MERITARIAN AXIOLOGIES AND DISTRIBUTIVE JUSTICE

Gustaf Arrhenius

Department of Philosophy, Stockholm University
The Swedish Collegium for Advanced Study
Gustaf.Arrhenius@philosophy.su.se

ABSTRACT: Standard welfarist axiologies do not care who is given what share of the good. For example, giving Wlodek two apples and Ewa three is just as good as giving Wlodek three and Ewa two, or giving Wlodek five and Ewa zero. A common objection to such theories is that they are insensitive to matters of distributive justice. To meet this objection, one can adjust the axiology to take distributive concerns into account. One possibility is to turn to what I will call Meritarian axiologies. According to such theories, individuals can have a claim to, deserve, or merit, a certain level of wellbeing depending on their merit level, and the value of an outcome is determined not only by people's wellbeing but also by their merit level.

1. Introduction

Standard welfarist axiologies do not care who is given what share of the good. For example, giving Wlodek two apples and Ewa three is just as good as giving Wlodek three and Ewa two, or giving Wlodek five and Ewa zero (assuming an "apple-measure" of welfare). A common objection to

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

such theories is that they are insensitive to matters of distributive justice. To meet this objection, one can adjust the axiology to take distributive concerns into account. One possibility, which I shall investigate in the present paper, is to turn to what I will call Meritarian axiologies. According to such theories, individuals can have a claim to, deserve, or merit, a certain level of well-being depending on their merit level, and the value of an outcome is determined not only by people's well-being but also by their merit level. A recent example of such a theory is Fred Feldman's desert adjusted version of Total Utilitarianism, "Justicism".[1] Here, a person's desert (i.e., merit) 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. Shelly Kagan and Tom Hurka have developed similar theories.[2] Need-based theories might be taken as another example of a Meritarian axiology, since need can be taken as a kind of merit.[3]

We shall take a liberal view on what can count as a merit or desert base in the present paper. There has been a long-standing discussion on whether a necessary condition for something being a desert base is that that the deserving person is responsible for it.[4] We are here assuming that this view is false. We assume, for example, that it makes sense to say that somebody deserves compensation because she is, through no fault of herself, badly off, or more concretely, that she deserves an expensive treatment because

---

[1] See Feldman (1997), Arrhenius (2003, 2006).

[2] Hurka (2001, 2003), Kagan (2003).

[3] Cf. Danielsson (1974).

[4] See Feldman (1997), ch. 9; Sadurski (1985), p. 117; Rachels (1978), p. 157; Cupit (1996); Mcleod (2003), pp. 3-5; Miller (2003). Cf. Rawls (1971), p. 104.

she has, through no fault of herself, contracted a painful disease, and the like.[5] At any rate, since some people seem to have strong views on the connection between responsibility and desert, we are in the present paper mainly going to use the term “merit” instead of “desert” (although we are not sure that this is a useful change of terminology).

Let us also stress that we are neither going to defend any specific conception of merit or desert base in the present paper, nor that desert or merit has any useful role to play in moral theory. We are interested in the structural features of the desert claims that people seem to make.

Meritarian theories can differ depending on what is considered a merit and how merit is taken into account in the value function. Regarding the latter issue, we shall consider an idea --- the fit-idea --- according to which the degree of fit between merit and receipt is what matters.[6] We shall investigate a family of value function according to which the intrinsic value of a life is determined by the sum of the value of wellbeing and the value of the fit between wellbeing and the recipient’s merit level, that is, an additively separable value function. We shall introduce a formalism in which we can state possible constraints on the value function in a precise manner. This will make the theory and its implications take a clearer form as compared to earlier contributions, or so we hope. We shall then discuss different interpretations of what we take as two central principles underpinning such Meritarian axiologies. We shall prove that on some interpretations, these principles imply a number of interesting auxiliary principles that seem to capture some important intuitions about merit and distributive justice quite well. However, we shall also show that they

[5] For a convincing defence of this view, see Feldman (1997), ch. 9.

[6] See Arrhenius (2006).

conflict with some other intuitions about desert and merit. Lastly, we shall discuss the implications of this theory for the relation between so-called comparative and non-comparative desert claims.

2. Meritarianism

For the purpose of stating Meritarianism and the fit-idea more precisely, it will be useful to state some definitions and assumptions, and introduce some notational conventions.

We shall assume that we have a ranking of lives in terms of the relation "has at least as high welfare as". For the present discussion, it does not matter whether welfare is understood along the lines of experientialist, desire or objective list theories.[7]

We also need a ranking of lives in terms of how deserving they are. One might get the impression from the literature that what people have in mind is the relation "has at least as high merit as" and it is surely a natural starting point. Such a relation would, however, not give us any information about how much welfare people deserve. We would need further information on how to correlate a certain merit level with a certain welfare level. None of the protagonists of merit based theories has, to the best of my knowledge, showed how to connect such a merit ordering with a

---

[7] Since the concept of welfare used here is a broad one, many of the views presented in the debate on the currency of egalitarian justice as alternatives to welfare, for example Rawls' influential list of primary goods, will fall under the heading of welfare as the term is used in this paper. For this debate, see Rawls (1971), Sen (1980), Dworkin (1981a, 1981b, 2000), Cohen (1989, 1993), and Arneson (1989).

welfare ordering, or even discussed the matter. Clearly, this is a weakness in all the merit based theories that have been presented so far, and it deserves further investigation. In the present paper, however, we shall sidestep this issue by assuming that we have a ranking of lives in terms of the relation "merits at least as much welfare as".[8] As we mentioned above, we shall take a liberal view on the factors that are possible determinants of this ranking.

