

# Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice

FRED FELDMAN

*University of Massachusetts, Amherst*

1. *Introduction.* In a famous passage near the beginning of *A Theory of Justice*, John Rawls discusses utilitarianism's notorious difficulties with justice. According to classic forms of utilitarianism, a certain course of action is morally right if it produces the greatest sum of satisfactions. And, as Rawls points out, the perplexing implication is "...that it does not matter, except indirectly, how this sum of satisfactions is distributed among individuals any more than it matters, except indirectly, how one man distributes his satisfactions over time."<sup>1</sup> He concludes the passage by saying that "[u]tilitarianism does not take seriously the distinction between persons."<sup>2</sup>

As I understand him, Rawls is here alluding to a very well known problem. It has been illustrated by appeal to a host of remarkably striking and ingenious examples. Among these are the Organ Harvest,<sup>3</sup> the Small Southern Town,<sup>4</sup> and the Colosseum.<sup>5</sup> One of the most straightforwardly relevant examples is given by Ross in *The Right and the Good*. Ross sketches a case in

---

<sup>1</sup> John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971): 26.

<sup>2</sup> Op. cit., 27.

<sup>3</sup> The Organ Harvest is discussed in Judith Thompson's "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204-17. Philippa Foot discusses essentially the same case in her "Abortion and the Doctrine of the Double Effect," *Oxford Review* 5 (1967).

<sup>4</sup> Most of the elements of the case of the Small Southern Town can be found in E. F. Carrick's *Ethical and Political Thinking* (Oxford: Oxford University Press, 1947), in a passage that is cited in John Rawls, "Two Concepts of Rules," *The Philosophical Review*, 64 (1955): 3-32. Rawls' own "telishment" case is again quite similar. Kai Nielsen's case of the Magistrate and the Threatening Mob, described in his "Against Moral Conservatism" *Ethics* 82 (1972): 113-124, makes the same point. Nielsen's essay is reprinted in Louis Pojman, *Ethical Theory* (Wadsworth: Belmont, California, 1989): 181-188. The example appears on p. 183.

<sup>5</sup> The case of the Christians in the Colosseum is briefly sketched by Amartya Sen in "Rights and Agency," *Philosophy and Public Affairs*, 11, 1 (Winter, 1982); reprinted in Samuel Scheffler, *Consequentialism and its Critics* (Oxford: Oxford University Press, 1988): 191.

which someone has the choice of either (a) giving 1000 units of value to a very good man, or else (b) giving 1001 units of value to a very bad man.<sup>6</sup> Ross points out that utilitarianism implies (given the standard set of provisos) that it is morally obligatory that the value be given to the bad man, since the total value then enjoyed would be slightly greater. Ross believes that the implications of utilitarianism are mistaken.

Reflection on this case may clarify what Rawls must have had in mind when he said that “utilitarianism does not take seriously the distinction between persons.” In Ross’s example, some benefits must be distributed. One way of distributing the benefits would be on balance best, but that distribution gives all the benefits to the less deserving man. Utilitarianism pays no attention to facts about past behavior and meritorious character. It requires that the benefit be given in whatever way will be best, regardless of the history and character of the recipients. In this way, it may be said to fail to take seriously the distinction between people.<sup>7</sup>

Certain traditional forms of consequentialism are refuted by this sort of objection. However, I am convinced that the objection does not reveal any defect in the basic consequentialist insight—that we ought behave in such a way as to make the world as good as we can make it. As I see it, what the objection reveals is that there are defects in the axiologies traditionally associated with consequentialism. I want to show that it is possible to construct an axiology that is sensitive to matters of justice. If we combine consequentialism with such an axiology, we can maintain the core insight of consequentialism while rebutting the objection from justice.

In this paper, I first present a fairly typical form of consequentialism and I show in somewhat greater detail how it goes wrong with respect to justice. I then sketch a novel value theory that takes explicit account of justice. I attempt to show that when we combine consequentialism with the new value theory, we get more palatable results in the problem cases. In a final section, I discuss some objections.

2. *Consequentialism: A Preliminary Statement.* Typical forms of consequentialism are based on the idea that on each occasion of moral choice, there are several possible acts available to the agent. These are his “alternatives” on that occasion. For each alternative, there is a “total consequence.” This is the

---

<sup>6</sup> *The Right and the Good* (Oxford: Oxford University Press, 1930); reprinted in Pojman, op. cit., p. 258.

<sup>7</sup> Although it is not clear that it was originally presented in order to establish this point, Williams’ case of Jim and the Indians could be used to illustrate the problem about justice. See Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973): 77–150. Parfit’s notorious population puzzles might also be viewed from this perspective. See Part Four of Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

combination of all the things that would happen as a result if the alternative were performed.<sup>8</sup>

In order to state the bare bones of a typical consequentialist theory, we need a ranking of total consequences in terms of intrinsic value.<sup>9</sup> Let us assume for a moment that this has been given: for each possible total consequence, there is a number indicating the total intrinsic value of that consequence. Let the numbers be assigned in the standard way with higher numbers representing better outcomes, and say that an act “maximizes intrinsic value” if and only if no alternative has a total consequence with greater intrinsic value.

We can now formulate the central principle of this form of consequentialism:

C: An act is morally right if and only if it maximizes intrinsic value.<sup>10</sup>

In order to give real substance to the theory, we must add an axiology, or theory of value. This will specify what it is about a total consequence that gives it its intrinsic value.

In virtue of its simplicity and familiarity, let us provisionally make use of a traditional form of hedonism.

According to this view, the fundamental bearers of intrinsic value are episodes of pleasure and pain, where an episode is an event that consists in someone’s feeling some amount of pleasure (or pain) for some stretch of time. Let us also assume that for each episode of pleasure or pain there is a number indicating the amount of pleasure or pain contained in the episode. We assign the numbers in the standard way, so that we can evaluate any episode of pleasure or pain in terms of its “hedonic level”—a measure of the amount of pleasure or pain in that episode.

