Fred Feldman has proposed a desert-adjusted version of utilitarianism, 'justicism', as a plausible population axiology. Among other things, he claims that justicism avoids Derek Parfit's 'repugnant conclusion'. This paper explains the theory and tries to straighten out some of its ambiguities. Moreover, it is shown that it is not clear whether justicism avoids the repugnant conclusion and that it is has other counter-intuitive implications. It is concluded that justicism is not convincing as a population axiology.

I. INTRODUCTION

A common objection to total utilitarianism is that it is insensitive to matters of distributive justice. Fred Feldman has developed a desert-adjusted version of total utilitarianism, 'justicism', which he thinks fares better in this respect.¹ Moreover, Feldman claims that as a 'happy by-product, justicism also generates a plausible answer to Parfit's awesome question': How many people should there ever be?² As a theory of distributive justice, Feldman's theory has been discussed elsewhere.³ We shall focus on justicism's implications in population ethics. The next section explains Feldman's theory in detail and tries to straighten out some of its ambiguities. Section III discusses whether it gives a plausible answer to Parfit's question. The final section points out some further reasons why this is not the case.

II. FELDMAN'S DESERT-ADJUSTED UTILITARIANISM

In hedonism, the value of an episode of pleasure or pain is a function of its hedonic level. In justicism, the value of such an episode is determined not only by the hedonic level but also by the recipient's desert level: 'the intrinsic value of an episode of pleasure or pain is a function of two variables: (i) the amount of pleasure or pain the

---

² Utilitarianism, Hedonism, and Desert, p. 195.
recipient *receives* in that episode, and (ii) the amount of pleasure or pain the recipient *deserves* in that episode. A person’s desert level is determined by factors such as her excessive or deficient past receipt of pleasure or pain, her moral worthiness, her rights and legitimate claims, her past conscientious efforts, and so forth. A person is said to have ‘positive desert’ if she deserves some pleasure, ‘negative desert’ if she deserves some pain, and ‘neutral desert’ if she deserves neither pleasure nor pain. Feldman partly describes the relationship between pleasure, pain, desert, and intrinsic value with the following six principles:

M1. Positive desert enhances the intrinsic goodness of pleasure.
M2. Negative desert mitigates the intrinsic goodness of pleasure.
M3. Neutral desert neither enhances nor mitigates the intrinsic goodness of pleasure.
M4. Positive desert enhances the intrinsic badness of pain.
M5. Negative desert mitigates the intrinsic badness of pain.

Unfortunately, in his discussion Feldman does not consistently abide by his own principles. He claims that ‘receipt of much less [good] than you deserve is not good for the world’ and that the intrinsic value of a life led by person who deserves 100 units of pleasure but receives only one unit is -49. These claims are clearly inconsistent with M1. Moreover, these claims are crucial for Feldman’s results in population axiology. As Ingmar Persson has pointed out, Feldman oscillates between two ideas: the merit-idea and the fit-idea. According to the former, the higher the desert level, the higher the value of pleasure. The latter idea, on the other hand, focuses on the degree of fit between desert and receipt. The merit-idea corresponds pretty well with M1–6 above, whereas the fit-idea does the work in Feldman’s discussion of population axiology. We shall therefore replace M1–6 with some new principles that better accord with Feldman’s intuitions in this area. Furthermore, we shall incorporate his idea that mitigations might yield that pleasure is intrinsically bad — Feldman refers to this

*Utilitarianism, Hedonism, and Desert*, pp. 162 f., emphasis in original. Feldman formulates justicism as a version of classical hedonism mainly for pedagogical reasons. It could equally well have been stated in terms of Feldman’s propositional theory of pleasure or in terms of some other theory of welfare. See ibid., p. 152.