Let w_1, w_2,… and so on be the numerical representation of a certain welfare level of a life, that is, the welfare in that life taken as a whole, and let m_1, m_2,… and so on, be the numerical representation of a certain merit level of a life, that is, the merit of that life taken as a whole.[9] We shall assume that the welfare and merit level can be measured on a ratio scale.[10] These are quite strong assumptions but are necessary for some intuitions regarding proportionality and merit to make sense (more on this later).

Let (w, m) be a life with welfare level w and merit level m. Thus, (1, 2) represents a life in which a person receives one unit of welfare but merits two units of welfare. Let A, B, C, and so forth be populations of lives,

---

[8] I am indebted to Wlodek Rabinowicz for pressing this point.

[9] Notice that we are not taking a stand on how to aggregate episodes of welfare into a measure of the welfare in a life as a whole. It could be done by just summing the episodes, but there are other approaches that we might find more in line with our intuitions. Likewise for the merit of a life. Cf. Arrhenius (2005).

[10] A *ratio* scale is unique up to a similarity transformation, which means that the ratios of scale values are preserved. The admissible transformations are all functions of the form $f(x) = \alpha x$, $\alpha > 0$. Expressions such as "Wlodek has many times higher (lower) welfare (merit) than Toni" are meaningful. One can meaningfully compare ratios of welfare and merit.

represented by vectors $\langle(w_1, m_1), (w_2, m_2),\ldots, (w_n, m_n)\rangle$. For example, if A = $\langle(1, 5), (3, 4)\rangle$, then 3 is the numerical representation of the welfare level of the second person in population A.

Let us call the value of the fit between merit and receipt of welfare in a life its *fit value* and the value of the welfare in a life its *welfare value*. Let *WV* be a function that returns the numerical representation of the welfare value of a life or a population, let *FV* be a function that returns the numerical representation of the fit value of a life or a population, and let *IV* be a function that returns the numerical representation of the intrinsic value of a life or a population,

Any version of Meritarianism can be divided into three parts. One part is the value function that tells us how to aggregate welfare and fit value into a measure of the intrinsic value of lives and populations. The other two parts tell us about how the welfare value of a life depends on the welfare of the life, and how the fit value of a life depends on the fit between the welfare and the merit of a life. As we said above, we are going to investigate a version of Meritarianism according to which the intrinsic value of a life equals the sum of its welfare and fit values, and where the fit value is understood along the lines of the fit-idea. Let us call this theory Additively Separable Fit Meritarianism or ASFM for short. The value function of ASFM is defined by the following principles:

V1: $IV(w, m) = WV(w, m) + FV(w, m)$

It should be noticed that it is not self-evident that we should formulate Meritarianism as an additively separable value function. It seems a quite natural move if we consider the fit between merit and receipt another intrinsic value in addition to welfare. If one holds the view that there is only one type of basic carrier of intrinsic value, that is, compound states of affairs consisting of a person's welfare and her merit level, then one might think that it does not follow in a natural way that the intrinsic value of a life is a sum of its welfare and fit values. It could instead be, say, the product of its welfare and fit values, or some more complicated function.[11] Making it an additively separable function has the advantage of simplicity, however, and the purpose of the present paper is to investigate how far we can get with such a value function in combination with the fit-idea.

We are going to make the intrinsic value of a population A equal to the sum of the intrinsic value of the lives in A:[12]

V2: $IV(\text{A}) = IV\langle(w_1, m_1), (w_2, m_2), \ldots, (w_\text{n}, m_\text{n})\rangle =$
$IV(w_1, m_1) + IV(w_2, m_2) + \ldots + IV(w_\text{n}, m_\text{n})$

It follows from the two definitions above that the welfare value of a populations equals the sum of the welfare values of the lives in the population, and that the fit value of a population equals the sum of the fit values of the lives in the population.

We are also going to assume that the welfare value of a life equals its welfare level:

[11] For a discussion, see Arrhenius (2006).

[12] Cf. Feldman (1997), p. 169, who writes: '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'. 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 …'.

WV1: $WV(w, m) = w$

It follows from the above definitions that the welfare value of a population equals the sum of the welfare levels of the lives in the population. There are, of course, alternatives to the above formulations that we could have considered. For example, we could have made the welfare value of a life a marginally decreasing function of the welfare of the life and in such a way captured the intuition that we should give priority to the worst off. Likewise, instead of summing each life's welfare value to get a measure of a population's welfare value, we could instead have used, say, averaging to avoid Derek Parfit's repugnant conclusion.[13] However, the attractive feature of the Meritarian approach is that the merit part of the equation might be able to handle all of these intuitions, that is, introducing a merit component in the axiology might make tinkering with the aggregation of welfare superfluous. That is clearly worthwhile investigating, hence the definitions above.