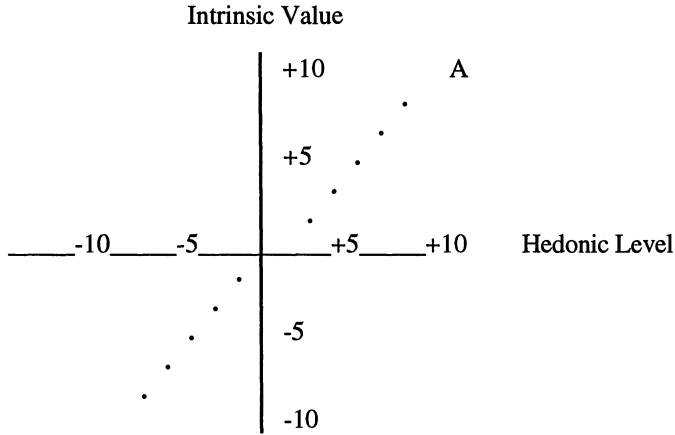
---

<sup>8</sup> For present purposes, I presuppose a traditional conception of alternatives and consequences. A particularly clear presentation of this sort of view can be found in Lars Bergstrom, “Utilitarianism and Alternative Actions,” *Noûs* 5 (1971): 237–52. In my *Doing the Best We Can* (Dordrecht: Reidel, 1986) I tried to show why this approach to actions and alternatives leads to difficulties. A more refined version of the view presented here would follow the pattern of MO in *Doing the Best We Can*.

<sup>9</sup> I am not proposing an account of the “essence of consequentialism.” Very many normative theories that I would classify as consequentialist do not make use of rankings of total consequences in terms of intrinsic value. For example, consider the form of utilitarianism that says we ought to maximize preference satisfaction. One could maintain this view and insist that nothing has intrinsic value. Another case is supplied by egoism. A typical form of egoism tells me to maximize my own welfare—it says nothing about intrinsic value. Finally, expected utility utilitarianism tells us that our moral obligation is to maximize expected utility, not intrinsic value.

<sup>10</sup> For a more rigorous formulation of the central consequentialist doctrine, see the discussion of MO in *Doing the Best We Can*. It would be interesting to compare C to Moore’s “ideal utilitarianism” as presented in *Principia Ethica*.

According to this simple form of hedonism, the intrinsic value of an episode of pleasure or pain is a function of its hedonic level. We can display the function in a simple graph. The North-South axis of the graph indicates amounts of intrinsic value, with positive numbers representing intrinsic goodness, and negative numbers indicating intrinsic evil. The East-West axis indicates hedonic level, with amounts of pleasure (increasing to the East) and pain (increasing to the West). The import of the graph is this: the intrinsic value of an episode of pleasure or pain is equal to the hedonic level of that episode.



Graph A

Now we can state a principle concerning the intrinsic value of a whole consequence. According to this view, the intrinsic value of a consequence is the sum of the intrinsic values of the episodes of pleasure and pain that occur in that consequence. Let us call this axiology 'H' (for 'hedonism'). When we combine H with principle C we get perhaps the most familiar form of consequentialism—hedonic act utilitarianism.

3. *Three Clean Cases.* Some of the examples discussed in the literature involve puzzling and controversial combinations of extraneous moral issues, and so their impact is not entirely clear. For purposes of illustration it would be better to have "cleaner," less complicated cases. Let us attempt to construct some simple cases of this sort.

Suppose I am required to give away a ticket that will entitle its bearer to a free lunch. Suppose I can give this ticket either to A or to B. Each of them would enjoy a free lunch, getting (let us say) ten units of pleasure from the experience. Suppose, furthermore, that B would be slightly disappointed if A were to get the ticket. A, on the other hand, would be somewhat more disappointed if B were to get the ticket. Suppose that A and B are in all relevant

respects quite similar, except that A has already enjoyed hundreds of free lunches, whereas B has never gotten even one. Suppose, finally, that no third party would be affected by my choice of A or B.

In this case, which we may call the *First Free Lunch*, C + H implies that the ticket should go to A. Here is a chart that helps to explain why:

---

<i>Actions</i>	<i>Value for A</i>	<i>Value for B</i>	<i>Total value</i>
Give ticket to A	+10	-1	+9
Give ticket to B	-2	+10	+8

---

Chart One

Provided that A would be slightly more disappointed than B by failure to receive the ticket, and A and B would be equally pleased to receive the ticket, and no one else would be affected by the distribution of the ticket, total intrinsic value is maximized by giving the ticket to A, even though he has already enjoyed far more free lunches than B. C + H then yields the result that it is my moral obligation to give the ticket to A. But this result is counterintuitive. Surely it would be more just to let B enjoy at least one free lunch—after all, B is in all relevant respects just like A except that he has so far had no free lunches, and A has had hundreds of them. Utilitarianism seems to ignore an important difference between people.

The *Second Free Lunch* is very similar to the first. Again I am required to give out a lunch ticket. Again I have my choice of giving it to A or giving it to B. Again each would enjoy the lunch, and each would be disappointed to fail to get the lunch—A a bit more than B. Let us imagine this time that A and B are alike with respect to past receipt of tickets for free lunches. However, in this case let us imagine that A has stolen and destroyed hundreds of lunches. Hundreds of decent people have gone hungry as a result of A's malicious thievery. B on the other hand is a decent fellow who has never stolen anyone's lunch.

