



## Utilitarianism and Personal Identity

DAVID W. SHOEMAKER

*Department of Philosophy, University of Memphis, 327 Clement Hall, Memphis, TN  
38152, USA*

### 1. Introduction

Ethical theories must include an account of the concept of a person. They also need a criterion of personal identity over time. This requirement is most needed in theories involving distributions of resources or questions of moral responsibility. For instance, in using ethical theories involving compensations of burdens, we must be able to keep track of the identities of persons earlier burdened in order to ensure that they are the same people who now are to receive the compensatory benefits. Similarly, in order to attribute moral responsibility to someone for an act, we must be able to determine that that person is the same person as the person who performed the act.

Unfortunately, ethical theories generally include a concept of a person and criteria of personal identity either as notions implied or presupposed by the already worked out theory, with little or no argument given in support of them. But both approaches are unsatisfactory, for each runs the risk of ignoring certain fundamental features of the way persons actually are. What is needed first is a plausible metaphysical account of persons and personal identity to which an ethical theory might then conform and apply.

Derek Parfit has attempted to provide such an account in *Reasons and Persons*, where he argues for what he calls the reductionist view of persons and personal identity and then attempts to show how such a conception provides a metaphysical background that lends important partial support to utilitarianism.<sup>1</sup> But Parfit's metaphysical view is neutral between two possible conceptions of personhood, and utilitarianism presupposes the truth of the more radical of the two conceptions, which precludes the possibility of there being a class of goods crucial to any plausible ethical theory. Utilitarianism thus presupposes a wildly implausible conception of persons, and so is itself implausible.

## 2. Parfitian Reductionism

Parfit is a reductionist about personal identity. He holds that the facts of personal identity simply consist in more particular facts about brains, bodies, and interrelated physical and mental events.<sup>2</sup> In contrast, non-reductionism is the position that the facts of personal identity consist in some further facts about Cartesian egos, souls, or other kinds of “separately existing entity.”<sup>3</sup> Parfit also favors a psychological criterion of personal identity: X and Y are the same person if and only if X is psychologically related to Y in a certain way, and this relation has not taken a branching form.<sup>4</sup>

The two relations together that constitute the identity-preserving psychological relation between X and Y are psychological connectedness and psychological continuity, together known as Relation R. Psychological connectedness consists in direct psychological relations, such as direct memory connections, intentional connections, character connections, and connections of beliefs, desires, and goals. Such connections can hold to any degree. Strong connectedness, for instance, holds between X and Y when “the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person.”<sup>5</sup> Psychological continuity holds between X and Y when there are overlapping chains of strong psychological connectedness between them. Thus, X and Y are the same person if and only if X is uniquely psychologically continuous with Y.

Parfit goes on to claim, however, that the identity-relation is not the relation that matters in terms of our ordinary survival and our anticipation or concern for the future. As he argues in a discussion of the case he calls “My Division,” uniqueness does not really contribute anything significant to the identity-relation.<sup>6</sup> What matters most in terms of our attitude toward the future is captured by Relation R. Consequently, we are to look to Relation R as the relation that matters in identity. Doing so has a number of significant implications for rationality and morality. For one thing, the unity of our lives is no longer guaranteed. Our lives may be more or less unified, given the degree to which psychological connectedness holds. For instance, I am presently strongly connected to that stage of myself that existed yesterday, but I am fairly weakly connected to that long-ago existing ten-year-old stage of myself. Parfit suggests the language of successive selves to illuminate the different degrees to which this relation might hold. We might use the word “self,” then, to refer to a collection of person-stages united by strong psychological connectedness, such that my ten-year-old self could be viewed as a past self, while my eighty-year-old self would presumably be a future self. The parts of my life with which I am currently strongly psychologically connected are united as my present self.<sup>7</sup>

In this way, different selves occasionally resemble different persons, and Parfit indicates that, at certain times and places, selves might be thought of as

the appropriate objects of moral concern.<sup>8</sup> But this notion also implies that our lives may not be unified in certain important respects. Psychological connectedness is certainly not guaranteed to unify our entire lives, and so the reductionist view itself implies at least the partial disintegration of persons.<sup>9</sup>

