1. Introduction

One of the standard objections to traditional act utilitarianism is that it is insensitive to issues of

justice and desert. Traditional act utilitarianism holds, for example, that it is morally obligatory to

torture or kill an innocent person, when doing so increases the happiness of others more than it

decreases the happiness of the innocent person. Utilitarianism is, of course, sensitive to what

people <u>believe</u> about justice (for example, people might riot, if they believe a gross injustice has

been done), but it is not sensitive to justice itself.

In response to this problem, Fred Feldman has recently developed a version of

consequentialism designed to deal more adequately with issues of justice.[1] He does this by

developing a theory of the good that is sensitive both to individual welfare and to what people

deserve. On his theory, the goodness of states of affairs is determined by the total amount of <u>desert-

adjusted</u> welfare.

I agree with Feldman that the permissibility of actions, and the moral desirability of worlds,

depends both on people's welfare and on issues of justice. Feldman's work in this area is important,

because it explores terrain that has been largely ignored by consequentialists.[1] I shall argue,

however, that his theory is implausible because it doesn't take welfare promotion seriously enough.

---

1. W.D. Ross, of course, developed a theory on which both welfare promotion and justice are morally relevant. That theory, however, is not a complete account of the good (or the right), because those considerations are considered only prima facie relevant.

In his article in this issue[2], Feldman argues that his theory has the resources to block the Repugnant Conclusion (the view that a world with a given number of people with at least reasonably good lives is worse than a world with sufficiently more people with lives barely worth living). I shall argue, however, that he avoids the Repugnant Conclusion only by making his theory even more implausible.

2. Background

Strictly speaking, the concept of justice involves more than desert. One might hold that justice requires that one's historical entitlements (e.g. à la Nozick) be respected, even where one does not deserve the entitled benefits. I shall follow Feldman, however, and use "desert" for the generic notion of whatever is relevant for justice. And for the sake of argument, I shall assume, with Feldman, that the relevant desert considerations are based on (a) past benefits and burdens, (b) past good and bad actions, (c) past agreements and historical entitlements, and (d) meritorious character. These considerations, we shall assume, determine a person's deserved welfare.

Also following Feldman, we'll assume that both welfare and desert are interpersonally cardinally measurable with absolute (i.e., non-arbitrary) zeros. Although a version of Feldman's theory could be formulated on the basis of ordinal notions of welfare and desert that are interpersonally comparable only in terms of their levels (with no absolute zeros), this theory would lack most of the features that Feldman takes to be important. So, we'll join him in assuming the strong measures of welfare and desert.

There are actually two relevant notions of desert—absolute and marginal—and Feldman's

presentation fails to distinguish these adequately. <u>Absolute (positive) desert</u> is the highest level of welfare that is fully deserved. <u>Marginal desert</u> is the desert for a particular increment in welfare (e.g., from 1 to 2). It would be natural to assume that there is decreasing marginal desert. For example, if a person's absolute desert level is 5, it would be natural to assume that marginal desert for an increment of welfare from 0 to 1, is greater than the marginal desert for an increment in welfare from 4 to 5. If we assume interpersonal comparisons, then this means that if there are two people with equal absolute desert levels (e.g. each deserves 5), but one person is at 0 welfare and the other at 4, then the first person has a stronger desert claim for an increment in one unit of welfare.

Marginal desert is the more general notion. For absolute desert is simply the welfare level at which marginal desert becomes zero. Positive (negative) absolute desert is the greatest (least) welfare level such that increments above (decrements below) are not deserved. So, we can without loss of generality simply be concerned with marginal desert.

Feldman does not distinguish between absolute and marginal desert. He writes as if absolute desert (which is what he means by "desert") is the only relevant notion. I suspect that this is because he was implicitly assuming that marginal deserts are constant and proportional to absolute desert (e.g., for a person with an absolute desert level of 5, the marginal desert for the increment from 0 to 1 is the same as that for the increment from 4 to 5, and these are less than the corresponding marginal desert levels for someone whose absolute desert level is 6). Given this assumption, there is no need to distinguish between absolute and marginal desert.[3]

An adequate theory of desert needs to be sensitive to marginal desert, and it is implausible

to hold that for any absolute desert level marginal desert is constant. Fortunately, the core of Feldman's theory can be formulated so as to be sensitive to non-constant marginal desert. To do this, I will interpret his references to desert as references to marginal desert, and where relevant for illustration purposes, I will follow him in assuming that marginal desert is constant and proportional to absolute desert.

