

## **A Set of Solutions to Parfit's Problems**

In Part Four of Reasons and Persons Derek Parfit searches for “Theory X,” a satisfactory account of well-being.<sup>1</sup> Theories of well-being cover the utilitarian part of ethics but don't claim to cover everything. They say nothing, for example, about rights or justice. Most of the theories Parfit considers remain neutral on what well-being consists in; his ingenious problems concern form rather than content. In the end, Parfit cannot find a theory that solves each of his problems.

In this essay I propose a theory of well-being that may provide viable solutions. This theory is hedonic—couched in terms of pleasure—but solutions of the same form are available even if hedonic welfare is but one aspect of well-being.

In Section 1, I lay out the “Quasi-Maximizing Theory” of hedonic well-being. I motivate the least intuitive part of the theory in Section 2. Then I consider Jesper Ryberg's objection to a similar theory. In Sections 4-6 I show how the theory purports to solve Parfit's problems. If my arguments succeed, then the Quasi-Maximizing Theory is one of the few viable candidates for Theory X.

### **1. The Quasi-Maximizing Theory**

The Quasi-Maximizing Theory incorporates four principles.

1. The Conflation Principle: One state of affairs is hedonically better than another if and only if one person's having all the experiences in the first would be hedonically better than one person's having all the experiences in the second.<sup>ii</sup>

The Conflation Principle sanctions translating multiperson comparisons into single person comparisons. For example, "Would Cindy and Bill get more pleasure from being at the Braves games than Jim and Carol?" becomes "Would it be hedonically better for one person to have Cindy's and Bill's relevant experiences or Jim's and Carol's?" In this example, don't try to imagine that a single person would have Bill's and Cindy's token experiences; imagine that she would have experiences qualitatively identical to Cindy's and Bill's. Also, don't imagine that she would be bored by having Bill's experiences, if Cindy's were similar; she must have Bill's experiences, not a bored version of them.

The Conflation Principle assumes that one person's having Cindy's and Bill's experiences suffices to capture those experiences' hedonic value. Some philosophers, however, believe that factors extrinsic to experience can affect hedonic value. Pleasurable and painful experiences, some philosophers think, vary in value with: the value of their intentional object<sup>iii</sup> or associated behavior;<sup>iv</sup> the extent to which they are deserved;<sup>v</sup> and (for pleasures) the extent to which they depend on false belief or cognitive error.<sup>vi</sup> Moreover, many philosophers either believe that an experience's being painful consists in its representing bodily damage (or being thought to);<sup>vii</sup> or in its inclining the subject to fight its continuation;<sup>viii</sup> or in its being disliked<sup>ix</sup>—and similarly for an experience's being pleasurable. Such philosophers may revise the Conflation Principle so that not only will the

one person have the others' experiences, but her experiences will have the same intentional objects (etc.) as the others'. These issues are irrelevant to the cases I'll discuss.

The Conflation Principle also assumes that one person can have experiences just like Cindy's and Bill's. That assumption too may be questioned. If I had Cindy's experiences, then Bill's, it might seem as though no single person persisted through these events; it might seem as though my body were inhabited by one person, then another. However, the Conflation Principle may be formulated without that assumption: "One state of affairs is hedonically better than another if and only if one person's having experiences relevantly similar to those in the first would be hedonically better than one person's having experiences relevantly similar to those in the second." On one interpretation, the experiences must be "relevantly similar" just in having pleasures and pains of the same duration and intensity. Some might deem more similarities necessary,<sup>x</sup> but again, these niceties are irrelevant here.

The Conflation Principle is substantive, since multiperson comparisons are often more difficult to make than the corresponding single person comparisons. For example, it is not obvious how two pleasurable lives compare to four somewhat less pleasurable lives. But given Conflation, it is obvious that the four lives are hedonically better. The principle of hedonic assessment for one life is roughly that of maximization. And the Conflation Principle is consistent with maximization. However, the two are distinct. Maximizers, for example, believe that additional moments of pleasure always improve an outcome. On Conflation, whether that is true turns on whether additional moments of pleasure always improve a life, and one needn't think they do.<sup>xi</sup> Moreover, as we shall see, the Quasi-Maximizing Theory, which incorporates Conflation, is not maximizing for a different reason.

2. Lexicality: Pleasures sufficiently different in intensity differ lexically. This means that no finite duration (no matter how long) of the less intense pleasure would be as good for someone as some relatively short duration—say, a month—of the more intense pleasure.<sup>xii</sup>

This principle is supported by the strong preferences of competent judges. Most people who have experienced both kinds of pleasure would greatly prefer a month of ecstasy to any duration of muzak and potatoes. So, those pleasures differ lexically. Note that this is not the distinction between higher and lower pleasures. A pleasure is lexically better than another by virtue of being much more intense, but higher pleasures may be less intense than lower pleasures. Higher pleasures are usually distinguished from lower pleasures by being more “dignified” (to use Francis Hutcheson’s phrase<sup>xiii</sup>) or by falling into a preferred category (for example, being mental rather than bodily).

3. Duration: Someone’s feeling pleasure for time  $t$  is hedonically better than someone’s feeling pleasure which is slightly more intense but lasts 1% of  $t$ .

Duration is also supported by the strong preferences of competent judges. Duration, I think, is obviously true: a slight increase in pleasure-intensity can’t offset a huge loss in duration. Duration is used to support the next principle.

4. Intransitivity: According to Transitivity, if  $x$  is hedonically better than  $y$ , and  $y$  is hedonically better than  $z$ , then  $x$  is hedonically better than  $z$ . Intransitivity is the denial of Transitivity.

Lexicality and Duration entail Intransitivity. The proof goes like this. Duration entails that 1 year of ecstasy (=A) is hedonically worse than 100 years of pleasure slightly less intense (=B); that B is worse than 10,000 years of pleasure slightly less intense (=C); that C is worse than 1,000,000 years of pleasure slightly less intense (=D); and so on to Z, which is  $10^{50}$  years of extremely mild pleasure. Given these premises, Transitivity entails that A is hedonically worse than Z. However, according to Lexicality, A is hedonically better than Z.<sup>xiv</sup>

On maximizing theories, each state of affairs can be given a number representing its hedonic value, and all hedonic comparisons can be made by referring to those numbers. But outcomes cannot accurately be assigned numbers when Transitivity fails; this would require, in the above example, assigning B a higher number than A (since B is better), assigning C a higher number than B (since C is better), and so on, but assigning A a higher number than Z (since A is better). This is impossible, since “being a higher number than” is transitive. The Quasi-Maximizing Theory is therefore not a maximizing theory.

I call the view expressed by 1-4 Quasi-Maximizing: “Maximizing” because of the Conflation Principle, but only “Quasi” because of Intransitivity. The Quasi-Maximizing Theory may be supplemented by welfare principles governing values other than pleasure. However, these principles need not change the structure of the theory, since a Quasi-Maximizer could accept the following principles. (In them, “better than” means “better than in terms of non-hedonic well-being.”)

The Non-Hedonic Conflation Principle: One state of affairs is better than another if and only if one person's living all the lives in the first would be better than one person's living all the lives in the second.

Non-Hedonic Lexicality: Lives worth living that differ sufficiently in non-hedonic quality differ lexically. This means that no finite duration (no matter how long) of the inferior life would be as good for someone as some relatively short duration—say, a month—of the superior life.

Non-Hedonic Duration: Someone's enjoying a life that is worth living for time  $t$  is better than someone's enjoying a life that is slightly better but lasts 1% of  $t$ .

Non-Hedonic Intransitivity: It is not the case that if  $x$  is better than  $y$ , and  $y$  is better than  $z$ , then  $x$  is better than  $z$ .

If the Quasi-Maximizing Theory is true, then perhaps these formally similar principles are true. I won't pursue this issue, except to say that non-hedonists about well-being may still use my strategy for solving Parfit's problems if they accept these principles.

## **2. A Brief Defense of Intransitivity**

Initially, Intransitivity seems wildly implausible. But consider the following.