### 3. Utilitarianism and Reductionism

Utilitarianism is an ethical theory for ranking various outcomes from an impersonal standpoint. Utilitarians hold that the best state of affairs among relevant alternatives contains the greatest net balance of aggregate individual welfare. Utilitarianism is impersonal insofar as it involves a focus solely on the total amounts of utility at stake in various outcomes, and “[i]t makes no moral difference how these amounts are distributed as between different people.”<sup>10</sup> Many utilitarians claim that the impersonality of the theory is entailed by a close analogy that obtains between cases of intrapersonal and interpersonal maximization. As Parfit remarks: “Since their attitude to sets of lives is like ours to single lives, [u]tilitarians ignore the boundaries between lives.”<sup>11</sup> Parfit further believes that utilitarians accept this analogy because they accept a reductionist view about personal identity. If a person’s life is less deeply integrated than it would be on a non-reductionist view, then while principles of distributive justice central to non-utilitarian views ought to be given greater scope, targeting past, present, and future selves, they nevertheless ought to be given less weight. After all, if a person’s life is less unified than we normally think, and this undermines the hard and fast boundaries between lives as well, then distributive principles relying on the separateness of persons and the individual unity of a person’s life as deep facts will have far less moral importance than they would on a non-reductionist view. Some critics have claimed that utilitarians ignore the boundaries between lives because they think of all people as together constituting a collective super-person, but this charge is false, according to Parfit. Because of the partial disintegration of persons suggested by reductionism, utilitarians “may be treating benefits and burdens, not as if they all came within the same life, but as if it made no moral difference where they came.”<sup>12</sup> Thus, Parfit suggests that reductionism may lend significant support to utilitarianism, simply because utilitarians can claim to be treating sets of lives like single lives, given that single lives are not deeply unified and are, in fact, metaphysically like sets of lives.

This is only one of the arguments Parfit offers in support of utilitarianism, and it is based on what we may call a moderate reductionist view of personal identity, a version of reductionism in which selves are considered the morally significant units. Persons are only partially disintegrated on this view, for the selves existing within a person’s life as a whole remain salient and distinguishable unities. There is no metaphysically deep entity such as a soul

providing life-long unity for persons. Instead, the only relation that matters in our identity is psychological connectedness, and this relation cannot guarantee the unity of a complete life. Consequently, because there can be several successive unified selves within the life of one person, and the unities are like different persons in important respects, single lives are analogous to sets of lives.

Moderate reductionism is but one way to view the disunity of persons. Another way comes from what we may call extreme reductionism. It is consistent with Parfit's view to say that, given the truth of reductionism, persons are completely metaphysically disintegrated. An advocate for such a view might argue that the only relation that could possibly provide any unity for persons is a non-reductionist unity-relation such as the unity provided by a soul or Cartesian ego. But because reductionists reject such unity-relations, persons are completely disunified, and an ethical view that attempts to justify certain rules or distributions based on the existence of any individual unities is false.

Extreme reductionism might lend support to utilitarianism in the following way. Many people claim that we are justified in maximizing the good in our own lives, but not justified in maximizing the good across sets of lives, simply because each of us is a single, deeply unified person, unified by the further fact of identity, whereas there is no such corresponding unity across sets of lives. But if the only justification for the different treatment of individual lives and sets of lives is the further fact, and this fact is undermined by the truth of reductionism, then nothing justifies this different treatment. There are no deeply unified subjects of experience. What remains are merely the experiences themselves, and so any ethical theory distinguishing between individual lives and sets of lives is mistaken. If the deep, further fact is missing, then there are no unities. The morally significant units should then be the states people are in at particular times, and an ethical theory that focused on them and attempted to improve their quality, whatever their location, would be the most plausible. Utilitarianism is just such a theory.<sup>13</sup>

Thus, there seem to be two ways to support utilitarianism by appealing to reductionism. Moderate reductionists may provide support for utilitarianism by focusing on selves as the morally significant units. Extreme reductionists, however, may provide support for utilitarianism by focusing on momentary experiences as the morally significant units. Which view do utilitarians in fact hold, and which view is the more plausible of the two? Parfit remains neutral on both questions, implying that utilitarians may appeal to either view and calling both versions of reductionism "defensible."<sup>14</sup> Nevertheless, utilitarians treat sets of lives like single lives because they are in fact extreme reductionists. They specify the only significant moral and metaphysical units to be the states people are in at particular moments. But this view of persons is both psychologically problematic and metaphysically unjustified.

#### 4. Utility and Quantities of Good

Regardless of their particular specifications of the good to be maximized, all utilitarians would accept the following utilitarian principle of distribution:

One alternative is at least as good as another if and only if it has at least as great a total of people's good.<sup>15</sup>

The question for utilitarians to answer, however, is precisely how various goods distributed at various times or to various people together comprise that total good. Answering this question involves evaluating the relative size of various goods and how much such goods are supposed to count in the assessment of total good. Utilitarians must specify how we are to weigh goods in all contexts.<sup>16</sup> The reason they must do so is that in order to be a theory with any practical applications, utilitarianism must include a workable method for assessing the goodness of possible outcomes and thus the rightness or wrongness of particular actions. Otherwise, maximizing that good would be a hopeless task. If they cannot assess and compare certain goods, they have no way of determining which state of affairs is best.