3. Feldman's Theory

Central to Feldman's theory is the distinction between welfare (what he calls "personal intrinsic value") and moral value (what he calls "world intrinsic value"). Both can be understood person-relatively. The welfare value of a world for a person is a measure of how good the world is for that person from that person's personal perspective. It is a non-moral notion, and is often interpreted in terms of happiness or preference satisfaction. The moral value of a world relative to a person is the value of the person's life in that world from the perspective of morality (or justice). Nice and Nasty might both have lives with 100 units of welfare, and yet the moral value of Nice's life might be greater than that of Nasty on the grounds that Nice deserves 100 and Nasty doesn't.

Feldman's theory has exactly the same structure has act utilitarianism. An action is judged permissible just in case it has maximally valuable consequences. Consequences are understood in terms of worlds that might come about. The value of worlds is determined by taking the total over people of the value of each person's life. The only difference between Feldman's theory and act utilitarianism is that his theory focuses on moral value (world intrinsic value) whereas utilitarianism focuses on welfare (personal intrinsic value).

Although I agree with Feldman that an adequate moral theory must be sensitive to issues of justice, I think it more plausible to incorporate justice directly at the deontic level via rights—instead of incorporating justice into the theory of the good. On this view, individual welfare should be promoted as much as possible, subject to the constraint that no one is to be made worse off than if no rights were violated. Obviously, I can't defend (or even explain!) this view here, so for the sake of argument I'm going to grant that justice is relevant for the theory of moral goodness.[4]

In what follows, then, I shall focus solely on Feldman's theory of moral value. And I shall grant that the moral value of a world is the sum of the moral values of the lives of the people in that world. The central question then is that of how the moral value of a person's life is determined. Feldman holds, and I shall grant, the following theses:

D: The moral value of a person's life in a world is his/her desert-adjusted welfare in that world.

DAW1: For a given change in welfare for a given person, the change in his/her desert-adjusted welfare is fully determined by his/her marginal desert for that increment.

DAW2: For a given increment in welfare for a given person, the greater the marginal desert of that individual for that increment, the greater the increment in desert-adjusted welfare for that individual.

5

DAW3: If a person has neutral/positive/negative marginal desert for a given welfare increment (decrement), then the <u>increment (decrement) in desert-adjusted welfare</u> is the same as/greater than/less than the <u>welfare increment (decrement)</u>.[5]

D just makes explicit that moral value is to be assessed on the basis of desert-adjusted welfare. DAW1 makes explicit that the change in desert-adjusted welfare induced by a given change in welfare is determined solely by that change in welfare and by the person's marginal desert for that change (nothing else is relevant). DAW2 makes explicit that desert-adjusted welfare is positively responsive to marginal desert: the greater the marginal desert, the greater the increase (decrease) that is induced by a given increase (decrease) in welfare. For example, an increase from 10 to 20 units in welfare increases desert-adjusted welfare more when the marginal desert is 2 than when it is 1. Finally, DAW3 connects desert-adjusted welfare to welfare. It says, for example, that increases in desert-adjusted welfare are (1) equal to the increase in welfare, when marginal desert is neutral, and (2) greater (less) than the increase in welfare, when marginal desert is positive (negative). For example, an increment in welfare from 0 to 10 increases desert-adjusted welfare 10 units if the person has neutral marginal desert, and more (less) than 10 if the person is positively (negatively) deserving of the increase.

If, as I am granting for the sake of argument, the moral value of worlds is determined by deserted-adjusted welfare, then the above claims are plausible. Feldman, however, goes well beyond these claims. Although he does not give us a full theory of moral value, he does sketch the

outlines of one. In Graph 1 and in his discussion, he endorses the general shape of the following formula for determining the desert-adjusted welfare (daw) of individual i, with welfare w, given that she has an absolute desert level D:

daw(w,i,D) = 2w - (D-w)/2, if D>w>0.

That is, for positive welfare and greater positive absolute desert, the desert-adjusted welfare is equal to (1) twice the amount of the welfare, minus (2) half the shortfall of that welfare from the absolute desert level (D-w). Thus, getting 100 when one deserves 100 produces 200 units of desert-adjusted welfare. And getting 0 when one deserves 100, produces -50 units of desert-adjusted welfare.