In the following, I shall take the above definitions for granted and when I sometimes claim, without any reference to the above definitions, that a certain version of the fit-idea has a certain implication in respect to the intrinsic value of a life or a population, then that is just short for saying that it has this implication in conjunction with the above definitions.

---

[13] See Parfit (1984) and Arrhenius (2000).

3. The First Central Fit-Idea

As I said above, I think there are two central parts to the fit-idea. Here is the first one:

> *The First Central Fit-Idea*: The better the fit between receipt and merit, the higher the fit value.

How should we formulate this idea more exactly? Actually, there are several options available here, and it will be worthwhile to consider some different alternatives. Here is a first try:

F1-1: If $m \geq w_1 > w_2$, then $FV(w_1, m) > FV(w_2, m)$; and
if $w_1 > w_2 \geq m$, then $FV(w_1, m) < FV(w_2, m)$.

In words: If two lives with the same merit both have less welfare than they merit, or both have more welfare than they merit, then the life with the least difference between merit and receipt of welfare has the highest fit value. Moreover, if one life receives exactly what she merits, and another life receives more or less than she merits, then the former life has the highest fit value.

This principle partly captures the first central idea and in a rather weak way. Let us call a person's welfare "under-merited" ("over-merited") if her receipt of welfare is clearly more (less) than she merits. F1-1 is compatible with there being an asymmetry between the value of over- and under-merited welfare such that, for example, $FV(4, 3)$ is greater than $FV(2, 3)$. One might find this an attractive feature since one might reasonably think that it is better that a person gets one unit of welfare more than she merits rather than one unit less. F1-1 is, on the other hand, also compatible with $FV(4, 3)$ being smaller than $FV(2, 3)$ which seems hard to defend. We can avoid this latter possibility by adding a further requirement to F1-1, for example:

A1: If $|w_1 - m| = |w_2 - m|$ and $w_1 > w_2$, then $FV(w_1, m) > FV(w_2, m)$.[14]

In words: If the difference between merit and receipt is the same for two lives with the same merit, and one has more and one has less welfare than she merits, then the former life has the highest fit value.

Those who believe that from a *pure* merit perspective, it is better that one get one unit more rather than one unit less than one deserves should probably stick with F1-1 and A1 as the best explication of the first fit-idea. The question is, however, whether we should capture the intuition that it is better that one get one unit more rather than one unit less than one merits by introducing an asymmetry between the value of over- and under-merited welfare, or by assigning increasing value to increases in the welfare level, other things being equal. Given our definitions of the intrinsic value of a life and the value of welfare (V1 and WV1), it follows that if $FV(4, 3) = FV(2, 3)$, then $IV(4, 3)$ is greater than $IV(2, 3)$ since $WV(4, 3) = 4$ is greater than $WV(2, 3) = 2$. Consequently, introducing an asymmetry in the value of over- and under-merited welfare along the lines of A1 is not necessary to achieve the desired ranking, given V1 and WV1, and is in that sense superfluous. Thus I think that assigning increasing value to increases in the

[14] '$|a\text{-}b|$' represents the absolute difference between the numerical values a and b. If $a \geq 0$, then $|a| = a$; if $a < 0$, then $|a| = -a$. For example, $|5| = 5$ and $|-5| = 5$.

welfare level sufficiently accounts for the intuition that it is better that a person gets more than she deserves rather than less than she deserves.

One might, however, think that F1-1 is too strong in another respect. One might hold the view that what matters from the perspective of justice is that people get at least as much welfare as they have claim to, and when people have received that amount, all the demands of justice are met. For example, assume that the merit base in question is need. One might think that when all needs are satisfied, all is fine from the perspective of justice and it does not matter how welfare is distributed above the level where all needs are satisfied. F1-1, on the other hand, implies that for any merit and welfare level, the better the match between merit and receipt, the higher the fit value.

This property of F1-1 will be shared by the other interpretations of the first central fit-idea that we shall discuss below. It is easily remedied, however. If we think that there are no distributional concerns left when everybody has received at least as much as they merit, then we can just adjust the part of F1-1 that concerns under-merited welfare so that it implies that the fit value of under-merited welfare is always the same:

F1-1′: If $m \geq w_1 > w_2$, then $FV(w_1, m) > FV(w_2, m)$; and
if $w_1 \geq w_2 \geq m$, then $FV(w_1, m) = FV(w_2, m)$.

F1-1′ claims that if two lives with the same merit both have less welfare than they merit, or one has less and one has a perfect match between merit and receipt, then the life with the least difference between merit and receipt of welfare has the higher fit value. If two lives with the same merit both have at least the welfare that they merit, then they have the same fit value.

It follows from F1-1′ that if one person has at least the welfare that she merits, and another person has less than she merits, then the former has the higher fit value. For example, the first clause of F1-1′ implies that $FV(10, 10) > FV(9, 10)$, and the second clause implies that $FV(11, 10) = FV(10, 10)$. By transitivity, $FV(11, 10) > FV(9, 10)$.