5 Ibid., pp. 161 f., 202 f.
6 Ibid., pp. 163–9.
7 Ibid., pp. 206, 163.
8 Carlson, 315, makes the same point.
9 Persson.
phenomenon as 'transvaluation' — in the principles. Call a person's pleasure 'deserved' if it roughly corresponds to her desert level, that is, if she receives exactly what she deserves or fairly close to what she deserves. If a person's pleasure does not roughly correspond with her desert level and thus is much more (less) than she deserves, then this pleasure is 'under-deserved' ('over-deserved'). The following principles probably capture Feldman's intuitions about desert and pleasure better than M1–6:

F1. Positive desert enhances the intrinsic goodness of deserved pleasure.
F2. Positive desert mitigates the intrinsic goodness of under-deserved pleasure.
F3. Positive desert mitigates and might transvaluate the intrinsic goodness of over-deserved pleasure.
F4. Negative and neutral desert mitigate the intrinsic goodness of pleasure.
F5. Positive and neutral desert enhance the intrinsic badness of pain.
F6. Negative desert mitigate the intrinsic badness of pain.

As we noticed above, if pleasure is over-deserved, then we might get transvaluation of the intrinsic goodness of pleasure — hence the formulation of F3. What about under-deserved pleasure? Feldman is not very clear on this point, but we shall interpret his talk about 'decreasing marginal intrinsic value' such that positive desert can mitigate but not transvaluate the intrinsic goodness of under-deserved pleasure. Consequently, F2 does not say anything about transvaluations. We have also reformulated M3 and M6 since from the perspective of the fit-idea, it is reasonable to claim that neutral desert mitigates the goodness of pleasure and enhances the badness of pain (see F4 and F5).¹¹

Finally, according to justicism, the intrinsic value of a person's life is the sum of the desert-adjusted intrinsic value of the episodes of

¹⁰ *Utilitarianism, Hedonism, and Desert*, p. 165. Feldman also discusses the possibility that mitigations yield that pain is intrinsically good (ibid., p. 167). I shall return to this idea in section IV.

¹¹ In personal communication, Feldman has confirmed that the original formulation (M3 and M6) was confused and that his view on neutral desert is as in F4 and F5. Feldman also claims (ibid., p. 168) 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'. This idea is compatible with M5 (F6) but I would suggest reformulating this principle in terms of deserved pain since it seems odd, from the perspective of proportional justice, that negative desert mitigates the intrinsic badness of very under-deserved pain, that is, pain that goes far beyond the deserved pain. I shall not pursue this matter further here, however.
pleasure and pain that occur in her life. The value of a population is the sum of the values of all the lives in the population.¹²

III. JUSTICISM AND THE REPUGNANT CONCLUSION

Feldman claims that his theory does not imply Derek Parfit's repugnant conclusion:

The Repugnant Conclusion: For any perfectly equal population with very high positive welfare, there is a population with very low positive welfare which is better, other things being equal.¹³

Diagram 1

<table>
<thead>
<tr>
<th>Very high positive welfare</th>
<th>Very low positive welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population B is much larger than A</td>
<td></td>
</tr>
</tbody>
</table>

In diagram 1, the width of each block represents the number of people, the height represents their lifetime welfare. All the lives in the above diagram have positive welfare, or, as we also could put it, have lives worth living. People's welfare is much lower in B than in A, since the A-people have very high welfare whereas the B-people have very low positive welfare. The reason for the very low positive welfare in the B-lives could be, to paraphrase Parfit, that there are only enough ecstasies to just outweigh the agonies or that the good things in life are of uniformly poor quality, e.g. working on an assembly line, eating potatoes, and listening to Muzak.¹⁴ However, since there are many more people in B, the total sum of welfare in B is greater than in A. Hence, total utilitarianism ranks B as better than A – an instance of the repugnant conclusion.

How is justicism supposed to avoid the repugnant conclusion? Feldman's clever and original proposal is that there is 'some modest

¹² Feldman writes (ibid., p. 169): 'The intrinsic value of a whole consequence is the sum of the justice-adjusted intrinsic value of the episodes of pleasure and pain that occur in that consequence.' On p. 208 he says that 'the relevant ... value of a world ... is the sum of the values of the lives lived there, adjusted for desert.'