It is also essential that utilitarians be able to provide a method of weighing goods in all contexts that is derived from the meaning of the notion of a quantity of good in the contexts and how such a notion generally acquires such meaning. Utilitarianism is supposed to be a theory about what outcomes are best, and so the method for determining which outcomes are better than others must take into account what the phrase "quantity of good" means in different situations. There are three basic contexts in which comparisons of goods might be made: first, the case in which one person, in the face of uncertainty, is deciding between different prospects with different payoffs, where the payoff in the prospect chosen will be distributed to that person at one time; second, the interpersonal case, in which two persons are deciding between prospects involving different payoffs, where the payoffs in the prospect chosen will be distributed to them at one time; and third, the intrapersonal case, in which one person is deciding between different prospects with different payoffs, where the payoff in the prospect chosen will be distributed to that person at different times. In each case, we may ask what the phrase "quantities of good" means and how it acquires such meaning. If it acquires its meaning in different ways or means something different in different contexts, then weighing the various goods together will be meaningless in terms of total good. In order for a complete assessment of goods accrued from various contexts to be made, it is crucial that what "quantity of good" means is the same in each of the contexts. Otherwise, in assessing total good, we will be forced to compare, essentially, apples and oranges. Thus, it is important to explore first how the notion of a quantity of good acquires its meaning in relation to the concept of utility.

John Broome presents a compelling case for how such meaning is acquired in each of the three basic contexts that, if correct, provides a method of weighing total good that lends overwhelming support for the utilitarian principle of distribution. His general argument runs as follows. If the betterness relations involved in each of the three contexts are coherent, they will satisfy the axioms of expected utility theory and can thus be represented by utility functions. Consequently, if the good assigned by each party to potential payoffs can be completely represented by certain utility functions, then all we have to do to assess the good of various alternatives is assess the good for the individuals involved, translate that good into utility functions, and sum up the utilities. The alternative with the greatest amount of utility, then, will always be the alternative that has the greatest amount of good.

Consider first, then, the betterness relation involved in the first context, the case in which one person decides between two uncertain prospects: "A is at least as good as B for person j." This betterness relation satisfies the axioms of expected utility theory, so it is coherent. As a result, it can be represented by a utility function, such that  $U_j(A)$  is at least as great as  $U_j(B)$  if and only if A is at least as good for j as B.<sup>17</sup> Not only do these utility functions represent the order of goodness for j, they also represent the quantities of good for j, because it is in precisely these kinds of comparisons that the notion of quantities of good gets its meaning. When weighing different alternatives in states of uncertainty, we weigh possible gains and losses against one another. If we can determine which prospect is better than another, what are we doing other than saying that the good of one prospect counts more than another? Thus, in this first case we can simply find the right utility function for the person and add up the utilities of each prospect. Because the right utility function will capture all that is good for the person, the prospect with the greatest amount of utility will then be the prospect that has the greatest amount of good, making it the best alternative for that person.

Similarly, for contexts in which we must compare the good of two possible payoffs to two different people, we can also represent the quantities of good involved with certain utility functions. The betterness relation involved here is "A is at least as good as B," where A and B are drawn from the same set of uncertain prospects. Assuming that interpersonal comparisons of goodness can be made, this relation is also coherent, so it can be represented by utility functions that will serve to represent the actual quantities of good at stake in both prospects. To determine which of two prospects is better, we can merely sum up the various utilities involved.<sup>18</sup> Consequently, there are at least two basic contexts in which the assigned utility functions serve the same purpose. In the context of comparing goodness for one person in the face of uncertainty at a particular time, and in the context of comparing goodness for two or more people in the face of uncertainty at a particular time, the better alternative is guaranteed to be the one with greater total utility. But what about comparisons

across time? Does utility represent quantities of good when we weigh goods that come at different times in people's lives? Perhaps the notion of quantities of good gets its meaning in a different way from these sorts of comparisons.

Consider, then, the betterness relation involved in this third context, the relation known as a dated betterness relation: "A is at least as good as B for j at time t," where A and B are drawn from the same set of uncertain prospects as before. This relation has to do with intrapersonal assessments of goodness in which a person decides between two or more prospects that will involve payoffs coming at different times in that person's life. It turns out that this relation as well is coherent, and thus can be represented by utility functions. It seems, then, that we should be able to represent quantities of good in every conceivable context. If so, then four very important conclusions follow. First, the same utility functions would determine how to weigh differences in good in and across the three contexts. Second, the utility functions would in fact represent quantities of good. The notion of quantities of good could get its meaning only from within the three contexts. Indeed, what other context could there be? Third, whenever weighing of goods would have to be done, what is best would always be determined by the total of utility. Whenever deciding what action would lead to the best outcome, all we would have to do is find the one with greatest total of good. Finally, then, the utilitarian principle of distribution would be wholly justified. As Broome puts it, "[T]he best alternative would always be the one with the greatest total of good."<sup>19</sup>