The same adjustment can be made to the other principles below. Moreover, the implications of these principles that we shall show below will still follow albeit only for cases involving over-merited welfare.

With this qualification in mind, let us consider an explication of the fit-idea that implies symmetry between the values of under- and over-merited welfare:

F1-2: If $|w_1-m| < |w_2-m|$, then $FV(w_1, m) > FV(w_2, m)$; and
if $|w_1-m| = |w_2-m|$, then $FV(w_1, m) = FV(w_2, m)$.

In words: If two persons merit the same amount of welfare, and the difference between the first person's receipt and merit is less than (equal to) the difference between the second person's receipt and merit, then the fit value of the first person's life is greater than (equal to) the fit value of the second person's life.

F1-2 implies a symmetry such that $FV(4, 3) = FV(2, 3)$, or in general that $FV(y+x, y) = FV(y-x, y)$. It also follows from F1-2 that the maximal fit value of a life with a given merit level is the life where the person gets exactly what she deserves since then $|w_1-m| = 0$ (which is the minimal value of the function $|x|$).

One might consider that apart from the symmetry implied by F1-2, there should also be a symmetry such that $FV(4, 5) = FV(4, 3)$, that is, if $|w-m_1| =$

$|w\text{-}m_2|$, then $FV(w, m_1) = FV(w, m_2)$. The following principle is a combination of this idea and F1-2:

F1-3: If $|w_1\text{-}m_1| < |w_2\text{-}m_2|$, then $FV(w_1, m_1) > FV(w_2, m_2)$; and
if $|w_1\text{-}m_1| = |w_2\text{-}m_2|$, then $FV(w_1, m_1) = FV(w_2, m_2)$.

In words: If the difference between receipt and merit in a certain life is smaller than (equal to) the difference between receipt and merit in another life, then the fit value of the former life is greater than (equal to) the fit value of the latter life.

This principle implies a further and quite interesting symmetry. It implies that the fit value only depends on the absolute difference between receipt and merit and not on the magnitude of the receipt or the merit. For example, it implies that $FV(2, 2) = FV(100, 100)$, $FV(1, 2) = FV(99, 100)$, and that $FV(1, 2) > FV(98, 100)$. An intuitive support in favour of this implication of F1-3 could be that when a person receives exactly what she deserves, then perfect justice is done, and perfect justice has one and the same value in all situations. Likewise for discrepancies between merit and receipt, that is, the imperfect justice of someone receiving, say, two units less than she deserves has the same value in all situations.[15]

---

[15] If Feldman were to agree with our definition of the intrinsic value of a life, then he would probably disagree with this symmetry. Feldman (1997, p. 212) seem to hold that the value of a life enjoying a deserved one unit of pleasure is two. However, 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 (Feldman 1997, p. 206). Now, since $IV(1, 1) = 2$ and $WV(1, 1) = 1$, it follows that $FV(1,1) = 1$, and since $IV(100, 100) = 200$ and

F1-3 in conjunction with V1 also implies two principles that seem to fit our intuitions about merit and receipt. Since the intrinsic value of a life equals the sum of its welfare and fit values, F1-3 implies that for a given welfare level, the better the fit between receipt and merit, the higher the intrinsic value of a life:

D1: If $|w\text{-}m_1| < |w\text{-}m_2|$, then $IV(w, m_1) > IV(w, m_2)$.

Moreover, F1-3 implies that for any given welfare level, the life with the highest intrinsic value is the life with a perfect match between the welfare level and the merit level. Any deviation between the receipt and the merit decreases the intrinsic value:

D2: If $m_1 = w$ and $m_2 \neq w$, then $IV(w, m_1) > IV(w, m_2)$.

F1-3 has an implication, however, that might be a reason to reject it, given our definition of the intrinsic value of a life. Firstly, F1-3 implies that $FV(1, 2) = FV(99, 100)$ and that $FV(2, 2) = FV(100, 100)$. Assume that we can give one unit of pleasure either to (1, 2) or to (99, 100). It follows from WV1 above that the increase in welfare value will be the same in both

---

$WV(100, 100) = 100$, it follows that $FV(100, 100) = 100$. According to F1-3, however, $FV(1, 1) = FV(100, 100)$. On the other hand, Feldman might very well reject V1 above and instead opt for a definition of the intrinsic value of a life according to which $IV(w, m) = WV(w, m) \times FV(w, m)$. Given this definition, the above evaluations are compatible with F1-3. It has, however, in combination with WV1 above, the odd feature that the intrinsic value of a life with zero pleasure is always zero, irrespective of the fit value. I shall not purse this matter further here.

cases, namely one unit. Since F1-3 implies that $FV(1, 2) = FV(99, 100)$ and that $FV(2, 2) = FV(100, 100)$, the increase in fit value will be the same in both cases (that is, $FV(2, 2) - FV(1, 2) = FV(100, 100) - FV(99, 100)$). Since, from V1 above, $IV(w, m) = WV(w, m) + FV(w, m)$, F1-3 implies that giving one unit of pleasure to (1, 2) increases intrinsic value equally much as giving one unit of pleasure to (99, 100). Yet, it seems reasonable to claim that from the perspective of proportional justice, (1, 2) is much worse off than (99, 100) since she only has half of the pleasure that she deserves, whereas (99, 100) has almost all she deserves. Hence, we should give the one unit of pleasure to (1, 2) rather than (99, 100). Thus, the fit value does depend on the magnitude of the receipt and the merit.