¹³ See Derek Parfit, Reasons and Persons, Oxford, 1984, p. 388. My formulation is more general than Parfit's and he does not demand that the people with very high welfare are equally well off.

level of happiness that people deserve merely in virtue of being people.\textsuperscript{15} He assumes that this modest level corresponds to 100 units of pleasure and that people with very low welfare enjoy only one unit of pleasure. Feldman has not given us any exact formula for calculating the desert-adjusted value of a life, but in his discussion of the repugnant conclusion he says that the life of a person who deserves 100 units of pleasure and receives exactly that amount of pleasure has an intrinsic value of 200. As we noticed above, if a person deserving 100 units only receives one unit of pleasure, then the intrinsic value of her life is -49.\textsuperscript{16} It follows that any population consisting of people with very low welfare and desert level 100 has negative value, whereas any population with very high welfare has positive value.\textsuperscript{17} Consequently, it seems that justicism avoids the repugnant conclusion.

The intuition behind Feldman's explanation of the unacceptability of the repugnant conclusion – that there is some level of welfare that people deserve merely in virtue of being people – is compelling and probably shared by many people. Moreover, Feldman's theory can explain ideas such as Blackorby et al.'s critical level according to which there is a positive welfare level that a life has to attain to have positive contributive value.\textsuperscript{18} But I'm not sure Feldman's theory really delivers what it promises. Feldman's reasoning involves a questionable interpretation of the \textit{ceteris paribus} clause in the repugnant conclusion. He implicitly assumes that the \textit{ceteris paribus} clause is satisfied whenever the people in the compared populations have the same desert level. This interpretation – let us call it the 'same merit interpretation' – is questionable for two reasons. First, given the fit-idea, which is crucial for Feldman's solution to the repugnant conclusion, the same merit interpretation seems out of place. Rather, closer at hand is the view that the \textit{ceteris paribus} clause is satisfied if there is the same fit between what people deserve and what they receive in compared populations. Again, Feldman oscillates between the merit- and the fit-idea. More importantly, it is not at all clear why we should focus on each individual's desert level. A more plausible reading of the \textit{ceteris paribus} clause, and one which I think Parfit had in mind when he formulated the repugnant conclusion, is that it is satisfied if and only if the compared populations are (roughly) equally good in regard to other axiologically relevant aspects apart from welfare. Consequently, what we are looking for are cases where the

\textsuperscript{15} Utilitarianism, Hedonism, and Desert, p. 194.
\textsuperscript{16} Ibid., pp. 206, 209.
\textsuperscript{17} Given the assumption that positive desert cannot yield transvaluation of the intrinsic goodness of under-deserved pleasure.
compared populations are, in some sense, equally good in regard to desert.

Admittedly, it is not completely clear how we should understand the ceteris paribus clause in relation to justicism since the two axiologically relevant aspects are entangled in a complex manner. As we saw above, the intrinsic value of an episode of pleasure or pain depends on the amount of pleasure or pain the recipient deserves in that episode. Strictly speaking, this is not compatible with Feldman’s own idea of intrinsic value. As he writes in another context, ‘[s]urely, if something is intrinsically good, it must be good in virtue of the way it is in itself, not merely because of some extrinsic relation it happens to bear to some other thing’. But in justicism, as stated by Feldman, it seems that the intrinsic value of an episode of pleasure or pain depends on contingent facts regarding the desert level of the recipient. However, this should only be taken metaphorically. What Feldman actually means is that the only carriers of intrinsic value are compound states of affairs consisting of a person’s experience of pleasure or pain and their desert level.

Now, we want to compare populations which are (roughly) equally good in regard to other axiologically relevant aspects apart from welfare. We thus need to separate the contributive value of welfare and the contributive value of the fit between desert and receipt to the intrinsic value of a life. To that end, we shall define a life’s desert value in terms of the difference between the value of that life and its welfare value, that is, in terms of how much the fit between desert and receipt contributes or detracts from the intrinsic value of a life. We are, so to say, factoring out the desert component of the intrinsic value of a life. For example, Feldman says that the intrinsic value of a life enjoying a deserved one unit of pleasure is two. The contributive value of the pleasure in this life is thus one unit, and the contributive value of the fit between desert and receipt, the desert value, is also one unit. These two values taken together yield that the intrinsic value of this life is two units, in accordance with Feldman’s view.