The Theory Behind Transitivity Might Be False

Transitivity seems obviously true to most philosophers. The idea that value is linear may inform that intuition. Imagine a line drawn on a chalkboard. If all outcomes can be represented by points on that line, with better outcomes represented to the right of worse ones, then being better than would be transitive because being to the right of is transitive for those points. However, this linear view of value is not sacrosanct. Many philosophers already reject it.<sup>xv</sup> Some philosophers, for example, believe that certain outcomes cannot be compared in terms of value.<sup>xvi</sup> Such outcomes couldn't be represented on that line. Also, some believe that certain outcomes are only roughly equal in value—that although neither is worse than the other, only one of them is worse than some third state of affairs.<sup>xvii</sup> Such outcomes, again, could not be accurately plotted on the line.

### Why Transitivity Might Fail

On any better-than relation, if x is better than y, then y must be worse than x, because the factors determining that an X is better than a Y must also determine—without room for maneuver—that Y is worse than X. But with Transitivity, there is room for maneuver. According to Transitivity, if x is better than y, and y is better than z, then x is better than z. Transitivity might fail because the factors determining how some X&Z compare are not identical with the factors determining how X&Y and Y&Z compare. Alternatively, Transitivity might fail because factors determining how some X&Z compare differ in significance in determining how X&Y and Y&Z compare.<sup>xviii</sup> This shows—or strongly suggests—that the denial of Transitivity is coherent.

### Why Transitivity Fails

Above I presented a counterexample to Transitivity consisting of twenty-six judgments (that B is hedonically better than A, that C is better than B, and so on, but that A is better than Z). In making these judgments, the difference in pleasure intensity always matters, but its significance is dramatically greater when we compare A and Z; a small difference in pleasure-intensity may be outweighed by greater duration, but a sufficiently great difference in intensity may not be. So, in this case, a factor determining how A&Z compare differs dramatically in significance from how it figures in comparing A&B, B&C, C&D, and so on. This explains why Transitivity fails.

### The Money-Pump Objection to Intransitivity Fails

Suppose we reject Transitivity in favor of the following thesis: for some possibilities, X is hedonically better than Y, Y is better than Z, but X is worse than Z.<sup>xix</sup> On a variant of the “money-pump” objection,<sup>xx</sup> an informed agent, who holds the thesis and is otherwise rational, would pay a small amount to trade X for Z (since Z is better), then pay a small amount to trade Z for Y (since Y is better), then pay a small amount to trade Y for X (since X is better)—the same X she started with. So, according to this objection, the thesis must be rejected. But the objection fails. The rational agent will not behave like this for exactly the reason why doing so seems irrational: because, from the standpoint of self-interest, one might as well put dollar bills down the garbage disposal. The money-pump objection



assumes that a rational agent would always prefer what is better and act on those preferences, no matter what. But that assumption would be rejected along with Transitivity.

### The Inductive Evidence for Transitivity Can Be Outweighed

Transitivity is inductively supported: we have often noted that an X is better than a Y both of which are better than a Z. But a sufficiently powerful counterexample can outweigh such evidence, and I have tried to give one. The preferences of competent judges, I believe, support Lexicality and Duration more than the inductive evidence supports Transitivity. But I needn't insist on that here; I need only insist that the Quasi-Maximizing Theory shouldn't be faulted at this stage for incorporating Intransitivity.

### If Transitivity Still Seems Necessarily True

To many philosophers, Intransitivity will smack of contradiction, despite my arguments. On their view, one can't sensibly be talking about the better than relation when one says that B is better than A, C is better than B, but A is better than C. These philosophers, however, need not reject Quasi-Maximization, for they may understand "better than" in its principles to denote a different relation [than the better than relation], which is [also] normative and comparative but [is] nontransitive. The Quasi-Maximizing Theory may thus be interpreted to employ a new concept, [X? which is controversial]], rather than an old concept, [X? used controversially]].<sup>xxi</sup> However, I don't favor this interpretation.

### 3. Ryberg's Objection to Combining the Conflation Principle and Lexicality

Jesper Ryberg, in a fine paper, argues against combining what I call the Conflation Principle and Lexicality.<sup>xxii</sup> On Lexicality, a month of ecstasy is better within a single life than any duration of mild pleasure. Ryberg says that even if a lot of ecstasy is better within a single life than any duration of mild pleasure, a little ecstasy isn't; so, conflating lives that each contain a little ecstasy into a superlife containing a lot can distort evaluation. I'll explain in an endnote why I think that Ryberg's argument, which exploits this idea, doesn't work.<sup>xxiii</sup> Here I'll consider another argument, which also uses Ryberg's insight. Consider these two states of affairs:

AA: 10 billion people live for five seconds each, experiencing ecstasy.

ZZ: 10 billion people live for seventy years each, enjoying mild pleasure.

And now the objection to Quasi-Maximization: "Each life in ZZ is hedonically better than each life in AA. So, ZZ is hedonically better than AA. However, Quasi-Maximization entails that AA is hedonically better than ZZ. After all, given Lexicality, the superlife containing all the experiences in AA is hedonically better than the superlife containing all the experiences in ZZ."

The Quasi-Maximizer should respond by denying that each life in ZZ is hedonically better than each life in AA.<sup>xxiv</sup> On this view, ecstasy trumps mild pleasure in the sense that the smallest possible duration of ecstasy is hedonically preferable to any duration of mild

pleasure. What is the smallest possible duration of ecstasy? I'll leave this question mostly open, but please note that the answer needn't be (and presumably isn't) "the smallest unit of time." For an ecstatic experience might consist of mental states which are not themselves ecstatic—for example, a minute of ecstasy would consist in 120 non-ecstatic mental states, if ecstasy can't be as brief as half a second. I will YOU ALREADY HAVE—SEE YOUR FORMULATION OF AA [continue to] assume that one can feel ecstasy over a five-second period.

On Quasi-Maximization, AA is better than ZZ and ecstasy trumps mild pleasure. Do these implications count against the theory? Some people think so. However, I am unsure, for three reasons. First, when I am enjoying great pleasure, or when I just have, I'm amazed at how good it is. This must be many people's experience. Dostoevsky, for example, told a friend:

In certain moments, I experience a joy that is unthinkable under ordinary circumstances, and of which most people have no comprehension. Then, I feel that I am in complete harmony with myself and the whole world, and this feeling is so bright and strong that you could give up ten years for a few seconds of that ecstasy—yes, even your whole life.<sup>xxv</sup>

Even Dostoevsky's eloquence, however, can't evoke in us very vivid memories of our best pleasures. Some joys are indeed "unthinkable under ordinary circumstances." Hence, it is hard to judge under ordinary circumstances whether the belief that ecstasy trumps mild pleasure tells against Quasi-Maximization.

Second, the intuition that ZZ's lives are better than AA's may concern something other than pleasure. I might prefer seventy years of mild pleasure to an ecstatic life of five seconds partly because I don't want to die in five seconds. Our fear of death may interfere with our comparing AA to ZZ merely in terms of pleasure. Also, ZZ may seem better than AA partly because a world of seventy-year lives may include goods which a world of five-second lives would lack, for example, scientific discovery, artistic creation, space exploration,<sup>xxvi</sup> mathematical achievement, friendship, appreciation of the good, love, human cultures, courage, compassion, and so on. Such factors may interfere with our comparing AA to ZZ merely in terms of pleasure.

Third, I am persuaded by the following argument:

1. The value of a month of ecstasy is exhaustively explained by describing each of its 518,000 five-second subperiods. (After all, a period of ecstasy is good because of what it's like, and what it's like consists in what its five-second subperiods are like.)
2. Therefore, if five seconds of ecstasy aren't better than some finite duration  $d$  of mild pleasure, then a month of ecstasy wouldn't be better than  $518,000(d)$  of mild pleasure.
3. Therefore, if five seconds of ecstasy aren't better than some finite duration of mild pleasure, then Lexicality would be false.
4. But Lexicality is true, as evidenced by the strong preferences of competent judges.
5. So, five seconds of ecstasy are better than any finite duration of mild pleasure.

This argument, like so much in this area, is apt to be controversial; but it helps Quasi-Maximizers defend the view that ecstasy trumps mild pleasure.

#### 4. Parfit's Nonparadoxical Problems

Anyone who proposes a theory of well-being must grapple with Parfit's arguments in Part Four of Reasons and Persons. There Parfit remains neutral about what makes life go best; but to test the Quasi-Maximizing Theory, I'll present Parfit's arguments as objections to possible theories of hedonic well-being.