Broome has shed some light on the foundations of utilitarianism. He has provided a metric of goodness, based on how the notion of quantities of good acquires its meaning, and he has shown that if utilitarians want to maximize the good, regardless of how that good is to be distributed, and regardless of the contexts in which the weighing of that good takes place, then this is the method by which they must perform that task. The ultimate conclusion, then, is "that the notion of a quantity of good acquires its meaning in such a way that it turns out to be best to maximize the total of people's good."<sup>20</sup>

Nonetheless, an important obstacle to such a conclusion remains, for there is perhaps reason to worry that dated betterness relations do not capture everything that is good for persons in the intrapersonal case. If they do not do so, then there would exist one context in which individual utility functions would not represent quantities of good. If dated betterness relations do not capture everything that is good for persons, then it would be possible for there to be a case in which a prospect with greater utility for someone than another could actually turn out to be the worse prospect for that person. This possibility would imply that the utilitarian principle of distribution rests on an incomplete and dubious foundation.

Consider one way in which dated betterness relations might not capture all of a person's good. Suppose Mary is deciding between two equally likely prospects, each with two possible payoffs which will be distributed to her at

two different times. Suppose, further, that both prospects are equally good for her at each of the payoff times. The utilities for each prospect, then, would also be equal. Nonetheless, one prospect may actually be better for her from the standpoint of her life as a whole, given, say, the ordering of the payoffs. For example, it may be better for her life as a whole if greater payoffs are distributed to her after smaller payoffs. But the superior value of that ordering “appears from the standpoint of the person as a whole, and does not show up in either of the times taken separately.”<sup>21</sup> Thus, while the prospects are equally good for her at each time of the possible payoffs, when the perspective switches to the standpoint of her life as a whole, one prospect may in fact be better for her than another.

There is, however, no such analogous problem with the betterness relation involved in the interpersonal case, given that there is no corresponding standpoint of the persons as a whole that could yield a different assessment of goodness. The only targets of consideration in the interpersonal case are goods for the different parties at one time, and there is no holistic good for both parties together that could justify an alternative assessment of the good of the various prospects.

There remains, then, a seeming disanalogy between the interpersonal and intrapersonal cases. The intrapersonal case is problematic because we are able to judge prospects from a standpoint that links the various goods of one person at different times together, whereas there is no analogous standpoint linking the good of different people together that threatens the interpersonal case. But unless the intrapersonal case and the interpersonal case are found to be analogous, the utilitarian principle of distribution is seriously threatened. The only way the intrapersonal and interpersonal cases can be made analogous, and thus the only way utilitarianism can be made into a workable moral theory, is if utilitarians view persons as individuated, momentarily existing experiencers, and not as enduring entities. They must target as the morally and metaphysically significant units only momentary states that people are in at any given time and regard as irrelevant the larger contexts within which the states take place. Utilitarians must be extreme reductionists.

## **5. Utilitarianism and Holistic Goods**

Utilitarianism’s foundational tenet is that maximizing utility maximizes the good. This task involves being able to weigh quantities of good on the same scale of measurement in and across every possible context. If we are unable to do so, then we cannot fully determine which possible alternative maximizes utility and thus the good. But utilitarians claim that they can indeed determine what maximizes utility, so they must have in place a method for weighing quantities of good in and across every possible context. Broome has offered

a powerful method for doing so, derived from the way in which the notion of quantities of good gets its meaning. His claims about how utility works in the face of uncertainty at a particular time for the individual and interpersonal cases are extremely plausible, and so there seem to be two contexts in which utility does represent quantities of good. The question, though, is whether or not utility represents quantities of good in the intrapersonal context, given the possible incompleteness of dated betterness relations. If such relations do capture all that is good or bad for a person, however, utilitarianism would seem to rest on extremely firm ground.

Something may certainly be bad for a person at one time and good for that person at another time. For example, the pain and sickness immediately resulting from a vaccination shot may be bad for a person at the time but good for that person at some later date.<sup>22</sup> The question, however, is not whether or not dated betterness relations exist. They do. The question is whether or not they capture all that is good or bad for a person. Are there some goods for persons that are not good for them at any particular time? One example might be happiness, for it seems possible to be a happy person in general without being in a happy state at any particular time. As J.J.C. Smart remarks, we can hardly talk “of a man being happy at a quarter past two precisely.”<sup>23</sup> Nevertheless, for hedonic utilitarians like Smart, happiness reduces to enjoyment, which is in fact a state a person is in at particular moments. As he notes, “[H]appiness involves enjoyment at various times, just as a wet climate involves rain at various times.”<sup>24</sup> Thus, while happiness may initially seem to be a good that persons do not receive at any particular time, the constituent elements of happiness, states of enjoyment, are indeed such goods, and because utilitarians are able to reduce happiness to the sum of moments of enjoyment, these summed individual moments end up capturing all that is good or bad for a person.