Erik Carlson has suggested a theory that implies this kind of symmetry. The intrinsic value of a life is determined by the following formulas in Carlson's theory:[16]

$IV(w, m) = m + (w - m)^k$ if $m \leq w$, $0 < k < 1$.

$IV(w, m) = m - (m - w)^k$ if $m > w$, $k > 1$.

It follows from above that $IV(1, 2) = 2 - (2 - 1)^m = 1$, $IV(2, 2) = 2 + (2 - 2)^k = 2$, $IV(99, 100) = 100 - (100 - 99)^m = 99$, and that $IV(100, 100) = 100 + (100 - 100)^k = 100$. Again, assume that we can give one unit of pleasure either to (1, 2) or to (99, 100). Since $IV(2, 2) - IV(1, 2) = 2 - 1$ and $IV(100, 100) - IV(99, 100) = 100 - 99$, the increase in intrinsic value will be the same in both cases according to Carlson's theory, namely one unit. Consequently, his theory implies that giving one unit of pleasure to (1, 2)

[16] Carlson (1997), p. 312. I have reformulated Carlson's theory in terms of the notation used here.

increases intrinsic value equally as much as giving one unit of pleasure to (99, 100).

For comparisons of proportions to make sense, we need to measure merit and welfare on at least a ratio scale. Hence the assumptions regarding measurement of merit and welfare in the beginning of this paper. There are pretty well worked out theories about how to construct ratio measurement of welfare, but no one has, to the best of my knowledge, showed how to construct such a scale for measurement of merit, or even discussed the matter. Hence, Carlson might defend his theory with some argument to the effect that we cannot measure merit on a ratio scale, or just some general scepticism in this matter. We do agree that this is a very weak spot in the merit based theories that have been presented in the literature and it might be that talk of proportional justice draws on intuitions that lack a secure foundation. However, since many intuitions regarding desert and merit seem to be about proportional justice, we shall for the moment proceed as if it is possible to measure merit on a ratio scale (more on this later).

Moreover, there is a related argument against F1-3 that is not vulnerable to the above rejoinder. One might very well consider that the fit value of a very deserving person, say a saint, getting exactly what she deserves is greater than the fit value of a barely virtuous person getting what she deserves, or that of a vicious person getting what she deserves. To put it more generally, the higher a person's merit, the higher the fit value of her getting exactly what she deserves:[17]

A2: If $w_1 > w_2$, then $FV(w_1, w_1) > FV(w_2, w_2)$.

---

[17] I am indebted to Tom Hurka for this point.

In other words, assuming that a saint deserves 100 units of welfare and a barely virtuous person deserves 2 units, $FV(100, 100)$ should be greater than $FV(2, 2)$ whereas F1-3 implies that $FV(2, 2) = FV(100, 100)$.[18] Notice that this argument does not presuppose measurement of merit on a ratio scale. It is enough that we can order lives in terms of their merit.

What does Carlson's theory say about the above case? Since he has not separated welfare and fit value in his value function, his theory does not yield any ranking of lives in terms of fit value. We can, however, reformulate his theory along the lines of V1 and WV1 and thus, so to say, factor out the merit component of the intrinsic value of a life:

$$IV(w, m) = w + (m - w + (w - m)^k) \text{ if } m \leq w, 0 < k < 1.$$
$$IV(w, m) = w + (m - w - (m - w)^k) \text{ if } m > w, k > 1.$$

In the above reformulation of Carlson's theory, the expression inside the parenthesis represents the fit value part of the intrinsic value of a life (the

[18] One might think that A2 implies that we should give the one unit of pleasure to (99, 100) rather than (1, 2), contradicting the intuition of proportionality discussed above, since one might think that it follows from A2 that the marginal increase in fit value is greater in the move from (99, 100) to (100, 100) as compared to the move from (1, 2) to (2, 2). This is false, however. Here is a counterexample: Assume, for the sake of example, that $FV = w^{½} - |w-m|$ if $m \leq w \geq 0$. It follows that $FV(1, 2) = 0$, $FV(2, 2) = 2^{½}$, $FV(99, 100) = 99^{½} - 1$, and $FV(100, 100) = 10$. Thus, the increase in fit value (and consequently intrinsic value) is greater if we give one unit of wellbeing to (1, 2) rather than to (99, 100) since $FV(2, 2) - FV(1, 2) = 2^{½} \approx 1.41$ which is greater than $FV(100, 100) - FV(99, 100) = 100 - (99^{½} - 1) \approx 1.05$. Still $FV(100, 100) = 10$ is greater than $FV(2, 2) = 2^{½}$.

first factor, *w*, represents the welfare value part of the intrinsic value). It follows that $FV(2, 2) = 2 – 2 + (2 - 2)^k = 0$ and $FV(100, 100) = 100 – 100 + (100 - 100)^k = 0$. Thus, like F1-3, Carlson's theory assigns the same fit value to a saint and a barely virtuous person getting exactly what they merit. Moreover, since $FV(-100, -100) = -100 – (-100) + (-100 – (-100))^k = 0$, Carlson's theory also assigns the same fit value to a saint and a vicious person getting exactly what they merit. This might strike one as a bit counterintuitive.