##### The Non-Identity Problem

Thomas Schwartz believes that "what is bad must be bad for someone."<sup>xxvii</sup> But suppose that if a 14 year old conceives now, her child's life would be barely worth living, while if she waits, her child would be much better off.<sup>xxviii</sup> Conceiving now would be bad for no one, since the now-child would not exist if she waited. Thus, Schwartz's theory cannot explain why waiting would be preferable,<sup>xxix</sup> it cannot solve the Non-Identity Problem. On the Quasi-Maximizing Theory—more specifically, on the Conflation Principle—waiting would be hedonically better because one person would be better off having the later child's experiences.

##### The Repugnant Conclusion

Parfit considers many theories that solve the Non-Identity Problem, but each runs afoul of other problems. According to the Impersonal Total Principle, for example,

If other things are equal, the best outcome is the one in which there would be the greatest quantity of whatever makes life worth living.<sup>xxx</sup>

This principle entails the 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.<sup>xxxi</sup>

A total principle of well-being thus entails the repugnant view that enough lives of piddling value can be better, in terms of welfare, than ten billion very good lives.<sup>xxxii</sup>

To test Quasi-Maximization, let's say that the lives in the Repugnant Conclusion are better or worse only in terms of pleasure. Is the Repugnant Conclusion false on Quasi-Maximization? Presumably, someone with a "very high [hedonic] quality of life" will experience at least a minute of intense pleasure; if so, then ten billion highly pleasurable lives will include at least 19,000 years of intense pleasure.<sup>xxxiii</sup> Given Lexicality, we may safely say that the superlife including those 19,000 years would be hedonically better than any life of merely mild pleasure.

According to the Average Principle, "it is worse if there is a lower average quality of life, per life lived."<sup>xxxiv</sup> On this principle, the Repugnant Conclusion is false, for the lives barely worth living are worse, on average, than those of very high quality.<sup>xxxv</sup> However, the

Average Principle is untenable: it entails, for instance, that a population in agony would be improved in terms of welfare if people were born whose agonies were just slightly less.<sup>xxxvi</sup>

The Quasi-Maximizing Theory, of course, would not consider that an improvement.

### The Absurd Conclusion

On the Average Principle, only “quality” or average well-being matters; on the Total Principle, only “quantity” or total well-being matters. Each view is too extreme. The best theory, Parfit thinks, will value both.<sup>xxxvii</sup> But merely valuing both, he recognizes, won’t avoid the Repugnant Conclusion; for even if its smaller, better-off population gets points for quality, so long as no limit is placed on quantitative value, enough lives barely worth living would be better overall.<sup>xxxviii</sup> So, Parfit considers valuing quality and quantity but limiting how much good lives can contribute to a world’s quantitative value. On the view he considers, good lives within any century cannot have more quantitative value than that of ten billion very good lives.<sup>xxxix</sup> On this view, the two populations in the Repugnant Conclusion would be of equal quantitative value, but the smaller population would be better overall because of its greater qualitative value.<sup>xl</sup>

Parfit further develops the view: “It would always be bad if an extra person has to endure extreme agony. And this would be just as bad, however many others have similar lives.”<sup>xli</sup> So, thinks Parfit, even if we limit how much lives worth living can contribute to a world’s quantitative value, we shouldn’t limit how much disvalue horrible lives can contribute. Given this asymmetry, he asks us to consider these two future populations:

D includes: (i) Earthlings like Earth's present population; and (ii) vastly many people living concurrently, all of whom have a very high quality of life, except that one person in each group of ten billion has a painful disease that makes life not worth living.

E includes: same as D, except that each group of ten billion in (ii) exists in a different future century.

On this view, D would be very bad, since only ten billion of its fortunate denizens improve it, while each person with a painful disease worsens it—and there may be trillions of them. E, however, would be very good, “even though, in both outcomes, there would be the very same number of extra future people, with the same very high quality of life for all except the unfortunate one in each ten billion.”<sup>xlii</sup> This is the Absurd Conclusion.

Let's extend the Conflation Principle as follows to cover outcomes that are equally good in terms of hedonic value: “Two states of affairs have equal hedonic value if and only if one person's having the experiences of one would be neither better nor worse hedonically than one person's having the experiences of the other.” On this extension, the two outcomes in the Absurd Conclusion are equally good; and so, by appealing to it, Quasi-Maximizers may reject the Absurd Conclusion.

### Variants of the Absurd and Repugnant Conclusions

Parfit considers two more ways to avoid the Repugnant Conclusion. Each denies that there is a single scale of value. On the Appeal to the Valueless Level, lives barely worth



living cannot be as good as ten billion blissful lives because lives below a certain level may have personal value but no moral value.<sup>xliii</sup> On the Lexical View, lives below the Mediocre Level always improve a state of affairs, but no number of them are as good as one life above the Blissful Level. Parfit criticizes these views with similar arguments; I will focus on the Lexical View, which Lexicality entails, if we stipulate that lives above the Blissful Level contain at least a month of ecstatic pleasure, while Mediocre lives contain only pleasures that are lexically worse. Let's stipulate that, so Parfit's arguments will apply to the Quasi-Maximizing Theory.

The Lexical View, Parfit says, entails an unacceptable variant of the Absurd Conclusion:

(A) Suppose that, in some history of the future, there would always be an enormous number of people, and for each one person who suffers, and has a life that is not worth living, there would be ten billion people whose lives *are* worth living, though their quality of life is not quite as high as the Mediocre Level. This would be worse than if there were no future people.<sup>xliv</sup>

The conclusion of (A) is too strong; a theory of well-being cannot entail that one outcome is worse simpliciter than another. Here the claim is that an outcome P with persons is worse than an outcome W without persons. But this doesn't follow from P's being bad in terms of well-being, since P's population might bring with it goods outside the domain of well-being. For example, P might include love, courage, artistic creation, human cultures, scientific discovery, and so on, which might have value independently of well-being. P might be better

than W because of such goods despite being worse in terms of well-being. (A) should conclude, “This would be worse in terms of well-being than if there were no future people.”

In what follows, I will assume that (A) has been revised in that way.

How does the Lexical View entail (A)? Parfit says,

The existence of ten billion people below [the Mediocre Level] would have less value than that of a single person above the Blissful Level. If the existence of these people would have less value than that of only one such person, its value would be more than outweighed by the existence of one person who suffers, and has a life that is not worth living.<sup>xlv</sup>

So, if any number of Mediocre Lives are worse than one Blissful Life, then any number of Mediocre Lives should be outweighed by “one person who suffers, and has a life that is not worth living.” For this argument to succeed, this bad life must be very bad; it must be the unfortunate analogue of one Blissful Life. Such a life is worse than the phrase “who suffers and has a life not worth living” suggests; suffering might be compensated, at least partly, by the good things in life, while “a life not worth living” might connote a life barely worse than neutral. If the one life is *very* bad, then the Lexical View does indeed entail that it could outweigh the ten billion Mediocre Lives in terms of well-being. However, I don’t find this implication absurd; I readily accept it.

“On the Lexical View,” says Parfit, “when we consider lives above the Mediocre Level, quantity could always outweigh quality.”<sup>xlvi</sup> Is this true on Quasi-Maximization—would some number of the best lives be hedonically worse than, say, a hundred times as

many lives barely above the Mediocre Level? Let's now interpret the Mediocre Level so that they would be. Mediocre-plus experiences, let's say, are good enough that having a long enough duration of them would be better for one person than having a shorter duration of very intense pleasure. So, Parfit says, the Lexical View entails a variant of the Repugnant Conclusion:

(R) If there were ten billion people living, all with a very high quality of life, there must be some much larger imaginable population whose existence would be better, even though its members have lives that are barely above the Mediocre Level.<sup>xlvi</sup>

But why is (R) repugnant? The Average Principle entails its denial; (R) seems less repugnant once one internalizes the reasons for rejecting that principle. And even if average well-being deserves some weight, enough Mediocre-plus lives should outweigh that influence.

Perhaps (R) seems false because each Blissful life is better than each Mediocre-plus life. One might reason as follows: "Behind a veil of ignorance, I would rationally prefer a Blissful world to a Mediocre-plus world, given that I will be someone and would rather be Blissful than just above Mediocre." But this decision procedure, as Parfit points out, entails the absurdity that Hell One is worse than Hell Two. In Hell One, ten people suffer great agony for fifty years. In Hell Two, ten million people suffer great agony for fifty years minus a day.<sup>xlvi</sup> So that procedure cannot be trusted.

Now I'll offer two arguments for (R), bolstering my contention that (R) is not repugnant.