Perhaps, though, being respected is a good that escapes this sort of reduction. Suppose Louise works extremely hard for such respect. Whether or not she achieves it depends on whether or not her colleagues do in fact respect her. But such respect may come only after her death. While postmortem respect may still count as a good for her by making, after the fact, her life worthwhile, there would seem to be no particular time during the course of her life when she could be said to have received this good. Utilitarians might respond to this case, however, in one of two ways. They might simply deny that being respected is really a good, insisting instead that Louise’s feeling or belief that she is respected is what counts as the relevant good, and this feeling or belief would then have a particular temporal location. Alternatively, even if they admit that being respected is a good for Louise, utilitarians may simply date this good to all the times on which she was working for that respect.<sup>25</sup> It is easy to see how such responses could be offered to account for many other seemingly non-dated goods as well.

It remains difficult to rule out the possible existence of non-dated goods, however, simply because of the natural standpoint of the person as a whole. This standpoint seems to link together the good of a single overall unit at different times and thus allows for there to be two different methods for weighing goods in the intrapersonal case. One method would involve the weighing of goods that are goods for a person at a particular time. The other method would involve the weighing of goods that are not goods for the person at any particular time and would consist in viewing such potential goods as holistic goods, or goods for a single enduring unit as a whole, as opposed to viewing them in terms of their value at any given moment. Consequently, a disanalogy between the interpersonal and intrapersonal cases remains, for the interpersonal case involves merely the linking together of the good of more than one unit. It links together the good of different people at a particular time, and there is no further possible standpoint from which the different people could be all the same.

The intuition behind this seeming disanalogy is that if persons are enduring metaphysical units, then it is possible for them to have non-dated, holistic goods. But what reasons do we have to trust this intuition? It is obviously true that if the only existing metaphysical person-units are momentary experiencers, then there can be no holistic goods, for such goods by definition involve benefits that are not good for the beneficiary at any particular moment, and person-units would only exist for a particular moment. But the mere existence of enduring metaphysical units may not in itself ensure the possibility of holistic goods. Perhaps there simply are no holistic goods, regardless of the temporal span of the metaphysical units, or perhaps what we think are holistic goods can always be reduced to mere sums of particular momentaristic goods. Such a view may seem implausible, but it is not readily apparent why. Indeed, we have already seen how such alleged holistic goods as happiness and being respected may plausibly be reduced to momentaristic goods.

Our first task, then, is to establish the independent possibility of non-reducible holistic goods. There are at least three that meet the requirements: success, autonomy, and aesthetic quality of lives or selves. Consider first the good of success in the context of marriages and careers. Success there is a matter of degree and is not reducible to a mere summation of good individual moments. Indeed, the pattern of good moments within the marriage or career makes a difference in a determination of its overall success. Suppose there are two marriages, both containing the same number of agreed upon good moments and the same number of agreed upon bad moments. The first marriage contains the bulk of its good moments in the first half of the marriage, while the last half is dominated by bad moments, whereas the second marriage contains its bulk of good moments in the last half, while the first half is dominated by bad moments. Most of us would agree that the second marriage is the more successful marriage of the two because of the pattern in which the

good and bad moments are distributed, despite there being the same number of individual moments in each of the two marriages. Consequently, the good moments contributing to that success, while equally good at a particular time for each of the parties in both marriages, may contribute more or less to the overall success of the marriages depending on where they occur within the context of the marriages as a whole.

The same point holds true for success in careers. Suppose Agnes and Bob work for the same law firm, in which the moments contributing to success involve winning a certain number of big money lawsuits. Over the course of their careers, both Agnes and Bob win twenty such lawsuits apiece. Agnes, however, wins eighteen in the first ten years of her career, while only winning two in the remaining twenty years, while Bob wins only two during his first ten years, but wins eighteen in the remaining twenty. Again, it seems Bob's is the more successful career because of the pattern in which the successful moments are distributed throughout his career, and his greater success is not reducible to a simple summation of the various moments of success involved. In order to judge the level of success, we must evaluate the career as a whole, and that overall success is not constituted solely by the good individual moments along the way.