One might be able to defend F1-3 with arguments analogous to those we used above in favour of symmetry between the values of over- and under-merited welfare. We could ask whether it is necessary, to capture the intuition that it is better that a saint gets what she deserves rather than that a barely virtuous person gets what he deserves, to incorporate A2 in our theory or if it is sufficient to assign increasing value to increases in the welfare level, other things being equal. Given our definitions of the intrinsic value of a life and the value of welfare (V1 and WV1), it follows that if *FV*(100, 100) = *FV*(2, 2), then *IV*(100, 100) is greater than *IV*(2, 2) since *WV*(100, 100) = 100 is greater than *WV*(2, 2) = 2. Consequently, adding A2 to our theory of desert might not affect the overall ranking of lives and states of affairs, given V1 and WV1, and might in that sense be superfluous.

Be that as it may, there is a further quite shattering problem for F1-3 and Carlson's theory. They both imply that we can make the world better by making ourselves or others less virtuous. Consider two lives (10, 10) and (10, 20). According to F1-3, *FV*(10, 10) > *FV*(10, 20) since there is a better fit in (10, 10) as compared to (10, 20). Since the two lives involve the same amount of welfare, it follows that *IV*(10, 10) > *IV*(10, 20). Likewise, according to Carlson's theory, *IV*(10, 10) = 10 whereas *IV*(10, 20) = $10 - (20\text{-}10)^k = 10 - 10^k < 10$. It follows that *IV*(10, 10) > *IV*(10, 20). Hence, we can make the world better by making people less virtuous. To put it otherwise, a world can better than another world just because people are less virtuous.

This objection presupposes, of course, that people's merit can vary. This seems true on most accounts of merit base. Recall two of Feldman's example of determinants of the merit level: moral worthiness and conscientious efforts. It seems odd that one can make the world better just by making less conscientious efforts or by bringing down one's moral worth.

On some other accounts, it is not as clear that people's merit can vary. Consider need or the idea that people merit a certain level of welfare just because they are persons. If these merit levels are fixed, then a merit theory that only incorporates these merit bases would be immune to the objection above. A lot of more work need to be done, however, to show that merit based on these properties cannot vary. Personhood seems to come in degrees and need seems to vary depending on particular properties of humans, such as the need for certain medicines.

In light of the above discussion, I think we should jettison F1-3 and with it Carlson's theory and consider the weaker F1-2 instead. It does not have the implications of F1-3 just discussed, but one might suspect that it has analogous consequences, for example that we should be indifferent between giving ten units of pleasure to (100, 0) and (-100 ,0) since, according to F1-2, *FV*(100, 0) = *FV*(-100, 0). However, since F1-2 does not imply that *FV*(110 ,0) = *FV*(-90, 0), it does not follow that we should be indifferent between giving ten units of pleasure to (100, 0) and (-100, 0).

Moreover, F1-2 is silent on the comparative intrinsic value of (10, 10) and (10, 20) since it only compares lives with the same merit level.

F1-2 has implications that are desirable from the perspective of distributive justice. It sometimes implies that we should redistribute welfare from people that have more than they deserve to people that have less than they deserve. For example, a population A = ⟨(100, 0), (-100, 0)⟩ is worse than a population B = ⟨(90, 0), (-90, 0)⟩ since WV(A) = WV(B) = 0, FV(90, 0) > FV(100, 0) and FV(-90, 0) > FV(-100, 0).

More importantly, F1-2, in conjunction with V1, V2, and WV1, satisfies a plausible adequacy condition suggested by Carlson. He has proposed that "[if] a consequentialist theory is to alleviate the objection from justice" then it should satisfy the following adequacy condition:

> J1: If *n* units of welfare are to be distributed among a certain number of people with a total merit level of *n*, then the distribution where each person gets exactly what she merits is better than any alternative distribution.[19]

As Carlson puts it, "[a] theory which does not satisfy J1 sometimes allows that some person get more that she deserves, and another person gets less, even though there is no reason in terms of maximizing net pleasure to allow this. Such a theory, it seems, permits us to depart from the requirements of justice for no good reason".[20] As Carlson has shown, Feldman's version of Justicism does not satisfy this condition.[21] Now, since the welfare value is

[19] Carlson (1997), p. 311. I have rephrased Carlson's condition.

[20] Carlson 1997, p. 311.

[21] Carlson (1997), p. 311.

going to be the same for any distribution of a fixed amount of welfare, the population with the highest intrinsic value will be the population with the highest aggregate fit value according to V1 and V2. As we pointed out above, it follows from F1-2 that the maximal fit value of a life with a given merit level is the life where the person gets exactly what she merits. Consequently, the aggregate fit value is maximized when everybody gets exactly what they merit. Hence, F1-2, V1, V2, and WV1 together imply J1.