Lives above the Bad Mediocre Level, let's say, always contribute disvalue to a state of affairs, but no number of them are as bad as one Agonizing life (that contains, say, at least a month of torture). Now consider a principle like (R):

(R\*) If there were ten billion people living, all with Agonizing lives, there must be some much larger imaginable population whose existence would be worse, even though its members have lives that are barely below the Bad Mediocre Level.

The Bad Mediocre Level, let's say, is characterized by experiences such that having those experiences for vastly long would be about as bad as having the experiences of ten billion Agonizing lives. Lives at the Bad Mediocre Level, I take it, are very painful; so, it seems plausible that  $10^{500}$  lives below that level would be worse than ten billion Agonizing lives. (R\*), therefore, is plausible. And insofar as (R\*) is plausible, (R) should seem acceptable.

(R\*) involves pain, while (R) involves pleasure. Painful experiences and pleasurable experiences are undeniably similar in some ways: both are conscious entities; both are normally known by introspection; both have normative significance because of what they're like. Nevertheless, many ethicists think they're significantly disanalogous. Should these philosophers reject the inference from (R\*) to (R)? I think not. Such ethicists typically stress ways in which pain is more significant than pleasure: our obligation to stop suffering, they may say, is greater than our obligation to increase pleasure; or, suffering is a greater evil than ecstasy is a good.<sup>xlix</sup> But to compare (R\*) to (R) is not to compare pain's moral import to pleasure's; rather, it is to compare the relative value of painful lives to the relative value of pleasurable lives. The claim is that enough very painful lives are worse than ten billion

Agonizing lives, and so enough pleasurable lives are better than ten billion Blissful lives.

The intuition that pain is more important than pleasure has no place here.

Here is the second argument for (R):

(1) The following two worlds are of equal hedonic value: (M) a world containing ten billion Blissful people, all of whom live on Earth; (N) a world containing ten billion Blissful Earthlings and a great many additional people in distant galaxies whose lives are hedonically neutral.

(2) The N-world would be hedonically improved if all its inhabitants became Mediocre-plus people.

(C) The Mediocre-plus world is hedonically better than the M-world. In other words, (R) is true.

This argument has the form: m and n are of equal hedonic value; k is hedonically better than n; so, k is hedonically better than m. On my view, such premises don't conceptually entail the conclusion, since factors determining how an M&N and K&N compare may differ (or differ in significance) from factors determining how K&M compare. However, the premises inductively support the conclusion, for we have often observed that, when one item is hedonically better than another, it is also hedonically better than an item equal in value to the other. These remarks parallel points I made about Transitivity.

(1) may be denied on the grounds that (N) is worse than (M) due to (N)'s inequality.

But, first, note that (N)'s inequality doesn't arise from social injustice, nor could it be lessened through redistribution. Hence, it is far from clear that (N)'s inequality is bad.

Second, note that (M) and (N) are merely said to be of equal hedonic value. Is inequality relevant to hedonic comparisons? That depends on the nature of inequality as a value. Larry Temkin says that “it is bad, unfair or unjust, for some to be worse off than others through no fault of their own . . .”<sup>i</sup> If (N)’s inequality is said to be unfair to those not living on Earth, or unjust, this is irrelevant to (1), since justice and fairness lie outside the scope of hedonic value. However, one might say: “The distribution of pleasure in (N) is not unfair or unjust, it’s just bad—bad in terms of pleasure. An outcome is hedonically worse if it contains less pleasure or if its pleasures are less evenly distributed.”<sup>ii</sup> So, (1) is false; (N) is hedonically worse than (M).” I find this view implausible, but I have nothing more to say about it.

Is (2) true? The change described in (2) is better in terms of both total and average hedonic well-being.<sup>iii</sup> Hence, (2) is plausible.

In light of all these points, (R) is acceptable, if not true.

## 5. The Mere Addition Paradox

Mere Addition is “when, in one of two outcomes, there exist extra people (1) who have lives worth living, (2) who affect no one else, and (3) whose existence does not involve social injustice.”<sup>liii</sup> The Mere Addition Paradox arises for these three states of affairs:

- A: 5 billion people, all of whom have a very high quality of life.
- B: 10 billion people whose lives are about four-fifths as good as the lives of the people in A.

A+: The 5 billion A-people and, by Mere Addition, 5 billion people whose lives are worth living though considerably worse than the lives of the A-people. The average quality of life is lower in A+ than in B.<sup>liv</sup>

According to Parfit, A is better than B, B is better than A+, but A is not better than A+. The paradox, on Parfit's view, arises because these beliefs are inconsistent with the transitivity of being better than.

Quasi-Maximizers reject Transitivity in favor of Lexicality and Duration. This might seem to solve the hedonic version of the problem. However, Lexicality and Duration are not involved in generating the Mere Addition Paradox; moreover, inductive evidence suggests that counterexamples to Transitivity are rare. I would put the paradox as follows: "These three beliefs—A is hedonically better than B, B is hedonically better than A+, and A is not hedonically better than A+—entail a violation of Transitivity; so they need support and explanation." Temkin, indeed, defends principles suggesting why this set of outcomes (A, B and A+) might violate the transitivity of being better than.<sup>lv</sup>

However, this issue doesn't arise on the Quasi-Maximizing Theory, which rejects Parfit's thesis that A is better than B. On the Conflation Principle, B is hedonically better than A, for having the experiences of the ten billion people would be better for one person than having the experiences of the five billion people whose lives are better by 25%. This thesis is like (R), according to which some number of Mediocre-plus lives are better than ten billion Blissful lives. However, this thesis may be more plausible than (R), since the B-lives are better than the Mediocre-plus lives. I defended (R) with two arguments. The idea that B is better than A could be defended with two similar arguments.

Parfit argues, against this view, that “B is better than A” leads to the Repugnant Conclusion:

There is a possible outcome C whose relation to B is just like B’s relation to A.  
 In C there are twice as many people, who are all worse off than everyone in B. . . .  
 If we conclude that B is better than A, we must conclude that C is better than B.  
 On the same argument, D would be better than C, E better than D, and so on  
 down the Alphabet. The best outcome would be Z: an enormous population all of  
 whom have lives that are barely worth living.<sup>lvi</sup>

On Quasi-Maximization—more specifically, on Conflation—C is hedonically better than B, D is hedonically better than C, and so on. However, Z is not better than A; Z is not best.<sup>lvii</sup>  
 In fact, no state of affairs among A-Z is best; no state of affairs is hedonically better than each of the others. These outcomes violate Transitivity because the A-pleasures are lexically better than the Z-pleasures, but the path from A to Z involves only changes for the better.

## 6. The Second Paradox

The Second Paradox, a descendant of the Mere Addition Paradox, may be Parfit’s most complicated argument.<sup>lviii</sup> For this reason, perhaps, it has gone largely unstudied. This is a shame. It is one of the great achievements of twentieth century ethics.

The Second Paradox offers up a series of possibilities that seem to get better and better, yet the last seems worse than the first. I’ll present the paradox so that the possibilities



seem to get hedonically better and better, yet the last seems hedonically worse than the first. The paradoxical conclusion, derived with Transitivity, is that the last possibility is hedonically better than the first. Parfit does not reject Transitivity; he tries to resolve the paradox differently. I will criticize Parfit's and Temkin's resolutions before suggesting that the paradox supports Intransitivity, the least plausible pillar of Quasi-Maximization.

### How the Second Paradox Goes

The first state of affairs is A+. (See the diagram.) A+ contains two groups of 10 billion people: one whose lives are at '100,' an ecstatic level, and another whose lives are at '50,' a level of pleasure well worth enjoying. The last outcome is Omega 100, a world that contains many, many lives each of which is barely worth living at each moment. In Omega 100 muzak and potatoes provide the only pleasures in life. Although A+ is hedonically better than Omega 100, A+ is transformed into Omega 100 via changes for the better.

Each change from A+ to Omega 100 takes one of two forms.

The first kind of change occurs as A+ becomes Alpha. This happens by raising both groups in A+ to a 105 level of pleasure and adding many, many groups of 10 billion people whose lives, at 45, are well worth living. Hedonic welfare is much improved going from A+ to Alpha because all the people in A+ hedonically benefit from the change, especially those in the 50 group, and the only "cost" of this benefit is adding people to the world who are glad to be alive.