We may assume that the numbers in Chart One apply again in this case. Again, C + H implies that it is my moral obligation to give the ticket to A. And again the implications of the theory conflict with my moral intuitions. In light of A's miserable past behavior, it seems to me that the ticket should go to B in spite of the fact that utility would not thereby be maximized. There is a "distributional impropriety" in giving the free lunch to A when he has so wantonly destroyed the lunches of others. Therefore, as I see it, the *Second Free Lunch* reveals another way in which C + H fails to take account of differences between people.<sup>11</sup>

---

<sup>11</sup> The *Second Free Lunch* is in some important ways quite similar to Ross' example involving the good man and the bad man. It is also quite like an example Rescher pre-

Let us now imagine a third example. Suppose this time that B is the legal owner of a ticket for a free lunch. Now A comes along and steals the ticket. He justifies his theft by appeal to a set of facts much like those illustrated in Chart One. Since he (A) would be more disappointed by failure to get the free lunch, utility is maximized by his getting it. Thus, he claims, when he stole the ticket he was doing the right thing.

Given the expected set of background assumptions, C + H implies in the *Third Free Lunch* that it was morally right for A to steal the ticket and that it would be right for him to get the lunch. And again it seems perfectly clear to me that the implications are false. If such a case were to occur, it would be unjust for the ticket and the free lunch to go to A. This injustice is sufficient, it seems to me, to make it morally wrong for A to steal the ticket. In this case the relevant difference between A and B is that B owns the ticket. C + H seems to ignore this fact.

I mentioned a number of examples at the outset. These include the Small Southern Town, the Colosseum, and the Organ Harvest. Although these cases are vastly more complex, each of them is in certain ways similar to the cases I have described. In every case, harms and benefits must be distributed in one of several ways. One distribution is stipulated to maximize utility. Unfortunately, that distribution also seems to involve substantial injustice. If the injustice is great enough, C + H yields the wrong moral judgment.

4. *A New Proposal.* Reflection on such examples leads some philosophers to say that consequentialism is false. As they see it, the examples reveal that sometimes, because of its injustice, the best outcome is not the one we ought to produce. But there is another way to interpret these cases. Following Brentano and others, we might take them to show that there is something wrong with the axiologies traditionally associated with consequentialism.<sup>12</sup> A different axiology—one sensitive to justice and injustice—might imply that

---

sents on p. 47 of his *Distributive Justice*. Rescher's example involves a case in which someone has the choice of conferring a certain benefit on a man who is "the very personification of virtue" or else conferring a slightly larger benefit on someone who is "the embodiment of vice."

<sup>12</sup> In his essay "Loving and Hating," Franz Brentano presents a complex axiology, designed to accommodate a variety of moral intuitions, including intuitions about the value of justice. In one passage, he indicates that he thinks that the world is not made better when people receive goods they don't deserve. In this respect, my view is like Brentano's. See Franz Brentano, *The Origin of Our Knowledge of Right and Wrong*, ed. by Oskar Kraus, English edition ed. by Roderick Chisholm and Elizabeth Schneewind (London: Routledge and Kegan Paul, 1969): 149. In his "Rights and Agency," *Philosophy and Public Affairs* 11, No. 1 (Winter, 1982), reprinted in *Consequentialism and its Critics*, ed. by Samuel Scheffler (Oxford: Oxford University Press, 1988): 187–223, Amartya Sen discusses "goal rights systems" of consequentialism. Such systems are in important ways similar to the sort of system I present here. More recently, David Sosa defended a similar view in "Consequences of Consequentialism," *Mind* 102, 405 (January, 1993): 101–22.

the acts that seem to be morally right in these examples have consequences that are in fact better than the consequences that would have been produced by their alternatives. And the greater value might arise, on the new axiology, from the amount of justice in the consequence.

In an attempt to develop these intuitions about the place of justice in axiology, I want to formulate an axiology that starts with hedonism, but incorporates the idea that the value of a pleasure or pain may be increased or decreased depending upon whether it is justly or unjustly experienced.<sup>13</sup> When we combine the new axiology with C, we will get a consequentialist normative theory that deals more successfully with the problems about justice.

My conception of justice is based on the ancient and plausible idea that justice is done when people receive goods and evils according to *desert*.<sup>14</sup> The closer the fit between desert and receipt, the more just the outcome. Other things being equal, the more just outcome is the better. Before turning to the formulation of the axiology, I should say a few words about desert.

I will not be able to provide any analysis of the concept of desert.<sup>15</sup> It functions here as a conceptual primitive. However, the idea should be familiar. We often hear people say that a particularly vicious criminal “deserves to have the book thrown at him”; or that a sick child “does not deserve to suffer in this way,” or that those who have labored in the background “deserve their day in the sun.” Roughly, to say that a person deserves some good is to say that it would be “distributionally appropriate” for him to get it. I assume that desert is a matter of degree, so that it makes sense to say that a certain person deserves a certain pleasure or pain *to a certain extent*. I will represent the various extents with numbers, in the usual way.

It is important to note that desert is not defined here in terms of moral obligation. The statement that a person deserves some good is not equivalent to the statement that she ought to get it, or to the statement that someone else ought to provide it. In some cases, although someone deserves some good, no one is in position to provide it; in other cases, for various reasons,

---

<sup>13</sup> In Chapter 11 of *Confrontations with the Reaper* (Oxford: Oxford University Press, 1992) I sketched an axiology that is similar to this one. There, however, in an attempt to find a consequentialist account of the morality of killing, I considered the idea that life itself might be an intrinsic good, and that each person deserves not only some measure of pleasure, but some measure of life.

<sup>14</sup> Aristotle claims that one central sort of justice is “a species of the proportionate.” His idea seems to be that distributions of goods and evils are just provided that they are proportional to the merits of the recipients. See *Nicomachean Ethics* V.3. In *Distributive Justice* (Bobbs-Merrill: New York, 1966), Nicholas Rescher defends a similar view, and cites supporting passages from Plato, various Roman jurists, Sidgwick, John Hospers, and others. Although there are important differences between Rescher’s approach and mine, I have benefitted enormously from his work—especially his very extensive bibliography on utilitarianism and justice.