4. The Second Central Fit-Idea

The next central principle of the fit-idea concerns the relative importance of increases in fit:

> *The Second Central Fit-Idea*: The contributive value of a given increase in fit by a change in the receipt decreases the closer to the merit level one gets.

The intuitive idea behind the second fit-idea is that the greater the mismatch between receipt and merit, the greater the urgency to increase the fit between receipt and merit by adjusting the receipt. Here is an example: Assume that $m = 10$, $w_1 = 5$, $w_2 = 4$, $w_3 = 3$. Then the increase in fit value from a change in pleasure from 4 to 5 is less than the increase in fit value from a change in pleasure from 3 to 4. In other words, $FV(4, 10) - FV(3, 10) > FV(5, 10) - FV(4, 10)$.

Here is how we can formulate the second fit-idea more exactly:

F2-1: If $|e_1|=|e_2|$, $|w_1 - m|>|w_1 + e_1 - m|$, $|w_2 - m|>|w_2 + e_2 - m|$, and

$|w_1 - m| > |w_2 - m|$, then

$FV(w_2 + e_2, m) - FV(w_2, m) < FV(w_1 + e_1, m) - FV(w_1, m)$.

In words: If we can increase the fit between merit and receipt in two lives with the same merit level by adjusting their welfare level up or down by a fixed amount, then the life with the greater difference between receipt and merit will get the greater increase in fit value from the adjustment of its welfare level.

F2-1, in conjunction with F1-2, V1, V2, and WV1, has some attractive implications in regard to distribution of welfare. For a start, it implies that the best distribution of a given amount of welfare between two persons with the same merit level is an equal distribution:

D3: For any m, if $w_1 > w_2 > w_3$ and $w_1 + w_3 = 2w_2$, then

$IV\langle(w_2, m), (w_2, m)\rangle > IV\langle(w_1, m), (w_3, m)\rangle$.

Here is an informal demonstration (a proof can be found in the appendix). Consider the distribution ⟨(3, 10), (0, 10)⟩ and whether it would better to give three units of welfare to (3, 10) to reach ⟨(6, 10), (0, 10)⟩ or to give it to (0, 10) to reach ⟨(3, 10), (3, 10)⟩, that is, whether the unequal or the equal distribution is the best one. Since we are considering a case where a fixed amount of welfare is to be distributed, the welfare value is going to be the same for any distribution. Consequently, it follows from V1 and V2 that the population with the highest fit value is the population with the highest intrinsic value. According to F2-1, if we can increase the fit between merit and receipt in two lives by adjusting their welfare level by a fixed amount, then the life with the greater difference between receipt and merit will get the greater increase in fit value from the adjustment. Clearly, (0, 10) has a greater difference between merit and receipt (i.e., 10) than (3, 10) so we get the greatest increase in fit value if we give (0, 10) three units of welfare rather than (3, 10). In other words, $FV(3, 10) - FV(0, 10) > FV(6, 10) - FV(3, 10)$, which is equivalent with $2FV(3, 10) > FV(6, 10) + FV(0, 10)$. It follows that $IV\langle(3, 10), (3, 10)\rangle > IV\langle(6, 10), (0, 10)\rangle$.

In general, F2-1, F1-2, V1, V2, and WV1, together imply that an equal distribution is the best distribution of any given amount of welfare to any given number of people with the same merit level, that is, a generalization of D3:

D4: For any m, and any populations A = $\langle(w_1, m), (w_2, m),\ldots,(w_n, m)\rangle$ and B = $\langle(q_1, m), (q_2, m),\ldots,(q_n, m)\rangle$, $n \geq 2$, if $w_i \neq w_j$ for some $i, j \leq n$, and $q_i = (w_1 + w_2 + \ldots + w_n)/n$ for all $i \leq n$, then IV(B) > IV(A).

Since the intrinsic value of a population is the sum of the intrinsic value of the lives in the population, that is, the value-function is additively separable, D4 follows from D3. For any population, we can by repeated application of D3 to different pairs of lives in the original population generate successively better populations until we reach a population with an equal distribution of welfare. The final population will be better than the original population by virtue of the transitivity of intrinsic value. Here is an example. Let us say that population A consists of four people: (10, m), (8, m), (6, m), and (4, m). The average pleasure in A is 7 units (28/4). D3 implies that $IV\langle(10, m), (8, m)\rangle < IV\langle(9, m), (9, m)\rangle$ and $IV\langle(6, m), (4, m)\rangle < IV\langle(5, m), (5, m)\rangle$. It follows from this and V2 that IV(A) < $IV\langle(9, m), (9, m), (5, m), (5, m)\rangle$. Applying D3 again, we get that

$IV\langle(9, m), (5, m)\rangle < IV\langle(7, m), (7, m)\rangle$. Again, it follows from this and V2 that $IV\langle(9, m), (9, m), (5, m), (5, m)\rangle < IV\langle(7, m), (7, m), (7, m), (7, m)\rangle$ which, by transitivity, is better than A.