The second kind of change occurs as Alpha is transformed into Beta. This occurs by lowering the two better-off groups in Alpha from 105 to 104 but raising as many worse-off

groups from 45 to 104. (Even after this change many groups are at 45.) This kind of change occurs down the Greek alphabet until we reach Omega. In Omega, many groups are at 90 but many more are still at 45.

Omega is transformed into Alpha 2 by improving all the lives in Omega to 95 (including the lives that were at 45) and adding many more groups at 40. This repeats the first sort of change. Alpha 2 is transformed into Beta 2 by lowering the better-off groups to 94 but raising the same number of worse-off groups to 94.



This repeats the second sort of change. By the time we reach Omega 2, the better-off groups are down to 80, though there are many more of them, while many groups are still at 40. At Alpha 3 all the people in Omega 2 are promoted to the level of 85 and many groups at 35 are added.

So at each Omega the average quality of life is lower than it was at the previous Omega, and the population has been greatly increased. At Omega 100, everyone's life is barely worth living at each moment. We want to say both that Omega 100 is hedonically worse than A+ and that each change from A+ to Omega 100 is hedonically for the better. Each change seems for the better because the quality of life is lowered only for those who are better off, and then only when this loss is more than offset by gains for the worse-off.

The Second Paradox may also be formulated in terms of painful lives. To do this, change the numbers in the Second Paradox to negative numbers; then the worlds keep getting hedonically worse, even though the last is hedonically better than the first. This version of the Second Paradox may be more difficult to resolve than Parfit's, if one refuses to let go of Transitivity. However, I won't pursue that argument here.

### Parfit's Suggestion

Parfit tries to resolve the Second Paradox by claiming that Alpha is better than Beta: 20 billion people at 105 (plus many more at 45) is better than 40 billion people at 104 (plus many more at 45, though 20 billion fewer than in Alpha). To defend this he appeals to "Perfectionism." Perfectionism is the view that "even if some change brings a great net benefit to those who are affected, it is a change for the worse if it involves the loss of one of

the best things in life.”<sup>lix</sup> In Alpha, Parfit says, the luckiest 20 billion listen to Mozart; but in Beta, 40 billion listen only to Haydn. Parfit might amplify his view with any of the following claims:

- (a) Mozart’s music has incomparably more intrinsic value than Haydn’s music.
- (b) The full appreciation of Mozart’s music has value apart from its effect on well-being. This value is incomparably greater than any similar value fully appreciating Haydn’s music might have.
- (c) The full appreciation of Mozart’s music contributes incomparably more to the appreciator’s well-being than would a full appreciation of Haydn’s music. This is not because Mozart’s music affords more intense pleasure, but because Mozart’s music is especially excellent, compared to Haydn’s.
- (d) The most intense pleasures contribute incomparably more to well-being than pleasures slightly less intense.

Are these true? We needn’t assess (a)-(c), for we may interpret the Second Paradox just in terms of pleasure: nothing like Mozart’s music exists in Alpha, let’s say; only pleasure-intensity is lost in the move to Beta. Although (a)-(c) cannot resolve this variant of the paradox, (d) could. And Parfit accepts something like (d).<sup>lx</sup> Indeed, the most intense pleasures are “one of the best things in life,” so Perfectionists should hold that even if some change brings great benefits, it is for the worse if the best pleasures are lost.

But in this respect, Perfectionism is wildly implausible. Perfectionism entails that a brief taste of the best pleasure is better than a very long duration of pleasure very slightly less

intense.<sup>lxi</sup> Perfectionism thus conflicts with Duration, according to which someone's feeling pleasure is better than someone's feeling a much briefer, slightly more intense pleasure.

Duration, as I said, is strongly supported by the preferences of competent judges. And Parfit himself says, "[Perfectionism] conflicts with the preferences that most of us would have about our own futures."<sup>lxii</sup>

Moreover, resolving an ambiguity may make (d) even less plausible. According to (d), the most intense pleasures contribute incomparably more to well-being than pleasures slightly less intense. What is the modal strength of "the most intense pleasures?" (A similar question arises for "the best things in life" in Perfectionism.) Most intuitively appealing are "the most intense possible pleasures" and "the most intense actual pleasures." But on the former interpretation, Perfectionism cannot resolve the variant of the paradox in which Alpha's pleasures are ecstatic but not the best possible. And on the latter, a brief taste of mild pleasure would be better than a very long duration of pleasure very slightly less intense, if the best actual pleasures are mild—a type of view Parfit rejects.<sup>lxiii</sup> Perhaps (d) is best interpreted as, "The most intense actual pleasures, provided they are sufficiently intense, contribute incomparably more to well-being than pleasures slightly less intense." But now (d) has lost its intuitive ring.

Parfit says that Perfectionism sometimes strikes him as "crazy."<sup>lxiv</sup> Why, then, does he advance it? Perfectionism, Parfit thinks, is the best of a bad lot. He asks rhetorically, "without Perfectionism how can we avoid the Repugnant Conclusion?"<sup>lxv</sup> This essay provides one answer.

### Temkin's Suggestion

Temkin suggests that many, if not all, of the steps in the Second Paradox may be worse in terms of inequality. Consider A+ and Alpha. A+ contains 10 billion people at 100 and 10 billion at 50; Alpha contains 20 billion at 105 and many more at 45. Since many more worse-off people exist in Alpha than in A+, Alpha may be worse in terms of inequality. Perhaps this outweighs other factors and Alpha is worse overall than A+. Or consider the comparison between Alpha and Beta. In Beta, 40 billion people are at 104 (plus many more at 45, though 20 billion fewer than in Alpha). The inequality between the better-off and worse-off groups may be worse in Beta because Beta has even more better-off people that the worse-off people could resent. And again, perhaps this licenses the judgment that Alpha is better overall than Beta.<sup>lxvi</sup> Thus, Temkin tackles variants of the Second Paradox in which each step is claimed to be for the better simpliciter. But, as I have interpreted the paradox, each move is only claimed to be better hedonically. And, as I suggested in Section 4, even if we think that more equal distributions of pleasure are intrinsically better (which we may not), equality is most plausibly thought to concern justice or fairness, not hedonic value. Hence, Temkin's proposal doesn't help resolve the paradox.<sup>lxvii</sup>

### The Quasi-Maximizing Proposal

On the Quasi-Maximizing Theory, the states of affairs in the Second Paradox get better and better hedonically and Omega 100 is worse than A+. The pleasures in A+ are lexically better than the pleasures in Omega 100, and so A+ is hedonically better than Omega

100. Parfit finds a path from A+ to Omega 100 involving only changes for the better; in doing so, he has produced an ingenious counterexample to Transitivity.

## 7. Conclusion

After a long and inventive, but ultimately unsuccessful, search for a satisfactory theory of well-being, Parfit despairs that

With more unsolved problems, we are further away from the Unified Theory. We are further away from the theory that resolves our disagreements, and that, because it achieves these aims, might deserve to be called the truth.<sup>lxviii</sup>

But having more unsolved problems needn't take us further away from the truth; solving one problem can point the way to solving the others and thus making theoretical progress. I believe that is the case here. The Second Paradox strongly suggests that hedonic betterness is not transitive; this allows us to affirm that B is better than A in the Mere Addition Paradox without entailing the Repugnant Conclusion. And if B is better than A—if ten billion lives are hedonically better than five billion lives of somewhat higher quality, as the Conflation Principle implies—then duration can swamp intensity when the difference in intensity is small; while the Repugnant Conclusion suggests that intensity swamps duration when the difference in intensity is large. All this points to something like a theory that includes Intransitivity, the Conflation Principle, Duration, and Lexicality.