Autonomy is perhaps an even clearer example of a holistic good, for it is coherently attributed as a good only to persons or selves as a whole. It would simply make no sense to say that, because a person made a certain number of seemingly autonomous choices over a certain period of time, that person's overall end of autonomy had been achieved. Autonomy has to do with the relation between certain choices and moments of action, it has to do with a person's ongoing reactions to certain contexts and that person's ongoing relationships with other people, and such elements are simply not captured by a mere summation of certain momentary goods. They make sense only within a larger context. As James Griffin puts it, "[A]lthough we can talk about the amount of some of our ends (e.g. pleasure) in isolation, with others it is more difficult (e.g. autonomy)."<sup>26</sup>

Perhaps, though, autonomy itself is not a good. Instead, it may be that what persons are really striving for is the feeling or belief that they are autonomous, which is captured by the notion of a momentary, dated good. However, such a response is implausible in the case of autonomy, for it is simply not this feeling or belief that persons generally want when choosing autonomy as an end. What they want is to be the kind of person or self who is a free chooser and self-responsible entity, whether or not they feel or believe this to be the case. Of course, being autonomous means occasionally feeling autonomous, but the feeling of autonomy is a separate good from autonomy itself. It is a feeling of self-satisfaction associated with, but not constitutive of, the achieving of an independent, holistic good. Moreover, as with success in marriages and careers, the pattern of choices and momentary goods associated

with autonomy makes a difference in the overall assessment of that good. As Griffin notes, when we assess the various ingredients involved in goods like autonomy, “the important estimate of how valuable they are is their contribution to a whole life. . . . The form of the judgment is, *this* combination makes a more valuable life than *that*.”<sup>27</sup>

A final example of a holistic good is the aesthetic quality of a life or self. We may view our lives in much the same way we view works of art and wish for our lives to have great aesthetic quality. For example, we may want our lives to have a certain dramatic structure, including elements of comedy and perhaps tragedy, with a thrilling climax and satisfying, but brief, denouement. Having a life of this form may be the greatest good we can gain.<sup>28</sup> In this case it is evident that such aesthetic quality can be only a holistic good. The good involved does not reduce to the summing of individual moments, but applies solely to the overall structure of the lived life. In fact, the value of the individual moments may be fairly trivial when viewed independently of this context. It is only when they come together, within a particular envisioned structure, that some further good arises for the person as a whole. In addition, the good being sought in cases like this is not the momentary good of a feeling or belief about aesthetic quality. What is being sought is a holistic good which, perhaps even in accordance with the dramatic structure envisioned, may never be felt or directly appreciated.

Thus, it seems clear that there are non-reducible, holistic goods. But what makes such goods possible? It is simply the existence of enduring metaphysical units. What makes them possible is that persons or selves exist, metaphysical entities that endure for a longer period of time than the momentary experiences which the persons or selves undergo. There must be a larger context within which momentary experiences take place in order for the notion of holistic goods to get off the ground. When the morally significant metaphysical units are of some duration, such goods may then be generated. This scenario leaves utilitarianism in dire straits, however, for as long as there are holistic goods, there is also a standpoint from which the good of one person or self at various moments can be linked together that remains disanalogous to the weighing method in the interpersonal case, which involves linking together the good of more than one unit. As long as the intrapersonal and interpersonal cases remain disanalogous, then, utilitarianism cannot provide a complete procedure for weighing good.

Consequently, there seems to be only one way for utilitarians to maintain that the good from intrapersonal distributions can be weighed against the good of other contexts, and that is simply to presuppose that the only assessment to be done in the intrapersonal case involves the linking of the good of different units as well. To fully support the general utilitarian principle of distribution, they must hold, either through metaphysical argument or brute assumption, that there in fact is no standpoint from which individual good at different times is unified. They must maintain instead that temporally distinct, momentarily

existing person-segments really are just like the different persons whose good is linked in the interpersonal case. If they do so, then the same utility functions will determine how differences in good are to be weighed against one another in and across every possible context, no good will be excluded, and the utilitarian principle of distribution will be saved. This means that utilitarians must be extreme reductionists about persons, not moderate reductionists or non-reductionists. If people are enduring and at least partially unified metaphysical entities, then it is possible for them to have holistic goods, in which case utilitarians could not provide a complete procedure for weighing good. Consequently, the only way to ensure that they can provide such a procedure is for them to assume that persons are nothing more than momentarily existing atoms, and what they have to target as the morally significant units are just those person-atoms, not non-reductionist lives or moderate reductionist selves. For the same utilities to count in all contexts of distribution, dated betterness relations must capture all that is good or bad for persons. For that to be the case, there must be no standpoint from which the good of different moments is unified into one thing, and within which “good” has a different meaning. Any theory targeting lives as both metaphysically and morally significant could yield such a standpoint, as could any theory targeting selves, those enduring entities unified by psychological connectedness. In each case there could be something good for the enduring unit as a whole, but not good for that unit at any particular moment. Therefore, the only way utilitarians can take the intrapersonal and interpersonal cases to be analogous, and thus assimilable into one another, is if they view persons in metaphysical terms as non-enduring entities and take the morally significant units to be the momentary person-atoms. They must view experiences as being things experienced by experiencers of no greater duration than the experiences themselves. The only context of any metaphysical or moral significance must be the momentary and immediate context within which the experiences take place. Utilitarians thus must view the only significant metaphysical and moral units to be states that people are in at particular times, regardless of the various connections those person-atoms might bear to any other person-atoms. Various diachronic connections of memory, intentions, belief, and body can play no role whatsoever in either their metaphysical worldview or their moral calculus. Utilitarians, then, must hold the extreme version of reductionism in order for utility to represent quantities of good in every context and in order for them to be able to maximize the good.