<sup>15</sup> For an extensive discussion of the concept of desert, see *Desert*, by George Scher (Princeton University Press: Princeton, 1987).

some other considerations override the demands of justice. So the statement that someone deserves something does not entail the statement that she morally ought to get it. Equally, the statement that someone ought to get a certain good does not entail that she deserves it. If the results would be good enough, there could be a case in which someone morally ought to give a certain good to a certain recipient, even though the recipient does not deserve it.

Many different factors influence the extent to which a given person deserves a certain pleasure or pain. One of these is *excessive or deficient past receipt*. Excessive past receipt lowers your desert level for a good; deficient past receipt increases it. To see how this works, imagine a case in which there are two possible recipients for some good. Suppose the potential recipients are alike in all relevant respects except that one of them has already received far more of that good than the other. Then, since other things are equal, the one who has so far been short-changed has greater desert. His desert level for the good is greater than the desert level of others who have already received more.

A recipient's *moral worthiness* may have an impact on his desert level, too. Suppose two potential recipients of good are alike in all other relevant respects, except that one of them has been good, whereas the other has been bad. Then the one who has been good has greater desert. If no other factors complicate the case, it would be more just for the good to be given to him.

A third factor is based on rights and claims. Other things being equal, someone with a *legitimate claim* on some good has greater desert relative to that good than does someone with less claim on it. Thus, if two recipients are in other respects similar, but one legally owns the means to a certain pleasure and the other does not, then the owner deserves more to have the pleasures arising from ownership of that object.

There are undoubtedly other factors that may influence the extent to which a person deserves some pleasure or pain, and in a full exposition of the theory of desert, each of them would be described in detail.<sup>16</sup> Furthermore, in real-life cases several of the factors may be jointly operative. The ways in which the factors clash and harmonize so as to yield an overall desert-level must also be investigated. However, since my aim here is primarily to show that it is possible to formulate an axiology that takes account of justice, and to show

---

<sup>16</sup> In *Confrontations with the Reaper*, I discussed the idea that a person's desert level relative to a certain good might be influenced by the amount of time and effort she has invested in acquiring that good. Thus, if two people are similar in other respects, but one of them has worked hard to cultivate a garden, whereas the other has done nothing in the garden, then the hard-working gardener has greater investment. She deserves more to enjoy the benefits deriving from the garden.

In *Distributive Justice* (Bobbs-Merrill: New York, 1966), Nicholas Rescher discusses seven "canons" of distributive justice. Each of these canons corresponds to a possible source of desert. See pp. 73–83.

how such a theory can be combined with consequentialism to rebut the objection from justice, I shall not pursue these lines of inquiry here.

According to the axiology I want to defend, the intrinsic value of an episode of pleasure or pain is a function of two variables: (i) the amount of pleasure or pain the recipient *receives* in that episode, and (ii) the amount of pleasure or pain the recipient *deserves* in that episode. Roughly, the theory maintains that pleasure is generally intrinsically good; but it is better if it is fully deserved, and it is less good if it is not deserved. In extreme cases, if it is undeserved, it may be worth much less—indeed, it may be worthless or even bad. Pain, on the other hand, is generally intrinsically bad. However, it is even worse when the person who suffers it does not deserve it. It is less bad—and may even be good—if the person who suffers it fully deserves it.

Part of the content of the axiology may be expressed by some principles about desert and receipt. Each principle governs a class of cases involving a range of receipt and desert levels. One principle governs cases in which a person experiences some pleasure when she deserves it. It seems to me that it is especially good if someone who has “positive desert” gets to experience precisely the pleasure she deserves. Following Chisholm and Moore, I use the term “enhancement” to indicate this axiological phenomenon.<sup>17</sup> This is the first principle of “just deserts”:

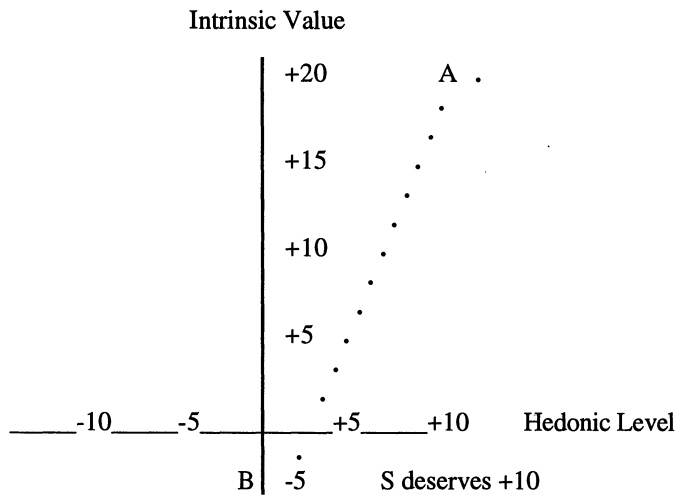
P1: Positive desert enhances the intrinsic goodness of pleasure.

Part of the import of P1 is illustrated in Graph B. In Graph B, we focus on the case of someone who deserves exactly 10 units of pleasure. The graph displays the outputs of intrinsic value generated by the function for various inputs of pleasure. However, this graph applies only to the case of a person who deserves to be getting 10 units of pleasure and so gives only a suggestion about cases involving people with other desert levels.

Point A on the graph indicates that when a person deserves 10 units of pleasure and gets 10 units of pleasure, then the episode as a whole has an intrinsic value of +20. Since justice is done when the deserving get what they deserve, intrinsic value is enhanced. Notice that the curve begins to flatten out as it moves northeast. This shows that as a person begins to receive more than she deserves, additional increments of pleasure have decreasing marginal intrinsic value. Again, this is a matter of justice. Point B is of interest, too. It indicates that it is intrinsically bad for a person who deserves 10 units of pleasure to get nothing instead.

---

<sup>17</sup> See *Brentano and Intrinsic Value*, pp. 85–87. Moore speaks of enhancement a few times in Chapter VI of *Principia Ethica*. See, for example, pp. 211 and 213.