This is not a full theory, since it only applies where the absolute desert level (D) is greater than the welfare level (w) and both are greater than zero. But it is enough to identify a major problem with Feldman's approach. Although Feldman is not committed to this particular formula, he is, I shall show, committed to the problematic features of this formula.

The problematic feature of this formula, and of Feldman's general view, is that it is committed to:

F1: For any given level of welfare, there is some absolute desert level such that for any greater absolute desert level the desert-adjusted welfare is negative.

The above formula is committed to the view that getting 10 units of welfare produces

7

negative desert-adjusted welfare when one deserves more than 50 (since 2x10 - [D-10]/2 < 0), for D>50). And it is not merely a quirk of this formula that produces this result. For it follows immediately from Feldman's principle iv. (stated just prior to Graph 1):

> iv. For present purposes one sort of case is especially important. This is the sort of case in which a person receives some good, but less than the amount he deserves. In this sort of case, it seems to me, the value for the world of his receipt depends entirely on how large a fraction of what he deserves he receives. If it is a very tiny fraction of what he deserves, then the case is[6] relevantly like the case (Case iii) in which a deserving person receives nothing. It is then bad for the world. If the fraction is large enough, it may be almost as good as the case (Case i) in which a person receives what he deserves.

In the context of Feldman's general moral theory (where the moral value of a world is the sum over people of the moral values of their lives), this is implausible. For it entails that a world of highly deserving people (e.g., deserving 300 units), leading flourishing lives (e.g., getting 50 units each), is worse than a world with no life at all (since 2x50 - [300-50]/2 = -25 < 0). This is taking justice too seriously! By assumption, every single person would be living a very good life, and yet this theory says that it's better that they not live (as compared with no life at all). And that's not taking welfare promotion seriously enough. This is like an approach to welfare promotion that tries to be sensitive to how equally welfare is distributed by being strictly egalitarian. Such an approach

says that everyone getting nothing is better than some getting 10 and some getting 20. And that's crazy. Just as there are more plausible ways of being sensitive to the equality of the distribution (e.g., maximinning), there are, I shall show below, more plausible ways of being sensitive to justice.

Feldman's theory takes justice too seriously, and welfare promotion not seriously enough. The objection just raised pointed out how Feldman's view implausibly implies that no life is sometimes better than flourishing (but inadequately rewarded) lives. I shall now identify a second manner in which Feldman's theory takes justice too seriously and welfare promotion not seriously enough.

The above formula and discussion are compatible with the view that increasing someone's welfare always increases the desert-adjusted welfare. For the implausible results discussed above concerned cases of non-existence versus existence. They did not concern cases of comparing two worlds in which a person exists in both and is better off in one of them. Feldman's approach, I shall show, also generates implausible results for this sort of case.

Any plausible theory of the moral good must, I claim, satisfy the following condition:

Pareto Superiority: If one world yields more welfare for at least one person, and at
least as much for everyone, as compared with a second world with exactly the same
people, then the first world is morally better.

This is a condition that requires that moral goodness be positively sensitive to individual welfare. Oversimplifying somewhat, it says that more welfare is morally better.

Nothing in Feldman's written discussion ensures that this condition is satisfied. The key issue is whether making a <u>negatively</u> deserving person better off increases his/her desert-adjusted welfare. In order to satisfy the Pareto Superiority condition, this must be so. One can agree with Feldman that increasing the welfare of a negatively-deserving person increases her desert-adjusted welfare <u>less</u> than it increases that of a similarly placed neutrally or positively deserving person, but in order to satisfy Pareto Superiority one must also hold that it <u>increases</u> her desert-adjusted welfare. All else being equal, making anyone–even a negatively deserving person–better off is better than benefiting no one.

In principle, Feldman could design his theory to ensure that Pareto Superiority is satisfied. For none of his written discussion commits him to the rejection of this condition. He could satisfy the condition by holding that, for people with negative desert D, w units of welfare produces w+D units of desert-adjusted welfare (where D is negative). Feldman has made it clear, however, that he rejects this condition. For he holds that giving a benefit to a negatively deserving person <u>decreases</u> her desert-adjusted welfare. And consequently, all else being equal, it makes the world morally worse. The idea, of course, is the strongly retributive view that the negatively deserving person deserves punishment, and so it is morally worse for the person to be rewarded rather than her being punished or getting nothing.