Given this definition of a life’s desert value, there is a straightforward interpretation of the ceteris paribus clause: The compared populations should be equally good in regard to desert value; we should not have a reason to choose one or the other of the compared populations because of the fit between desert and receipt. The only reason for choosing one or the other population should be the well-

---

19 Utilitarianism, Hedonism, and Desert, p. 138.
21 Utilitarianism, Hedonism, and Desert, p. 212.
being of their members. Now, to be able to determine how good a population is in regard to desert value, we need a method of aggregating this value. Given Feldman’s framework, the desert value of a population is determined by adding up the individual desert values of its members (recall that according to justicism, the value of a population is the sum of the desert-adjusted intrinsic value of the episodes of pleasure and pain that occur in the population). Consequently, two populations are equally good in regard to desert if and only if the total sum of desert value is the same in the compared populations. Let us call this the ‘same desert value interpretation’ of the *ceteris paribus* clause. In the light of Feldman’s ideas about desert, I find this a much more plausible reading of the *ceteris paribus* clause than the same merit interpretation. If the only relevant axiological aspects of a population are people’s welfare and their desert value, and two populations differ in regard to people’s welfare but not in regard to their desert value, then it is plausible to say that compared populations are equally good in regard to all other axiologically relevant aspects apart from welfare.

Does justicism avoid the repugnant conclusion if we read the *ceteris paribus* clause according to the same desert value interpretation? I do not think so, since if the desert value is the same in the compared populations, then the ranking is determined solely by the total sum of people’s welfare. For example, consider a population A with very high deserved welfare. Assume that the total desert value of this population is $x$ units. Again, the desert value of a life enjoying a deserved one unit of pleasure is one. Consequently, a population B consisting of $x$ lives enjoying a deserved one unit of pleasure will have the same total desert value as population A (we are simplifying our reasoning here by assuming that the desert value of a population can always be represented by an integer, but it should be clear how the argument could proceed without this assumption). Since lives with one unit of pleasure and neutral desert have negative desert value (from F4) and lives with a deserved one unit of pleasure have positive desert value, there is some desert level between 0 and 1 such that a life with this desert level and one unit of pleasure has neutral desert value. Now, if the total welfare of population B is less than the total welfare of population A, then just add a sufficient number of lives with neutral desert value and one unit of pleasure. The resulting population consists only of people with very low positive welfare, and is equally as

---

22 We are assuming here that the functions involved are continuous which Feldman also seems to assume (ibid., pp. 205 f.). Again, it should be clear how the argument could proceed without this assumption. Even if no individual could have neutral desert value, we could put together groups of lives with aggregate neutral desert value since their negative and positive desert value would cancel each other out.
good as A in regard to desert, but better than A according to justicism since the total welfare is greater.

We can proceed similarly with populations of people with under-deserved very high welfare, populations of people with over-deserved very high welfare, and populations of people with very high welfare but with varying desert. For any of these populations, we can take a number of lives with one unit of pleasure and a desert level such that the aggregate desert value corresponds to the aggregate desert value in the high welfare population. By adding a sufficient number of lives with neutral desert value and one unit of pleasure, we will get populations that have the same desert value as the populations with very high welfare, but which are better according to justicism since the total welfare is greater. Consequently, contrary to Feldman, justicism implies the repugnant conclusion.

In cases involving deserved welfare, as we showed above, justicism implies repugnant conclusions. Feldman discusses a case like this, although, since he reads the *ceteris paribus* clause according to the same merit interpretation, he does not think it exemplifies the repugnant conclusion. Nevertheless, he considers whether this ‘variant of the original example will prove just as repugnant’.\(^{23}\) He writes:

\[\text{[I]t is not entirely clear that } Z' \text{ [a population of two billion billion people with one unit of deserved pleasure] ought to be considered horrible. Note that the residents of } Z' \text{ are not like us. They deserve far less than we deserve. Each of them deserves just } +1 \text{ and each of them gets exactly what he or she deserves. Since } Z' \text{ is so incredibly populous, and since the total amount of good enjoyed by the residents is so huge, and since everything in } Z' \text{ is said to be just as it ought to be, it is not clear that we should find } Z' \text{ repugnant.}\]