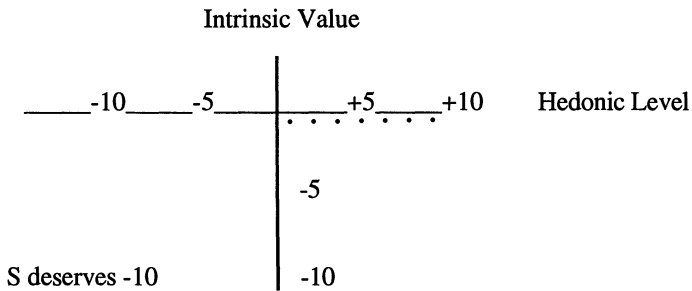


Graph B

The second principle concerns what happens when someone who deserves pain gets pleasure instead. If a person with negative desert enjoys pleasure, justice is not done, and the value of that pleasure is diminished, or “mitigated.”<sup>18</sup> The general principle here is:

P2: Negative desert mitigates the intrinsic goodness of pleasure.

Graph C illustrates one possible view about how negative desert might mitigate the value of pleasure. Again, we consider only one class of cases: the cases in which the recipient deserves exactly 10 units of pain. The graph shows outputs of intrinsic value for various inputs of pleasure:

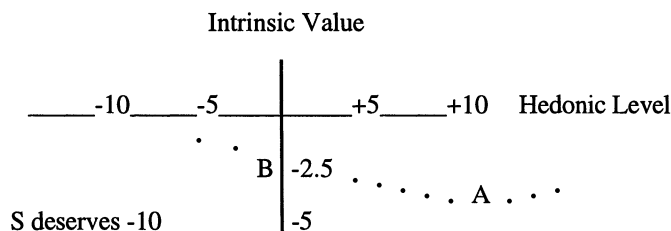


Graph C

<sup>18</sup> Chisholm calls this phenomenon “defeat.” See, for example, *Brentano and Intrinsic Value*, p. 83.

The significance of the graph is this: when a person has negative desert, positive receipt is worthless. In such cases, pleasures—of any intensity—have no intrinsic value.

According to a somewhat more extreme view, when a person on the whole deserves pain, but receives pleasure instead, then the injustice of the situation is so great as to make it intrinsically bad.<sup>19</sup> If we accept this view, we will want to go beyond saying merely that negative desert *mitigates* the intrinsic goodness of pleasure. We will want to say that in some cases, negative desert “transvaluates” the goodness of pleasure. By this I mean to indicate that if a person’s desert level is negative enough, positive receipt becomes intrinsically bad.<sup>20</sup> We can illustrate one version of this view as follows:



We have so far considered two sorts of cases: those in which a person with positive desert receives pleasure, and those in which a person with negative desert receives pleasure. What about the case in which a person has “neutral” desert—he neither deserves pleasure nor deserves pain? In such cases the value of the pleasure is neither enhanced nor mitigated. Hence, the intrinsic value of an episode of pleasure of this sort is directly proportional to the amount of pleasure it contains. The principle here is:

P3: Neutral desert neither enhances nor mitigates the intrinsic goodness of pleasure.

The fourth principle concerns the case in which someone deserves pleasure, but gets pain instead. As I see it, pain is bad enough; but when someone who deserves pleasure gets pain instead, the evil of that pain is made

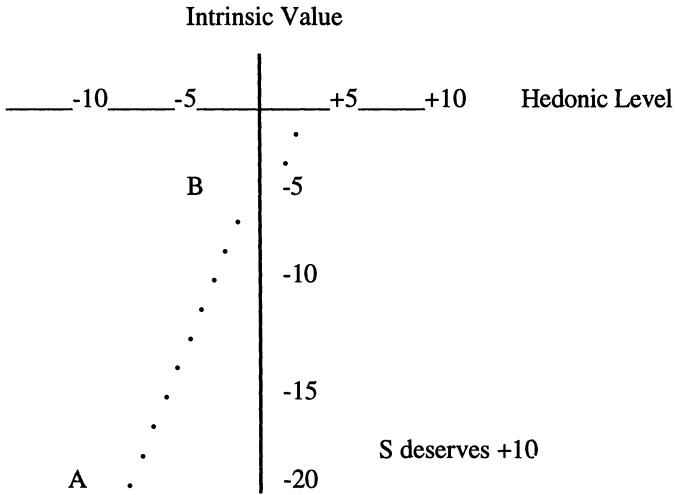
<sup>19</sup> In *The Origin of Our Knowledge of Right and Wrong*, p. 23, Brentano says, “Pleasure in the misfortunes of others (Schadenfreude) is bad because [it is not a correct emotion].” In *Brentano and Intrinsic Value*, p. 76, Chisholm says that Meinong also used the term “schadenfreude” to indicate joy in another’s pain. Following Brentano and Meinong, Chisholm maintains that joy is good, but joy in another’s pain is neutral in value. In *Principia Ethica*, pp. 208–9, Moore says that “The first class [of great positive evils] consists of those evils which seem always to include an enjoyment or admiring contemplation of things which are themselves either evil or ugly.” He goes on to cite cruelty and lasciviousness as examples.

<sup>20</sup> Chisholm, *Brentano and Intrinsic Value*, p. 83.

The fourth principle concerns the case in which someone deserves pleasure, but gets pain instead. As I see it, pain is bad enough; but when someone who deserves pleasure gets pain instead, the evil of that pain is made even worse, or (to use another Chisholmian term) “aggravated,” by its injustice:

P4: Positive desert aggravates the intrinsic badness of pain.

P4 is the principle that stands behind the axiological intuition that it is worse for bad things to happen to good people than it is for bad things to happen to bad people. Graph D illustrates one view about how positive desert might aggravate the evil of pain.