In his Concluding Chapter, Parfit says,



As I argued, we need a new theory about [well-being]. This must solve the Non-Identity Problem, avoid the Repugnant and Absurd Conclusions, and solve the Mere Addition Paradox. I failed to find a theory that can meet these four requirements. Though I failed to find such a theory, I believe that, if they tried, others could succeed.<sup>lxix</sup>

The Quasi-Maximizing Theory, I've argued, provides defensible solutions to Parfit's problems. Many philosophers, however, are firmly wedded to Transitivity. What should they think? Transitivity is incompatible with Lexicality and Duration. According to Duration, someone's feeling pleasure for time  $t$  is hedonically better than someone's feeling pleasure which is slightly more intense but lasts 1% of  $t$ . Duration can hardly be denied, but perhaps Lexicality is false. Perhaps a long enough duration of mild pleasure is preferable to a month, or even to a trillion months, of ecstasy. If so, then the Impersonal Total Principle of well-being can solve Parfit's problems. On this principle, the Non-Identity Problem doesn't arise; the Repugnant Conclusion is true; the Absurd Conclusion doesn't follow; the Mere Addition Paradox is resolved by affirming that  $A$  is worse than  $B$  and then—in response to Parfit's objection—by embracing the “repugnant” conclusion that  $A$  is worse than  $Z$ ; and the Second Paradox is resolved by affirming another supposedly repugnant conclusion, that  $A+$  is worse than Omega 100. Sikora, Anglin, Ng, Attfield, Ryberg, Norcross, Fotion, and Tannsjo defend the Repugnant Conclusion, so perhaps they would favor this set of solutions.<sup>lxx</sup>

I haven't, of course, shown that the Quasi-Maximizing Theory is completely adequate—that it is Theory X—but its success in dealing with Parfit's problems shows that we should explore it further.

Stuart Rachels

University of Alabama

#### References

Anglin, Bill. (1977) "The Repugnant Conclusion," Canadian Journal of Philosophy,

Vol. VII, No. 4, December, pp. 745-754.

Aristotle. (1980) The Nicomachean Ethics, David Ross, trans., revised by J. L. Akrill and J. O. Urmson (Oxford University Press).

Armstrong, D. M. (1968) A Materialist Theory of Mind (London: Routledge & Kegan Paul).

Attfield, Robin. (1991) The Ethics of Environmental Concern (2nd edition) (Athens, Ga.: The University of Georgia Press).

Baier, Kurt. (1958) The Moral Point of View (Ithaca: Cornell University Press).

Brandt, Richard. (1979) A Theory of the Good and the Right (Oxford University Press).

Brentano, Franz. (1973; originally published in 1952) The Foundation and Construction of Ethics, Franziska Mayer-Hillebrand, ed. and Elizabeth Schneewind, trans. (London: Routledge and Kegan Paul).

- Brink, David O. (1997) "Rational Egoism and the Separateness of Persons," Reading Parfit, Jonathan Dancy, ed. (Oxford: Blackwell Publishers).
- Broad, C. D. (1930) Five Types of Ethical Theory (London: Kegan Paul, Trench, Trubner & Co., Ltd.).
- Carson, Thomas L. (1981) "Happiness, Contentment and the Good Life," Pacific Philosophical Quarterly 62, pp. 378-392.
- Chisholm, Roderick M. (1986) Brentano and Intrinsic Value (Cambridge University Press).
- Churchland, Paul M. (1984) Matter and Consciousness (Cambridge, Mass.: The MIT Press).
- Cowen, Tyler. (1996) "What Do We Learn From the Repugnant Conclusion?" Ethics 106, July, pp. 754-775.
- Davidson, Donald; McKinsey, J. C. C.; and Suppes, Patrick. (1955) "Outlines of a Formal Theory of Value, I," Philosophy of Science 22, pp. 140-160.
- Feldman, Fred. (1995) "Justice, desert, and the Repugnant Conclusion," Utilitas 7, 2 (November), pp. 109-206, reprinted in Utilitarianism, Hedonism and Desert (Cambridge University Press, 1997).
- Fotion, Nick. (1997) "Repugnant Thoughts About the Repugnant Conclusion Argument," Contingent Future Persons, Nick Fotion and Jan C. Heller, eds. (Dordrecht: Kluwer Academic Publishers).
- Gallie, W. B. (1954) "Pleasure," The Aristotelian Society, Supplementary Volume, pp. 147-164.

- Griffin, James. (1979) "Is Unhappiness Morally More Important Than Happiness?" Philosophical Quarterly 29, pp. 47-55.
- Griffin, James. (1986) Well-Being: Its Meaning, Measurement and Moral Importance (Oxford University Press).
- Haezrahi, Pepita. (1960) "Pain and Pleasure: Some Reflections on Susan Stebbing's View That Pain and Pleasure Are Moral Values," Philosophical Studies, Vol. 11, pp. 71-78.
- Hall, Richard J. (1989) "Are Pains Necessarily Unpleasant?" Philosophy and Phenomenological Research Vol. XLIX, No. 4, June. pp. 643-659.
- Hare, R. M. (1963) Freedom and Reason (Oxford: The Clarendon Press).
- Hutcheson, Francis. (1897) A System of Moral Philosophy, in British Moralists, Vol. 1, L. A. Selby-Bigge, ed. (Oxford University Press).
- Kant, Immanuel. (1983; originally 1785) Grounding for the Metaphysics of Morals, Immanuel Kant: Ethical Philosophy, James W. Ellington, trans. (Indianapolis: Hackett Publishing Company).
- Kjetsaa, Geir. (1987) Fyodor Dostoevsky: A Writer's Life, trans. by Siri Hustvedt and David McDuff (New York: Viking).
- Korsgaard, Christine M. (1996) The Sources of Normativity (Cambridge University Press).
- Kuhn, Thomas. (1970) The Structure of Scientific Revolutions, second edition (The University of Chicago Press).
- Lemos, Noah. (1994) Intrinsic Value: Concept and Warrant (Cambridge University Press).

- Lewis, C. I. (1946) An Analysis of Knowledge and Valuation (La Salle, Ill.: Open Court Publishing Company).
- McTaggart, J. M. E. (1927) The Nature of Existence (Cambridge University Press).
- Moore, G. E. (1903) Principia Ethica (Cambridge University Press).
- Nagel, Thomas. (1979) "The Fragmentation of Value," reprinted in Mortal Questions (Cambridge University Press).
- Nelkin, Norton. (1994) "Reconsidering Pain," Philosophical Psychology, Vol. 7, No. 3, pp. 325-343.
- Ng, Yew-Kwang. (1990) "Welfarism and Utilitarianism: A Rehabilitation," Utilitas 2, pp. 171-193.
- Norcross, Alastair. (1997) "Comparing Harms: Headaches and Human Lives," Philosophy and Public Affairs, Vol. 26, No. 2 (Spring), pp. 135-167.
- O'Neill, John. (1992) "The Varieties of Intrinsic Value," The Monist, April, Vol. 75, No. 2, pp. 119-137.
- Parfit, Derek. (1984) Reasons and Persons (Oxford University Press).
- Parfit, Derek. (1986) "Overpopulation and the Quality of Life," Applied Ethics, Peter Singer, ed. (Oxford University Press).
- Persson, Ingmar. (1997) "Person-Affecting Principles and Beyond," Contingent Future Persons, Nick Fotion and Jan C. Heller, eds. (Dordrecht: Kluwer Academic Publishers).
- Pitcher, George. (1970) "Pain Perception," The Philosophical Review, July, pp. 368-393.
- Plato. (1961) Republic, Paul Shorey, trans., Plato: The Collected Dialogues, Edith Hamilton and Huntington Cairns, eds. (Princeton University Press).

- Rachels, Stuart. (1993) A Theory of Beneficence (unpublished undergraduate thesis, University of Oxford).
- Rachels, Stuart. (1998) "Counterexamples to the Transitivity of Better Than," Australasian Journal of Philosophy 76, pp. 71-83.
- Rawls, John. (1971) A Theory of Justice (Cambridge, Mass.: Harvard University Press).
- Raz, Joseph. (1986) The Morality of Freedom (Oxford University Press).
- Ross, W. D. (1930) The Right and the Good (Oxford University Press).
- Ryberg, Jesper. (1996a) "Parfit's Repugnant Conclusion," The Philosophical Quarterly, Vol. 46, No. 183 (April), pp. 202-213.
- Ryberg, Jesper. (1996b) "Is the Repugnant Conclusion Repugnant?" Philosophical Papers, Vol. XXV, No. 3, pp. 161-177.
- Ryle, Gilbert. (1958) Dilemmas (Cambridge University Press).
- Schwartz, Thomas. (1978) "Obligations to Posterity," R. I. Sikora and Brian Barry, eds. Obligations to Future Generations (Philadelphia: Temple University Press).
- Schwartz, Thomas. (1979) "Welfare Judgments and Future Generations," Theory and Decision 11, pp. 181-194.
- Sikora, R. I. (1975) "Utilitarianism: The Classical Principle and the Average Principle," Canadian Journal of Philosophy 5, pp. 409-419.
- Sikora, R. I. (1981) "Classical Utilitarianism and Parfit's Repugnant Conclusion: A Reply to McMahan," Ethics 92 (October), pp. 128-133.
- Smart, J. J. C. (1973) "An outline of a system of utilitarian ethics," Utilitarianism for and against (with Bernard Williams) (Cambridge University Press).