\(^{24}\)

I do not find this answer convincing. It rests on a kind of misunderstanding of the repugnant conclusion. The counter-intuitiveness of this conclusion does not essentially rest on categorical properties of populations with very low positive welfare, for example, that such populations are repugnant or very bad in themselves. The unacceptability of the repugnant conclusion arises from the fact that any population with very high welfare is *worse* than some population with very low welfare. It is this *comparative aspect* of the repugnant conclusion that we find hard to accept. And this counter-intuitiveness is not ameliorated by Feldman’s appeal to desert. On the contrary, I suggest that those who embrace Feldman’s explication of desert should have an even more firm belief about the unacceptability of the repugnant conclusion than those who do not: How could the fact that the people with very high welfare also have a *very high desert level*, and

\(^{23}\) Ibid., p. 212.

\(^{24}\) Ibid., p. 212.
the fact that the people with very low welfare also have a very low desert level, reverse our intuitive judgement about the repugnant conclusion? Since the people with very low desert level must be less morally worthy, made less conscientious efforts, and so forth, than the people with very high welfare, these differences in desert only serve to strengthen the dreadful character of the repugnant conclusion.

Feldman also claims that 'it is not clear that the description of Z' [the population of two billion billion people with deserved pleasure one] is coherent':

I stipulated that merely in virtue of being a person, each of us deserves +100. - - The people in Z' allegedly deserve much less. But why do they deserve much less? It must be because they did something wrong. If they did something wrong, that would make Z' worse. The description of Z' is therefore incomplete. These people must have done some evil deeds. Yet the evil of those deeds is neither described nor included in the calculations.25

I find this claim surprising. It seems clear to me that the 'evil of those deeds' is included in the calculations made when we evaluate worlds according to justicism. These deeds are reflected in the low desert-level of the Z'-people and in their low welfare. For example, the population with very high welfare could be inhabitants of a world where people cooperate and help each other. Their high desert level is a reflection of their cooperative and helpful characters whereas their high welfare is due to the fruits of cooperation. In Z', on the other hand, people do not cooperate but only look out for themselves. They live in a Hobbesian state of nature: 'war of every one against every one' which makes life 'nasty, brutish and short'.26 The low desert level in this world reflects the selfish character of its inhabitants whereas their low welfare is caused by lack of co-operation.

Given the same desert value interpretation of the ceteris paribus clause, justicism implies the repugnant conclusion. Assume, implausibly, that Feldman could muster some decisive argument to the effect that although justicism is based on the fit-idea, the same merit interpretation is the most plausible way to understand the ceteris paribus clause. Although justicism combined with the same merit interpretation avoids the repugnant conclusion, it would imply analogous conclusions. For example, since positive desert mitigates the value of under-deserved pleasure (F2), justicism implies that for any population consisting of people who deserve one unit of pleasure but who enjoy very pleasurable lives, there is a population consisting of people with a deserved one unit of pleasure and lower total welfare which is better. Moreover, with the same merit interpretation,

25 Ibid., p. 212.
justicism has some other very counter-intuitive implications to which we shall now turn.

IV. JUSTICISM AND THE WEAK NON-SADISM CONDITION

At one point, Feldman consider the possibility that 'it is slightly good (+ 2.5) for a person to receive 10 units of pain when this is precisely what he deserves' as a way of capturing retributivist intuitions.\textsuperscript{27} It follows from this view that a life involving only undeserved pain, or deserved pain and other pains and pleasures that cancel each other out, has positive intrinsic value and that we can make the world much better by adding a large number of lives with overall deserved negative welfare. Since we can assume that the people in the compared populations have the same desert level, justicism violates the following compelling principle given the same merit interpretation:

\textit{The Negative Mere Addition Principle:} An addition of people with negative welfare makes a population worse, other things being equal.

One might object here that other things cannot be equal since a person who deserves pain must have done something wrong, and this wrong must have involved bringing undeserved pain to others. Consequently, the existence of such a person would make a population worse.\textsuperscript{28} There is, however, nothing in Feldman's theory that implies such a necessary connection between negative desert and bringing undeserved pain to others. Moreover, even if there were such a connection, we could assume that the undeserved pain is outweighed by pleasure brought about by, for example, unintended and fortuitous circumstances.