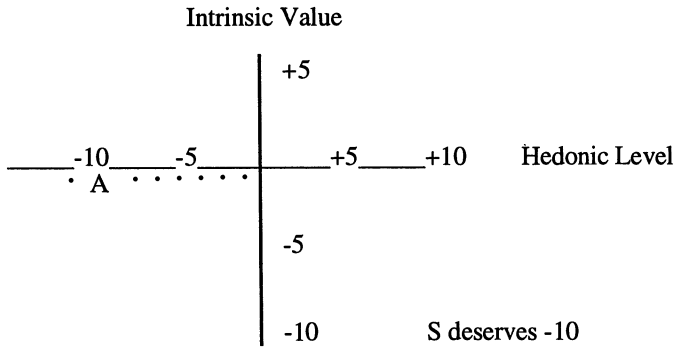
Graph D

Graph D applies only to cases in which someone deserves 10 units of pleasure. It illustrates a view about how this positive desert makes it especially bad for such a deserving person to get pain. Note, for example, Point A. This indicates that when a person deserves +10, but gets -10 instead, the whole episode is so unfair that it has an intrinsic value of -20.

The fifth principle concerns the case in which someone who deserves pain gets pain. Pain is generally bad; but it is not so bad for it to be experienced by someone who deserves it. In this case, since justice is done, the badness of the episode has been mitigated.

P5: Negative desert mitigates the intrinsic badness of pain.

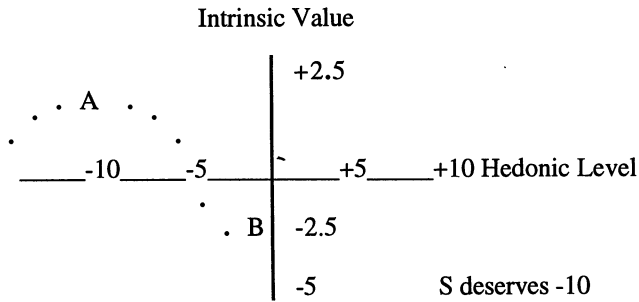
A graph illustrating one possible interpretation of P5:



Graph E

Point A indicates what happens when ten units of pain are received by a person who deserves exactly ten units of pain. Since in this case the person gets exactly what he deserves, justice is done. The evil of the pain is mitigated by the negative desert of the recipient. Although he suffers some pain, the world is not made worse. The intrinsic value of the episode is zero.

Some philosophers seem to believe that the world is made *better* when the guilty suffer precisely the harm they deserve.<sup>21</sup> This phenomenon may be called the “transvaluation of the evil of pain by negative desert.” One interpretation of this view is illustrated in Graph E’.



Graph E’

Point A on Graph E’ serves to illustrate the idea that it is slightly good (+2.5) for a person to receive ten units of pain when this is precisely what he

<sup>21</sup> In *The Metaphysical Elements of Justice*, Kant describes a case in which some isolated society is about to disband. One murderer is still in prison, yet to be executed. Kant seems to want to say that even though in a case like this there is no utilitarian justification for carrying out the sentence, still the criminal ought to be executed. If we approach the case from the perspective of the principle of the transvaluation of the evil of pain by negative desert, we might claim that the world is made better by the infliction of deserved pain on the murderer. Thus, though traditional forms of utilitarianism might not be able to explain it, a version of C + JH might be able to explain why retributive punishment is justified.

deserves. This expresses the retributivist axiological intuition that sometimes it is good for bad people to be punished. The curve slopes downward in both directions from Point A in order to illustrate the idea that it is not so good for a person who deserves pain to get either more or less pain than he deserves. This corresponds to the intuition that punishment must be proportional to the crime.

The views expressed in Graphs E and E' are reminiscent of views endorsed by Meinong, Brentano, Chisholm, and Moore. Brentano, for example, discussed the case in which someone takes sorrow in someone else's sorrow. Since sorrow is precisely the right reaction in such a case, Brentano declared the whole state of affairs to be good.<sup>22</sup> Moore's views about hatred of evil and ugliness are again similar.<sup>23</sup> In my view, each of these cases illustrates the same general thesis: the evil of pain is mitigated (and in some cases perhaps transvaluated) by negative desert.

The final principle concerns the case in which a person with neutral desert receives pain. As I see it, the intrinsic value of any such episode of pain is directly proportional to the amount of pain it contains. Thus the principle is:

P6: Neutral desert neither enhances nor mitigates the evil of pain.

According to "justice adjusted hedonism" (or "JH"), the fundamental bearers of intrinsic value are episodes of pleasure and pain. In each case, the intrinsic value of the episode of pleasure or pain is not a simple function of hedonic level. Desert plays an essential role, too. As a result, some pleasures are not good (as shown in Graph C) and some pains are not bad (as shown in Graph E). In every case, the justice-adjusted intrinsic value of the whole episode is a function of the receipt level and the desert level of the recipient. The intrinsic value of a whole consequence is the sum of the justice-adjusted intrinsic values of the episodes of pleasure and pain that occur in that consequence.<sup>24</sup>

---

<sup>22</sup> Brentano discusses this sort of case in "Two Unique Cases of Preferability," which is in the "Supplementary Notes" to *The Origin of our Knowledge of Right and Wrong*, English edition edited by Roderick Chisholm; pp. 90–91.

<sup>23</sup> Moore claims (PE, 217) that while the specific emotion of hatred is not intrinsically good, it is nevertheless intrinsically good for a person to hate what is evil or ugly. This is at least analogous to the idea that it is good to take pain in another's pain.

