that needs to be pursued is the exact nature of the anti-additive aggregationist principles that need to be employed in assessing individual lives. It is one thing to see that a simple additive aggregationist approach is not always adequate, and to recognize that one wants to give weight to such factors as the "shape" of a life, the "distribution" of benefits and burdens within a life, and/or the desirability of avoiding great lows or achieving great highs. But one needs to know more than that.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press

What makes one "shape" better or worse than another? What distributional considerations are most relevant? How does one determine when a trade-off in terms of the quantity of goodness in a life is warranted by the quality of the highs or lows to be realized or avoided?

Also, in developing an anti-additive aggregationist account of value, it is important to bear the following in mind. I have claimed that we should make room for anti-additive aggregationist principles *within* a life as well as *between* lives; but though I think the considerations in the two cases are analogous, I do not believe they are identical. In particular, I think that the separateness of persons *is* important, and that it has a bearing on the nature and implications of the anti-additive aggregationist principles that are applicable between lives, distinguishing those principles, in important ways, from those that are applicable within lives.

Finally, although I believe that anti-additive aggregationist principles are applicable within lives in certain contexts, I have not claimed, and do not believe, that they are relevant in all contexts. To the contrary, I believe that there are many cases where we rightly choose between different alternative lives primarily on the basis of which has the greater sum total of benefits and burdens within those lives. So, in essence, I think there are some cases where we *should* appeal to additive aggregationist principles in assessing lives, and some cases where we should *not,* and it is imperative to understand what factors determine which principles are relevant when. Moreover, if I am right that different factors come into play in different contexts, then, as I have argued at length elsewhere, this raises the specter of intransitivity in our "all things considered" judgments.[36] And this raises a host of deep and troubling questions about the very nature of moral ideals and the possibility of making coherent rankings of outcomes. These are big issues that I explore elsewhere, but cannot pursue here.

In some ways, then, I think the main claims of this essay may seem obvious and uncontroversial. But when they are fully developed, I think it will be seen that they raise fundamental questions about the very nature of the good, and the intelligibility of comparing lives, or outcomes, as better and worse. It is easy to see that a simple additive aggregationist approach sometimes has deeply implausible implications when comparing lives or outcomes. Unfortunately, it is much more difficult to come up with an alternative approach that does not have deeply implausible implications of its own.

Philosophy, Rutgers, The State University of New Jersey

[36] See, for example, my essays "A 'New' Principle of Aggregation"; "A Continuum Argument for Intransitivity"; and "Rethinking the Good." See also my "Worries about Continuity, Transitivity, Expected Utility Theory, and Practical Reasoning," in Dan Egonsson, Jonas Josefsson, Björn Petersson, and Toni Rønnow-Rasmussen, eds., *Exploring Practical Philosophy* (Aldershot: Ashgate Publishing Limited, 2001), 95–108; and my "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* 16, no. 2 (1987): 138–87.

https://doi.org/10.1017/S0265052509090013 Published online by Cambridge University Press