According to Feldman, the positive value of deserved pain 'expresses the retributivist axiological intuition that sometimes it is good for bad people to be punished'.\textsuperscript{29} I doubt that many retributivists would agree. An axiological retributivist thinks that if people commit crimes, then it is good that they are punished. In other words, it is worse if crimes are committed with impunity than if they are committed and punished, but neither of these states of affairs is good. Retributivists do not think that it is good that people commit crimes and are punished.

Since lives with positive welfare might have negative intrinsic value

\textsuperscript{27} Utilitarianism, Hedonism, and Desert, pp. 167 f.
\textsuperscript{28} Feldman made a similar suggestion in personal communication.
\textsuperscript{29} Ibid., p. 167.
and lives with negative welfare might have positive intrinsic value, it should come as no surprise that justicism violates the following condition:

The Weak Non-Sadism Condition: There is a negative welfare level and a number of lives at this level such that an addition of any number of people with positive welfare is at least as good as an addition of the lives with negative welfare, other things being equal.

As we noticed above, Feldman’s theory yields that the intrinsic value of a person who deserves 100 units of pleasure but receives only one unit is -49. What is the intrinsic value of a person who deserves 100 units of pleasure but receives terrible pain? As it is presented, Feldman’s theory does not give an exact answer to this question. At any rate, it follows from F5 that such a life is also going to have negative intrinsic value, presumably much below -49. Assume that the value of such a life is -k. Now, for any number n of people who suffer terrible pain but who deserve 100 units of pleasure, there is a number m of people with over-deserved positive welfare but with lower total intrinsic value, namely any number m > kn/49 of such lives. From the perspective of desert, I find this violation of the weak non-sadism condition perplexing. Surely, if everybody deserves positive welfare and one has a choice of adding people suffering terrible pain or people with positive (albeit over-deserved) welfare, then the latter addition must be the better one.

To be fair to Feldman, he does not commit himself to the transvaluation of the evil of pain by negative desert. His view seems rather to be that the intrinsic value of a life with deserved negative welfare is zero. Nonetheless, this version of justicism still violates the negative mere addition principle and it does not capture the view of the axiological retributivist since lives with deserved negative welfare have neutral rather than negative intrinsic value. Likewise, since the intrinsic value of people with over-deserved positive welfare can be negative, it still violates the weak non-sadism condition.

Given the same desert value interpretation, justicism does not violate the weak non-sadism condition and the negative mere addition principle since if the desert value is the same in the compared populations the ranking is going to be determined solely by the total

---

30 Feldman’s choice of numerical representation is, according to himself, ‘somewhat arbitrary’ but since the value of a person who deserves 100 units of pleasure and receives nothing is -50, and positive desert enhances the intrinsic badness of pain (see M4 and F5 above), a person who deserves 100 units of pleasure and receives pain must have negative intrinsic value below -50. See ibid., p. 206.

31 Ibid., p. 167. At p. 165, Feldman also suggests that the value of a person with negative desert but positive welfare is zero rather than negative.
sum of people's welfare. This is not much of a comfort, however, since justicism will still have the implications pointed out above, even though these implications will not formally count as violations of the weak non-sadism condition and the negative mere addition principle since other things are not equal. Moreover, with this interpretation of the ceteris paribus clause, justicism implies the repugnant conclusion. Consequently, Feldman's desert-adjusted utilitarianism does not constitute a step forward in our search for a reasonable population axiology.\textsuperscript{32}

Gustaf.Arrhenius@philosophy.oxford.ac.uk

\textsuperscript{32} I would like to thank John Broome, Krister Bykvist, Erik Carlson, Sven Danielsson, David Donaldson, Fred Feldman, Tom Hurka, Adeze Igboemeka, Jan Österberg, Wlodzimierz Rabinowicz, Howard Sobel, and Wayne Sumner for their helpful comments on earlier drafts of this paper, and Erik Arrhenius for his help in practical matters. Financial support from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) is gratefully acknowledged.