- Stearns, J. Brenton. (1972) "Ecology and the Indefinite Unborn," The Monist, Vol. 56, No. 4 (October), pp. 612-625.
- Tannsjo, Torbjorn. (1998) Hedonistic Utilitarianism (Edinburgh University Press).
- Temkin, Larry S. (1987) "Intransitivity and the Mere Addition Paradox," Philosophy and Public Affairs, Vol. 16, No. 2, Spring, pp. 138-187.
- Temkin, Larry S. (1993) Inequality (Oxford University Press).
- Temkin, Larry S. (1994) "Weighing Goods: Some Questions and Comments," Philosophy and Public Affairs 23, pp. 350-380.
- Temkin, Larry S. (1996) "A Continuum Argument for Intransitivity," Philosophy and Public Affairs 25, pp. 175-210.
- Tye, Michael. (1995) Ten Problems of Consciousness (Cambridge, Mass.: The MIT Press).
- Wilkes, K. (1977) Physicalism (London: Routledge & Kegan Paul).
- Zimmerman, Michael J. (1980-81) "On the Intrinsic Value of States of Pleasure," Philosophy and Phenomenological Research 41, pp. 26-45.

---

<sup>i</sup> Parfit 1984. Part Four is "Future Generations," pp. 351-441. "Theory X" is introduced on p. 366. Parfit uses "well-being" and "beneficence" interchangeably, but I will stick to "well-being," since "beneficence" connotes charity, which is not Parfit's topic (or mine). Also, Parfit sometimes says "human well-being" (pp. 370, 393, 394), but a satisfactory theory of well-being should also apply to nonhuman animals, so I omit that qualification.

<sup>ii</sup> For similar ideas, see Lewis 1946, pp. 546-547; Hare 1963, p. 123; and Smart 1973, p. 26.

I restrict the Conflation Principle to states of affairs containing finite durations of experience.

---

<sup>iii</sup> If I take sadistic pleasure in the belief that someone is suffering, that belief would be the pleasure's intentional object. See Moore 1903, p. 210; Broad 1930, p. 234; Ross 1930, p. 137; Haezrahi 1960, pp. 74-75; Brentano 1973 (originally 1952), p. 196; Zimmerman 1980-81, pp. 34-35; Carson 1981, pp. 387-389; Chisholm 1986, ch. 6; O'Neill 1992, p. 132; and Lemos 1994, pp. 73-77.

<sup>iv</sup> See Aristotle 1980, X. 5 and Haezrahi 1960, p. 72.

<sup>v</sup> See Kant 1983 (originally 1785), p. 7 and Ross 1930, p. 136.

<sup>vi</sup> See Brentano 1973 (originally 1952), p. 172 and Chisholm 1986, p. 67. Philosophers sometimes cite Plato's Republic in connection with this view, presumably referring to 585a-e, but there Plato doesn't commit himself to it.

<sup>vii</sup> See Armstrong 1968; Pitcher 1970, p. 371; Wilkes 1977; Nelkin, 1994, p. 332; and Tye 1995. See Gallie 1954 for a similar view of enjoyment.

<sup>viii</sup> See Brandt 1979, p. 38 and Korsgaard 1996, p. 147. Gilbert Ryle advances a behaviorist forerunner of this view in Ryle 1958.

<sup>ix</sup> See Baier 1958, p. 273; Churchland 1984, p. 52; and Hall 1989, p. 646. Frequently authors don't distinguish between this view and the view above, that an experience's being painful consists in its inclining the subject to fight its continuation. See, for example, Baier 1958, p. 272 and compare p. 493 and p. 501 of Parfit 1984. Brink endorses a combination of these views in Brink 1997, p. 112.

<sup>x</sup> If some pleasures are "higher" than others, then the one person's pleasures must also be "qualitatively" like the others'. Also, as above, one might hold that the pleasures and pains must: be taken in intentional objects of the same value; be associated with



---

behavior of the same value; be equally deserved; and (for pleasures) be taken in beliefs of the same truth-value or which arise from equally good cognitive processes.

<sup>xi</sup> For example, see Griffin 1986, p. 355, fn. 33. Incidentally, I disagree.

<sup>xii</sup> Lexicality helps resolve Methuselah's Paradox, the Paradox of the Chairs and Prometheus's Paradox (and possibly The Conundrum of the Cure), as delightfully laid out in Cowen 1996.

<sup>xiii</sup> See Hutcheson 1897, pp. 421-422.

<sup>xiv</sup> I introduced this argument in Rachels 1993. Also see Rachels 1998. The argument has also been advocated in Temkin 1996, sect. 4 and Persson 1997, pp. 50-52.

<sup>xv</sup> Einstein rejected an analogous view of time. On that view, all events can be represented by points on a single line, with later events represented to the right of earlier ones and simultaneous events represented by the same point (or set of points). This view entails absolute simultaneity, for any two events either are represented by the same set of points on that line, or they are not.

<sup>xvi</sup> See, for example, Raz 1986, ch. 13, where he defends incommensurability.

“Statements of incommensurability, i.e., statements that of two options neither is better nor are they of equal value, do not compare the value of options. They are denials that their values are comparable. Incommensurability is not yet another valuation of the relative merits of two options alongside such valuations as having greater value or having equal value. It is a rejection of the applicability of such judgments to the options in question.” (p. 329)

<sup>xvii</sup> See Parfit 1984, p. 431.

<sup>xviii</sup> See Temkin 1994, pp. 361-363 and 1996, pp. 193-194.

---

<sup>xix</sup> For simplicity, I adopt a stronger thesis than Intransitivity. Intransitivity entails that, for some possibilities, X is hedonically better than Y, Y is hedonically better than Z, and X is not hedonically better than Z (which is weaker than “X is hedonically worse than Z”).

<sup>xx</sup> The first such argument, to my knowledge, appeared in Davidson et al. 1955, p. 146. The authors say, “We owe the inspiration for this example to Dr. Norman Dalkey of the Rand Corporation.” In the coming discussion, I speak of “trading” X, Y and Z. Strictly speaking, one cannot trade possibilities, but one can trade the means to making them obtain.

<sup>xxi</sup> I am grateful to Derek Parfit for suggesting this point. But I owe Parfit more thanks than that; this paper began as an undergraduate thesis he helped me write (Rachels 1993). He was an ideal advisor.

<sup>xxii</sup> See Ryberg 1996a, pp. 210-212. Ryberg introduces the idea I call “Conflation” on p. 205, and his “discontinuity” is very like Lexicality (see pp. 203-204). Ryberg’s arguments are neutral about what well-being consists in, but I’ll put them in terms of pleasure.

<sup>xxiii</sup> Ryberg asks us to compare:

A: Ten billion people who each experience a lot of ecstasy.

Z<sup>^^</sup>: A much larger population in which each person tastes ecstasy briefly, has enough mild pain to counterbalance that ecstasy, and enjoys some mild pleasure.

According to Ryberg, since the ecstasy and the mild pain within Z<sup>^^</sup> cancel out, we may compare A and Z<sup>^^</sup> by comparing A to:

---

Revised Z<sup>^</sup>: A much larger population in which each person has only mild pleasure.

A is better than Revised Z<sup>^</sup>; so, Ryberg thinks that A is better than Z<sup>^</sup>. But according to Ryberg, a Quasi-Maximizing sort of view entails that Z<sup>^</sup> is better than A. After all, the Z<sup>^</sup>-superlife—that is, all the experiences in Z<sup>^</sup> conflated into one life—would contain a longer duration of ecstasy than the A-superlife, with inconsequential pleasures and pains left over.

However, Ryberg overlooks that a Quasi-Maximizer may prefer the A-superlife to the Z<sup>^</sup>-superlife, reasoning much as he does, that “even though the Z<sup>^</sup>-superlife contains a lot of ecstasy, each little bit is counterbalanced within it by some pain; and since A’s ecstasy is better than Z<sup>^</sup>’s mild pleasure, A is better than Z<sup>^</sup>.” It’s not clear how A and Z<sup>^</sup> compare on Quasi-Maximization—and it’s not clear which is better—so Ryberg’s objection fails. (Incidentally, Ryberg might be partly aware that the Quasi-Maximizing view doesn’t obviously entail that Z<sup>^</sup> is better than A. For on p. 211 he says, “the conclusion will be that Z<sup>^</sup> is better than A.” But on the next page he merely says that a Quasi-Maximizing sort of view is “consistent with” that conclusion.)