## **6. Problems with Extreme Reductionism**

What has been established to this point is merely that if utilitarianism is to be a viable ethical theory, an extreme reductionist account of persons must be

presupposed at its very foundation. However, this result may strike some utilitarians simply as a necessary, albeit counterintuitive, implication of what remains a highly plausible moral theory. As Smart remarks, “The utilitarian . . . will test his particular feelings by reference to his general principle, and not the general principle by reference to his particular feelings.”<sup>29</sup> Thus, what remains to be shown is exactly why such an atomistic presupposition makes utilitarianism itself seriously problematic.

Consider once more the contrast between the two specific versions of reductionism. On a moderate reductionist view of personal identity, selves are targeted as the morally significant units. As entities unified by strong psychological connectedness, selves typically endure over a certain stretch of time and so are not atomistic entities. In order to be so they would have to be entities that came into existence at a particular time, bore no relations of memory, belief, character, desires, or goals with any other entity, and then underwent an experience and disappeared. But this is clearly not the way things ordinarily work. Normal people do in fact bear these relations to previous stages of themselves. An extreme reductionist favoring an atomistic conception of persons would not deny that these connections typically occur. Such a person would have to claim that they neither serve to unify people nor are relevant for moral purposes. On this view, people are entirely disintegrated. As a result, the only units of any significance for the atomist are simply different momentary experiencers, and just because the experiencers may bear some connections to other experiencers at different times, that fact can be of no metaphysical or moral importance. Nevertheless, the connections enable the generation of holistic goods, which in turn render utilitarianism incomplete with regard to overall assessments of good. The connections give rise to the standpoint the utilitarian is unable to handle, the standpoint of the self as a whole, from which there can be goods that are not good for the self at any particular time. Benefits and burdens may have effects on an entire self, without having any such effects on contextless atoms, that are ignored when the significance of psychological connectedness is ignored. If utilitarians claim to be able to assess the good involved in every moral context, but they ignore the special connectedness that exists and can unify person-stages, then they also ignore the good that can arise because of the special connectedness between person-stages. As a result, they are not able to assess the good involved in every moral context.

But why are utilitarians wrong to ignore psychological connectedness and deny the metaphysical unities it can provide? After all, utilitarians may maintain that even though momentary experiencers bear the relation of strong psychological connectedness to other such experiencers, that relation still does not give rise to any enduring metaphysical unities. Their reasoning may run as follows: what we thought provided enduring metaphysical unities, the deep further fact, is in actuality always missing, given the truth of reductionism.

But if that was the only thing that could have provided enduring unity, and that fact is missing, then there are no enduring unities. The only remaining metaphysical units, then, from which we draw our moral units, are momentary, atomistic entities.

Such reasoning is too hasty. Even though the further fact of identity is always missing, there are other relations we can turn to that provide a source for enduring metaphysical unities. Indeed, the relations constituting Relation R, psychological continuity and connectedness, are just such relations. We are not justified in abandoning metaphysical unities solely because the further fact is missing if there are other relations that can take its place. Psychological continuity can provide the unity-relation for the person, while strong psychological connectedness can provide the unity-relation for the self. As a result, the slide from non-reductionism to extreme reductionism is not the slippery slope assumed by the utilitarian and can be halted if we merely switch our focus from the further fact to the relations that matter in personal identity. Moderate reductionism certainly provides us one such method for doing so. We also have very good pragmatic reasons for doing so. As Bernard Williams points out, if I did not conceive myself as an enduring unity of some kind, I would not have the psychological make-up for anything to count as a reason for action on any given occasion.<sup>30</sup> After all, what reason would I have to act at all, at any moment, if I did not think it would be me that would live through, and benefit from, that action? Having a sense of my own future is essential for my having reasons to act now. What gives me a reason for living are my present projects, and I cannot even understand them as my present projects unless I also understand them “*as the projects of one [person] who will . . . change.*”<sup>31</sup> Thus, for the projects which give meaning to my life to be comprehended as mine, I must conceive myself as an enduring entity propelled into the future by those present projects.<sup>32</sup> Of the two specific versions of reductionism set forth, only moderate reductionism provides plausible metaphysical reinforcement for such a conception.