<sup>24</sup> In a number of places, Sen has argued that "sum ranking" is an important (and problematic) component of traditional utilitarianism. Roughly, the idea is that the total utility of an outcome is the sum of the utilities of the individuals in that outcome. See, for example, his Dewey Lecture "Moral Information" *The Journal of Philosophy*, LXXXII, 4 (April, 1985) 175. My summation principle goes further: the total utility of an outcome is the sum of the utilities of the *basic utility-bearing states of affairs* in that outcome. A number of philosophers, influenced perhaps by Rawls and Bernard Williams, have claimed that this is the crucial element that obliterates the distinctions between people. I have attempted to retain an extreme form of sum ranking, while at

If we combine C with our new axiology, the resulting normative theory (“C + JH”) is sensitive to matters of justice. It takes account of the enhancing, aggravating, mitigating, and perhaps transvaluating factor of desert. My claim is that the resulting theory generates correct results in the problem cases.

5. *Reevaluation of Puzzle Cases.* In Section 3 I presented three cases in which C + H generated incorrect normative judgments. In order to apply C + JH to these cases, we have to recalculate the intrinsic values of the relevant consequences according to JH.

In the *First Free Lunch*, there were two potential recipients of a ticket for a free lunch. One recipient, A, had already received hundreds of such tickets. The other, B, had received none. But it was stipulated that A would suffer slightly greater disappointment if he failed to get the ticket. C + H implied that it was my obligation to give the ticket to A, and this seemed mistaken. Now let us consider what C + JH implies for the *First Free Lunch*.

Chart Two contains some numbers illustrating the application of the new axiology to the example:

---

<i>Actions</i>	<i>Value for A</i>	<i>Value for B</i>	<i>Total value</i>
Give ticket to A	+8	-2	+6
Give ticket to B	-2	+20	+18

---

Chart Two

If A gets the ticket, he will experience ten units of pleasure. But in virtue of his excessive past receipt of tickets, his desert level is low—perhaps even negative. He doesn’t deserve the pleasures derivable from more free lunches. As a result, the intrinsic value of his pleasure is mitigated. This is the import of principle P2. I have accordingly reduced the value of A’s pleasure from +10 to +8.

If the ticket goes to A, then B will be disappointed again. He will suffer one unit of pain. But B does not deserve this pain. In virtue of his deficient past receipt, he deserves the pleasures of a free lunch. In accordance with Principle P4, it follows that the evil of B’s pain is aggravated by his positive desert. I have therefore represented the value of B’s pain as -2 rather than as -1. The justice-adjusted intrinsic value of giving the ticket to A is +6 (rather than +9 as before).

Now let’s consider the results of giving the ticket to B. If B gets the ticket, A will suffer 2 units of pain. The happiness of B, if he gets the ticket,

---

the same time formulating the axiology in such a way as to respect the morally important distinctions between people.

is fully deserved. After all, he has had far less than his fair share of free lunches, and has suffered far more than his fair share of disappointment. According to Principle P1, the value of his happiness is enhanced by his positive desert. Following the suggestion of Graph B, I have assigned it a value of +20.

So the act that intuitively seems to be morally right—giving the ticket to B—also has better results according to JH. Thus, an example that refuted C + H does not refute C + JH.

The treatment of the *Second Free Lunch* and the *Third Free Lunch* according to JH is similar. In each of these cases JH implies that the best outcome is the one in which the ticket goes to the more deserving B. In the *Second Free Lunch*, this follows from the stipulated fact that A has behaved miserably, and has thereby lowered his desert level. In the *Third*, it follows from the fact that B is the legitimate owner of the ticket: he deserves the pleasures arising from ticket ownership; A does not.

I noted earlier that Rawls claimed that “utilitarianism does not take seriously the distinction between persons.”<sup>25</sup> As I understand it, the point of this remark is that utilitarianism advocates the distribution of goods and evils in whatever way will maximize total utility—with no regard for the character or the past behavior of the various recipients. Recipients of good and evil function merely as “vessels” into which value may be poured. The theory implies that the value should be poured out in whatever way will yield the greatest total. It should be clear, however, that this charge cannot be leveled against the axiology proposed here. The value of a given distribution of goods and evils depends crucially on the extent to which each recipient deserves his or her share. Past receipt and character play central roles in the determination of desert. Hence, the theory recognizes and pays careful heed to the morally relevant differences between persons.

Although I will not attempt to show it here, I believe that the other puzzle cases mentioned at the outset pose no special threat to C + JH. In each of those cases, C + H implied that some grave injustice should be committed. In some of those cases C + JH delivers normative results different from, and more palatable than those delivered by C + H. In such cases, once we adjust utility for justice, it will turn out that the best consequence is the one in which no injustice is committed. Thus, as I see it, C + JH does not have the defect that Rawls had in mind in the passage cited.

It is important to keep in mind that the actual implications of the theory in these cases would depend upon details about the receipt levels and desert levels of the participants in all the outcomes. In some cases C + JH will imply that a serious injustice is required in order to assure the best outcome, and so (sadly) ought to be committed. The theory does not imply that justice

---

<sup>25</sup> Rawls, *op. cit.*, p. 27.

must always be maximized; it implies that justice-adjusted intrinsic value must always be maximized.

6. *Some Objections.* A critic might object to C + JH, claiming that it is covertly circular. The circularity, it might be alleged, arises in this way: first we purport to explain the normative concept of *moral obligation* by appeal to the value-theoretic concept of *intrinsic betterness*. Then when it comes time to give substance to the concept of intrinsic betterness, we appeal to the concept of *justice*. Justice is explicated in terms of *desert*. And desert, finally, is explained by saying that a person deserves some pleasure or pain precisely when he *ought* to get it—and thus the circle is completed. We have made covert use of the concept of moral obligation in our attempt to explain moral obligation.

This objection turns on a misunderstanding of the proposal. The objection goes wrong at the final step, where it is alleged that we explain desert by appeal to moral obligation. I endorse no such explanation. The concept of desert is not defined by appeal to the concept of moral obligation. Indeed, it is not defined in any way here. It functions as a conceptual primitive for purposes of the present theory.