I find this view implausible. For it entails (in the context of Feldman's general theory) that a world of negatively deserving people with flourishing lives is morally worse than a world of the same negatively deserving people with horrible lives. That, I claim, is taking justice too seriously, and welfare promotion not seriously enough. We can grant that it's morally better to give a benefit

10

to a positively deserving person than to a negatively deserving person without holding that it's better to give no benefit at all than to give to a negatively deserving person.

In order to make more concrete where Feldman's theory goes wrong, I shall identify an alternative to Feldman's theory of moral value that is sensitive to welfare promotion and to justice, that satisfies D, DAW1-DAW3, and Pareto Superiority, and that avoids the problematic implications identified above. Like Feldman's theory, this theory holds that the moral value of a world is the sum over people of the moral value of their lives, and it holds that the moral value of a life is its desert-adjusted welfare. The theory differs from Feldman's theory in its account of desert-adjusted welfare.

The theory ascribes a <u>desert factor</u> to each person based on his/her marginal desert—with greater marginal desert factors for greater marginal desert. Neutral desert has a factor of 1, positive deserts have factors greater than 1, and negative deserts have factors less than 1 but greater than 0. For any given increment in welfare for a given person, the desert-adjusted welfare increment is the <u>product</u> of his/her marginal desert factor and the welfare increment. Assuming for illustration that marginal desert is constant and proportional to absolute desert, this can be represented as follows:

$daw(w,i,D) = w \times df(D)$, where $df(D)$ is the desert factor corresponding to desert level D.

This theory ensures that a person's desert-adjusted welfare increment is determined solely by her marginal desert and her welfare increment (DAW1). It ensures that the greater the marginal

desert, the greater the increment in desert-adjusted welfare (DAW2). And it ensures that the increment (decrement) in desert-adjusted welfare for a person with neutral/positive/negative desert is the same as/greater than/less than the welfare increment (decrement) (DAW3). It also satisfies Pareto Superiority, since all desert factors are positive (even for negative desert). And given the assumption that the moral value of a world is the sum over people of the moral values of their lives, it ensures that a world with people with lives worth living (above 0 in welfare) is morally better than a world with no people.

Such a theory, I claim, is more plausible than Feldman's. It takes desert into account without letting it overpower welfare promotion. My purpose here is not to defend this theory, but only to illustrate how one could accept much of the underlying motivation of Feldman's theory without accepting some of the implausible specifics.

4. The Repugnant Conclusion

I turn finally to how Feldman's theory blocks the Repugnant Conclusion. It is generally held that all total views of value (e.g., where the value of a world is the sum over people of the value of their lives) have the following implication:

Repugnant Conclusion: For any world with a finite number of flourishing lives, there is a world with many more people with lives barely worth living that is better.

This seems to be an implication of all total views, since enough lives with low value can add up to more than a smaller number of lives with high value. Feldman shows, however, that this is not so. For his theory is a total theory, and it does not have this implication. In Feldman's theory, the Repugnant Conclusion is blocked by the assumption that:

> F2: At the beginning of life each person deserves a <u>significantly positive</u> (as opposed to barely positive) life, say, 100 units of welfare.

In the context of Feldman's theory, this assumption blocks the Repugnant Conclusion. For, if each person starting life deserves 100 (say), then any person whose desert is still at 100 with welfare below a specified positive level (e.g., 20 in the Feldmanian formula discussed above) will have <u>negative</u> desert-adjusted welfare. And thus the effect of adding such people to the world will be to <u>decrease</u> the moral value of the world. More specifically, if, as above, $daw(w,i,D) = 2w - (D-w)/2$, for $D>w>0$, then, anyone deserving 100 with a life at less than 20 units of welfare will have negative desert-adjusted welfare.

Of course, there are cousins of the Repugnant Conclusion that are still applicable. For it is true that for any world with a finite number of people with flourishing lives (e.g., above 100) there is a world with sufficiently more people with welfare slightly above 20 that is judged morally better. Although this conclusion may still be distasteful, it is not repugnant. It is repugnant that enough lives <u>barely worth living</u> (e.g., at .0001 units of welfare) should be better than a large number of flourishing lives. It is at most merely distasteful that enough <u>reasonably good lives</u>

should be better than a large number of flourishing lives.