<sup>xxiv</sup> If each life in ZZ is better than each life in AA, Quasi-Maximization may be beyond repair. One could modify the theory to give weight to average hedonic well-being. Then a Quasi-Maximizer might say that ZZ is better than AA because ZZ’s greater average well-being outweighs AA’s superlife being better. But so long as average well-being isn’t trump, this strategy merely delays defeat. For we may increase the preferability of the AA’s superlife over ZZ’s indefinitely by increasing their populations, which wouldn’t change how they compare with respect to average well-being. Eventually, on this hybrid

---

theory, AA would be considered better than ZZ (for example, when AA consists in 100 trillion people who experience ecstasy for five seconds, and ZZ consists in 100 trillion people who enjoy mild pleasure for seventy years). This would be the wrong conclusion, if each life in ZZ is better than each life in AA.

<sup>xxv</sup> See Kjetsaa 1987, p. 149.

<sup>xxvi</sup> These first three examples come from Nagel 1979, p. 130.

<sup>xxvii</sup> Parfit 1984, p. 363 (in italics); see Schwartz 1978 and 1979.

<sup>xxviii</sup> Parfit 1984, pp. 357-361. I have changed the example slightly.

<sup>xxix</sup> Schwartz doesn't think it would be, if indeed nothing else is relevant to the 14 year old's decision. See Schwartz 1978 and 1979. Defenders of Schwartz's view must grapple with Parfit's argument on pp. 367-369.

<sup>xxx</sup> Op. cit., p. 387.

<sup>xxxi</sup> Ibid., p. 388. Authors before Parfit charged classical or total utilitarianism with entailing what he calls the Repugnant Conclusion. See Rawls 1971, pp. 162-163 and Stearns 1972, pp. 616-617. The earliest discussion of something like the Repugnant Conclusion might have been McTaggart 1927, pp. 452-453.

<sup>xxxii</sup> Tyler Cowen shows some perils of trying to avoid the Repugnant Conclusion in Cowen 1996, but he doesn't consider a position very much like Quasi-Maximization.

<sup>xxxiii</sup> If some possible population of ten billion highly pleasurable lives includes no intense pleasures (their pleasures would be not quite intense), then Quasi-Maximizers will accept the Repugnant Conclusion. Below I defend (R), a similar claim. The difference is this. While (R) raises the quality of life for the enormous group of people in the Repugnant Conclusion, this variant lowers the quality of life for the ten billion.

---

<sup>xxxiv</sup> Parfit 1984, p. 420.

<sup>xxxv</sup> But the Average Principle does entail a variant of the Repugnant Conclusion. See Bill Anglin's ingenious essay (Anglin 1977).

<sup>xxxvi</sup> Parfit 1984, p. 422. Parfit makes other strong objections to the Average Principle in this section.

<sup>xxxvii</sup> Ibid., p. 405. According to Parfit, the Average Principle is but "one version" of the view that quality alone has value (pp. 402, 405), so he wouldn't identify quality with average well-being, as I do. Perhaps a view that enjoins maximizing the well-being of only the best-off would entail, on Parfit's view, that quality alone has value. But I am not sure exactly what Parfit means by "quality."

<sup>xxxviii</sup> Ibid., p. 402.

<sup>xxxix</sup> Alternatively, suppose that good lives over all times cannot have more quantitative value than that of ten billion very good lives. Once that limit is reached, it would be considered bad in terms of welfare for ten billion and one additional persons to exist, ten billion of whom have excellent lives and one of whom has a life just barely worse than neutral. But such an addition would not be bad. Moreover, the view is absurd given that, before the limit is reached, adding those persons to the population would be considered a substantial improvement.

<sup>xl</sup> This view has an absurd implication Parfit doesn't mention. Consider two states of affairs:

K: 20 billion people, all of whom are very happy and live in the same century.

L: 20 billion people, all of whom are very happy, ten billion of whom live in one century, and ten billion in another.

---

On the view in question, L is twice as good as K. But timing as such shouldn't matter.

<sup>xli</sup> Parfit 1984, p. 406.

<sup>xlii</sup> Ibid., p. 411.

<sup>xliii</sup> Ibid., p. 412.

<sup>xliv</sup> Ibid., p. 415. Here and elsewhere I substitute “Mediocre” for “Valueless,” with Parfit’s permission (see p. 416).

<sup>xliv</sup> Ibid., p. 528, fn. 40.

<sup>xlvi</sup> Ibid., p. 528, fn. 40.

<sup>xlvi</sup> Ibid., p. 415.

<sup>xlvi</sup> Ibid., p. 393.

<sup>xlvi</sup> Incidentally, Quasi-Maximizers should agree. Suffering’s disvalue is greater than ecstasy’s value, I would say, for one is better off having a neutral experience than suffering and enjoying ecstasy for equal durations.

For a nice discussion of purported disanalogies between pleasure and pain, see Griffin 1979.

<sup>1</sup> Temkin 1993, p. 290. The emphasis is mine.

<sup>li</sup> Temkin tells me that many economists seem to believe this—for example, many who accept Atkinson’s measure of inequality. See Temkin 1993, pp. 135-141 for a discussion of Atkinson’s measure.

<sup>lii</sup> This idea can be expressed without “total” and “average”: “the change described in (2) is better in terms of both superlives and representative lives (where a life L is

---

representative of an alternative S containing n lives just in case having all the experiences of L n times is exactly as good hedonically as having all the experiences in S).”

<sup>liii</sup> Parfit 1984, p. 420.

<sup>liv</sup> See ibid., pp. 419-430. For simplicity, I have omitted Parfit’s “Divided B.” This won’t affect the arguments.

<sup>lv</sup> See Temkin 1987, pp. 147-151.

<sup>lvi</sup> Parfit 1984, p. 430.

<sup>lvii</sup> Temkin anticipates this type of resolution to the Mere Addition Paradox: “However A and Z compare to some intermediate world, or set of worlds, this does not entail how they compare if preferability is deeply intransitive.” (Temkin 1987, p. 157, fn. 24)

<sup>lviii</sup> Parfit 1986, pp. 156-164. Section 6 of this paper is based on Section VI of Rachels 1998. However, there are important differences.

<sup>lix</sup> Ibid., p. 163. The next sentence refers to p. 164.

<sup>lx</sup> Parfit told me this in conversation (if memory serves!), but also he implies this by his willingness to extend his remarks about music to “the other best experiences, activities, and personal relationships, and to the other things which give most to the value of life.” (Parfit 1986, p. 164)

<sup>lxi</sup> Ryberg agrees that this is implausible: “As long as one is still very well off, I think one would gladly sacrifice a little in quality in favor of a gain in quantity.” (Ryberg 1996a, p. 206)

<sup>lxii</sup> Parfit 1986, p. 164.

<sup>lxiii</sup> “We cannot claim that great benefits to those who are worst-off would not make the outcome better if they involved the loss of Ravel’s Bolero.” (Parfit 1986, p. 163)

---

<sup>lxiv</sup> Ibid., p. 164.

<sup>lxv</sup> Ibid., p. 164.

<sup>lxvi</sup> Temkin suggested this to me in correspondence. His published work bears on these issues, especially Temkin 1993, chs. 7 and 9.

<sup>lxvii</sup> Desert, like justice, fairness, and equality, seems to lie outside the scope of hedonic value. Hence, Fred Feldman's "justicist" approach doesn't seem to apply to these problems. See Feldman 1995, reprinted in Feldman 1997.

<sup>lxviii</sup> Parfit 1984, p. 452.

<sup>lxix</sup> Reasons and Persons (1987 reprinting only), p. 443.

<sup>lxx</sup> See Sikora 1975, pp. 409-419 and Sikora 1981, pp. 128-133; Anglin 1977, p. 754; Ng 1990; Attfield 1991, pp. 127-130; Ryberg 1996b; Norcross in conversation, but for a similar assertion in print see Norcross 1997, p. 146 (where he denies what he calls claim 3); Fotion 1997, pp. 95-96; and Tannsjo 1998, pp. 160-163.