In my discussion of desert, I have made liberal use of evaluative terminology. Thus, I have spoken of “fair shares,” “worthiness of receipt,” “moral rights,” and “fittingness.” However I have tried to avoid use of the concepts of moral rightness, wrongness, or obligatoriness. The statement that a person *deserves* some pleasure must not be confused with the statement that *it would be morally right* for him to get that pleasure. In order for the theory as a whole to avoid circularity, it is important that the concept of desert is not explicated by appeal to the concepts of right and wrong. For, as I see it, desert plays a role in the determination of justice, and justice plays a role in the determination of the intrinsic value of outcomes, and the intrinsic value of outcomes plays a central role in the determination of right and wrong.

My approach involves a somewhat unorthodox view about the place of the theory of justice in moral philosophy. Traditionally, the theory of justice has been thought to belong in normative ethics. I have always been puzzled about its implications for action. Suppose the theory of justice firmly establishes that some sort of behavior is unjust. Is it supposed to follow that we should not engage in that sort of behavior? If so, the theory of justice seems to have invaded the turf of ordinary normative ethics. If not, the import of the theory of justice is obscure.

My approach firmly locates the theory of justice in axiology. Our axiology determines the value of each outcome, in part, by the justice or injustice of the distribution of pleasures and pains within that outcome. Justice and injustice, in turn, are understood to be determined by reflections on the quality of the fit between pleasures and pains received and pleasures and pains de-

served in each outcome. Thus, as I see it, a full-blown axiological theory would have to include an account of justice. Normative ethics then takes the information provided by axiology, and generates prescriptions for right action. In my view, the prescription is very simple: behave in such a way as to make the world as good as you can.

I turn now to a second objection. It might be claimed that even if the theory presented here is acceptable, it cannot possibly figure in a defense of *consequentialism*—for the theory is not a form of *consequentialism*. The problem, according to the objection, is that my proposed axiology ascribes intrinsic value to complex states of affairs. The values of these states of affairs turn on normative features, such as desert and justice. No consequentialist theory, it might be claimed, makes use of normative features in this way.

In reply to this objection, I would first want to insist that C + JH is as much a form of consequentialism as the hedonistic utilitarianism of Mill and the pluralistic “ideal” utilitarianisms of Brentano and Moore. In each of these cases, the axiology is complex. In the case of Mill, the axiology is based on “higher” and “lower” pleasures, which allegedly have higher and lower intrinsic values. In the cases of Moore and Brentano, the axiology is based on such complex states of affairs as “the love of good” and “the hatred of evil.” For Moore and Brentano, the value of a state of affairs turns crucially on normative features internal to that state of affairs. If Mill and Moore and Brentano are consequentialists, then so am I.

I can go beyond this merely historical point. I think the definition of “consequentialism” is a matter of great controversy. The literature is filled with incompatible (and sometimes idiosyncratic) definitions.<sup>26</sup> It would be hard to find any characterization that would meet with universal approval. Perhaps it would be interesting to consider a typical, and obviously relevant characterization. In *The Theory of Justice*, Rawls gives his own account. He says:

The structure of an ethical theory is, then, largely determined by how it defines and connects these two basic notions [the right and the good]. Now it seems that the simplest way of re-

---

<sup>26</sup> Moral philosophers have widely divergent views about the essence of consequentialism. Some say that consequentialist theories must require the *maximization* of some value; others permit mere *satisficing*; and others merely say that such theories claim that rightness depends in some way on the values of consequences. Some say that consequentialist theories all agree that the normative status of actions depends upon *intrinsic values* of consequences; others leave intrinsic value entirely out of account, and mention such things as preference satisfaction, or happiness. Some say that consequentialist theories must be theories about the moral normative status (rightness, wrongness, obligatoriness) of *actions*; others say that such theories may concern other sorts of value (rationality, goodness, justice, etc.) of other sorts of entities. Parfit goes so far as to say that a form of consequentialism might evaluate eye color, or even climate! “...the best possible climate is the one that would make outcomes best.” (*Reasons and Persons*, p. 25.)

lating them is taken by teleological theories; the good is defined independently from the right, and then the right is defined as that which maximizes the good.<sup>27</sup>

Inspection will reveal that C + JH conforms to this characterization of teleological theories. I admit, of course, that in my "definition" of the good, I have made numerous references to matters involving justice. But, as I have already emphasized, my account of justice does not appeal to the central normative concepts of *right* and *wrong*. So far as I can tell, my characterization of the good is independent of my characterization of the right.<sup>28</sup>

---

<sup>27</sup> Rawls, *op. cit.*, p. 24.

<sup>28</sup> Earlier versions of this paper were presented at Washington University, the Georgia Philosophical Society and the Universities of Connecticut, Rochester, West Virginia, and Missouri. It was discussed in faculty seminars at Arizona State University and St. Cloud State University. I am grateful to the participants in all these discussions for their valuable criticisms and suggestions. Especially useful comments were made by John Troyer, John Bennett, Ed Weirenga, Richard Feldman, Shelly Kagan, Ned Markosian, Sharon Ryan, David Blumenfeld, Peter Markie, David Cowles, Peter de Marneffe, and John Bahde.

I am also grateful to Ted Sider, Owen McLeod, Ish Haji, Kevin Moon, Noah Lemos, Neil Schaefer, Bob Frazier, and two anonymous referees for *Philosophy and Phenomenological Research* for critical comments on earlier drafts. I have not had space to discuss all the objections these critics have presented. I hope to do so elsewhere.

I first got the idea of presenting this sort of axiology by means of graphs many years ago in conversations with Michael Zimmerman. His paper "On the Intrinsic Value of States of Pleasure," *Philosophy and Phenomenological Research*, 41 (1980): 26–45 originally contained several very elegant, multi-colored graphs illustrating outputs of intrinsic value for various inputs of pleasure and value of object of pleasure. The published version of the paper does not contain the graphs, primarily because *Philosophy and Phenomenological Research* was, at the time, unable to reproduce them in print. In any case, I am grateful to Zimmerman.