So, I grant that in an important sense Feldman has blocked the Repugnant Conclusion. He does so, however, at the cost of making his theory even more implausible. For in the context of his theory of the moral value (as the total over people of their desert-adjusted welfare), this assumption leads to the conclusion that (assuming for illustration the desert-adjusted welfare formula stated above) a world in which everyone's welfare is 19, and everyone's desert is unchanged at 100 since they came into being, is morally worse than a world with no life at all. This is so, since each person's desert-adjusted welfare will be negative (since $2 \times 19 - [100-19]/2 = -2.5$). But that's crazy. Each person has a life that is more than barely worth living, and so this world is morally better than a world with no life. The problem here is the same problem identified above in my discussion of F1. The problem is made worse, however, by the assumption of F2. For F2 increases (from 0 to 20, say) the minimum level of welfare necessary to have a positive impact on moral value. So although Feldman does block the Repugnant Conclusion, the cost is too great.

It will be instructive before closing to note how standard utilitarianism could adopt Feldman's idea and block the Repugnant Conclusion. Forget justice for the moment. Consider the standard total utilitarian view (not adjusted for desert) but with the zero point for moral value set at the threshold of what it takes to have a moderately good life (e.g., 20 on the absolute scale for welfare). Anyone with less than a moderately good life will be evaluated as having a life with negative moral value on this scale. Barely happy lives will have negative moral value. The standard Repugnant Conclusion is blocked.

For the standard total view, where the existing people are fixed, the zero points for welfare

and moral value are irrelevant. All that matters are the welfare gains and losses relative to alternative worlds. Where the existing people are not fixed, then the zero points become relevant. It matters because numbers must be assigned to non-existence. The usual assumption is that non-existence is worth zero in welfare and in moral value. The above approach separates the zero point for moral value from that of welfare. It holds that a person can have a life worth living (above 0 in terms of welfare), but still have a negative impact on the moral value of the world as a whole (e.g., if it is less than 20 units of welfare).

If this approach is taken, standard total utilitarianism can avoid the Repugnant Conclusion. For lives that are barely worth living have negative moral value, and thus adding such lives to the world makes things worse. No number of lives barely worth living can be better than a finite number of flourishing lives.

Unfortunately, this does not solve the problem of the Repugnant Conclusion for standard total utilitarianism. For, like Feldman's theory, this version of the theory has the problem that is says that many people with lives worth living but below the moderately good threshold (and thus below the moral zero point) is morally worse than no people at all. And that is not plausible. This is just another illustration of how the sort of assumption that Feldman makes in F2 is effective in blocking the Repugnant Conclusion, but implausible.

5. Conclusion

Feldman has insightfully developed some previously undeveloped possibilities. He has rightly emphasized the possibility and plausibility of developing a theory of moral goodness that is

sensitive to both justice and welfare promotion. And he has identified an interesting way of blocking the Repugnant Conclusion. I have argued, however, that his theory takes justice too seriously in that it does not ensure that if everyone is better off in one world than in a second, then the first world is morally better (Pareto Superiority). Violation of this condition shows that his theory is guilty of "justice-worship". A more plausible approach would recognize justice as relevant <u>but subservient to</u> welfare promotion.

Notes

1.Fred Feldman, <u>Confrontations with the Reaper</u> (Oxford: Oxford University Press, 1992), "Adjusting Utility for Justice", <u>Philosophy and Phenomenological Research</u>, forthcoming.

2.Fred Feldman, "Justice, Desert, and the Repugnant Conclusion", <u>Utilitas</u> ? (199?): ??-??.

3.Once the assumption of constant marginal desert proportional to absolute desert is dropped, some assumption connecting absolute and marginal desert is needed. A natural, and plausible, assumption is that increases in absolute desert increase marginal desert. More carefully: if one absolute desert level is greater than a second, then (1) for all levels of increment, the marginal desert associated with the first absolute desert level is at least as great as the marginal desert associated with the second, and (2) for at least one level, the marginal desert of the first is greater.

4.For an explanation and defense of this view, see Peter Vallentyne, "Rights-Based Paretianism," <u>Canadian Journal of Philosophy</u> 18 (1988): 89-101.

5.This summarizes P1-P6 of Feldman's "Adjusting Utility for Justice".

6.Roughly this point is also made in Ingmar Persson's "Feldman's Justicized Act Utilitarianism", <u>Ratio</u>, forthcoming.