One final recourse for a utilitarian is to say, “You may be right. People have a psychological need to conceive themselves in this way. Our theory will allow for people to conceive themselves as enduring agents, even though they are really not. We secretly maintain our extreme reductionism in order to weigh good in all contexts, but we also permit people to deceive themselves about their metaphysical existence to best preserve their psychological well-being and allow the utilitarian society to run as smoothly as possible.”<sup>33</sup> This reply, however, fails to take seriously the force of Williams’s point, for it also applies directly to a self-aware utilitarian. Indeed, why would such a utilitarian, or any other utilitarian privy to this secret, have any reason to act knowing the so-called truth about non-enduring existence? What reason would such a utilitarian have for promoting or, perhaps, concealing utilitarianism, or for doing anything at all? It seems this response belies a real belief in enduring

metaphysical unities, despite the protestations to the contrary, for to maintain this position over any amount of time would require the belief of being an enduring entity and not being an enduring entity. Such ongoing cognitive dissonance would be too much for even the most die-hard theorist.

## 7. Conclusion

Utilitarians can provide a complete procedure for weighing goodness only by presupposing that persons are not enduring metaphysical units. If they were to do otherwise, by presupposing either a non-reductionist or a moderate reductionist view, they would have to allow for the possible generation of holistic goods, goods which cannot be represented by the same utility functions used in other contexts. Their theory would then be one in which the same utility functions would not represent quantities of goodness in every context and they would thus be unable to assess total goodness. As a result, they must deny the very possibility of holistic goods, and the only view of persons allowing them to do so is extreme reductionism, a position in which experiencers are viewed as having the same temporal existence as their dated, momentary experiences.

There is a metaphysical problem and a pragmatic, psychological problem with viewing persons in this way. On the one hand, the slide from non-reductionism to extreme reductionism about identity and unity is not warranted. Moderate reductionism provides a substitute for the alleged further fact in the form of psychological connectedness. Such connectedness can provide a metaphysical unity formerly thought provided by the further fact. On the other hand, in order to have any reason to act at all, we must conceive ourselves as enduring and unified entities to some extent, and moderate reductionism lends plausible metaphysical reinforcement to this pragmatic conception of ourselves that extreme reductionism does not. All of this leads to an ironic conclusion. Parfit argues that his form of reductionism makes utilitarianism more plausible than it would have been if non-reductionism were accepted, which is indeed the case. Extreme reductionism, after all, is still one possible version of Parfit's form of reductionism. The much more plausible specification of Parfit's form of reductionism, however, is moderate reductionism, and this version of his view ultimately provides a very serious reason to doubt utilitarianism.<sup>34</sup>

## Notes

1. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).
2. *Ibid.*, pp. 210–211.
3. *Ibid.*, p. 210.
4. *Ibid.*, p. 207.

5. Ibid., p. 206.
6. Ibid., pp. 254–266.
7. Ibid., pp. 302–306.
8. Although he insists elsewhere that talk of successive selves is not to be taken literally. See Parfit's Postscript to a reprinted version of "Later Selves and Moral Principles," in Ted Honderich and Myles Burnyeat, eds., *Philosophy As It Is* (New York: Penguin Books, 1979), p. 211.
9. Parfit, *Reasons and Persons*, p. 336.
10. Ibid., p. 330.
11. Ibid., p. 331.
12. Ibid., p. 336.
13. For an argument similar to this one, see Parfit, *Reasons and Persons*, pp. 342–345.
14. See, for example, Parfit's discussion in *Reasons and Persons*, pp. 342–343.
15. John Broome, "Utilitarian Metaphysics?" in Jon Elster and John E. Roemer, eds., *Interpersonal Comparisons of Well-Being* (Cambridge, England: Cambridge University Press, 1991), p. 70.
16. Ibid., pp. 70–71.
17. Ibid., p. 72.
18. Ibid., pp. 73–79.
19. Ibid., p. 81.
20. Ibid.
21. Ibid., p. 83.
22. Ibid., p. 79.
23. J.J.C. Smart, in J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge, England: Cambridge University Press, 1973), p. 23.
24. Ibid.
25. Broome discusses a somewhat similar example and reply in Broome, p. 82, fn. 11.
26. James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Oxford University Press, 1986), p. 36.
27. Ibid.
28. I am grateful to Andrew Cross for this suggestion.
29. Smart, *Utilitarianism: For and Against*, p. 69.
30. Bernard Williams, "Persons, Character and Morality," in Amelie Oksenberg Rorty, ed., *The Identities of Persons* (Berkeley, CA.: University of California Press, 1976), pp. 205–207.
31. Ibid., p. 206.
32. Ibid., p. 209.
33. I am grateful to Andrew Cross for pointing out this final fall-back position.
34. I am indebted to Eric Cave, Andrew Cross, William Harms, Penelope Maddy, Alan Nelson, Brian Skyrms, Gary Watson, a University of California, Irvine colloquium audience, and the editor of the *Journal of Value Inquiry* for their extremely helpful remarks on earlier drafts of this paper.