We should appreciate that F2-1 implies D4. Carlson suggests that D4 is a crucial requirement for a theory to meet the objection from justice. As he puts it, '[i]f everybody has equal desert, it is a breach of justice to give more to one person than to another'.[22] As Carlson has pointed out, Feldman's theory violates D4 and directs us to distribute welfare unequally although everyone deserves the same level of welfare.[23]

As with the first fit-idea, there is a stronger interpretation of the second fit-idea:

F2-2: If $|e_1|=|e_2|$, $|w_1 - m_1|>|w_1 + e_1 - m_1|$, $|w_2 - m_2|>|w_2 + e_2 - m_2|$, and $|w_1 - m_1|>|w_2 - m_2|$, then
$FV(w_2 + e_2, m_2) - FV(w_2, m_2) < FV(w_1 + e_1, m_1) - FV(w_1, m_1)$.

In words: If we can increase the fit between merit and receipt in two lives by adjusting their welfare level up or down by a fixed amount, then the life with the greater difference between receipt and merit will get the greater increase in fit value from the adjustment of its welfare level.

F2-2 is much stronger than F2-1 since it is applicable to lives with different merit levels. According to F2-2, for example, a change from (5, 10) to (6, 10) yields a greater increase in fit value than a change from (2, 4) to (3, 4) since there is a greater difference between receipt and merit in the former case as compared to the latter.

---

[22] Carlson 1997, p. 311.

[23] Carlson (1997), pp. 309-311.

F2-2 has a very interesting implication. Assume that we are to distribute six units of welfare to two persons who currently enjoy zero welfare but who deserve eight and four units of welfare respectively: 〈(0, 8), (0, 4)〉. Since we are considering cases where a fixed amount of welfare is to be distributed, the welfare value is going to be the same for any distribution. Consequently, it follows from V1, V2, and WV1 that the population with the highest fit value is the population with the highest intrinsic value. Clearly, (0, 8) has the greatest difference between merit and receipt (i.e., 8) so we get the greatest increase in fit value if we give her the first unit of welfare, according to F2-2. This continues to hold true until we reach the distribution 〈(4, 8), (0,4)〉. In this distribution, we have the same difference between merit and receipt for both individuals (i.e., 4), so it does not follow from F2-2 which one to give the next unit of welfare in order to maximize the increase in fit value. Assume first that this is the case if we give it to the first individual to reach the distribution 〈(5, 8), (0, 4)〉. Now the second person has the greatest difference between merit and receipt (i.e., 4) so we get the greatest increase in fit value if we give her the next unit of welfare to reach the distribution 〈(5, 8), (1, 4)〉. Assume secondly, contrary to above, that we maximize the increase in fit value if we given one unit of welfare to the second individual in 〈(4, 8), (0, 4)〉 to reach the distribution 〈(4, 8), (1, 4)〉. Now the first individual has the greatest difference between merit and receipt (i.e., 4) so we get the greatest increase in fit value if we give her the next unit of welfare to again reach the distribution 〈(5, 8), (1, 4)〉. We have now distributed the six units of welfare so that we in each step have maximized the increase in fit value. Hence, 〈(5, 8), (1, 4)〉 has the highest fit value of all possible distributions of six units of welfare given the merit level of the involved individuals, and thus the maximal intrinsic value. In general, F2-2, in conjunction with V1, V2, and WV1, implies the following principle:

> J2: If a fixed amount of welfare is to be distributed among a certain number of people, then the distribution where each person gets the same difference between her merit and receipt is better than any alternative distribution.

Notice that D4 and J1 are special cases of J2. This principle is intuitively appealing but one might object to J2, and thus F2-2, in the same way we objected to F1-3: It ignores considerations of proportional justice. Like F1-3, F2-2 implies indifference between giving one unit of pleasure either to (1, 2) or to (99, 100) because the difference between receipt and merit is the same in both cases. Moreover, it has implications that are perpendicular to proportional intuitions. Consider again the distribution of six units of welfare between (0, 8) and (0, 4). As we showed above, F2-2 implies that the optimal distribution is 〈(5, 8), (1, 4)〉. One might object that the first individual is getting more than half of what she deserves whereas the second individual is getting one a quarter of what she deserves. Would the distribution 〈(4, 8), (2, 4)〉 not be better from the perspective of proportional justice since then both individuals get exactly half of what they deserve?
The problem with F2-2 is that incorporates a specific interpretation of what it means to get "closer" to the merit level, namely a measure of closeness that exclusively focuses on the gap between receipt and merit and which is the same over different merit levels. Those who believe in proportional justice will challenge this and claim that closeness should be understood in terms of the proportions between receipt and merit. Thus the correct

interpretation of the second central fit-idea should be stated in terms of the differences in proportions between receipt and merit, perhaps along these lines:

> F2-3 (roughly): If we can increase the fit between merit and receipt in two lives by adjusting their welfare level up or down by a fixed amount, then the life with the smallest proportion of receipt relative its merit will get the greater increase in fit value from the adjustment of its welfare level.

Consider again the case of giving one unit of welfare to either (4, 8) or (1, 4). F2-2 implied that we should give it to the first individual to reach the distribution ⟨(5, 8), (1, 4)⟩ since she has the greatest difference between merit and receipt (i.e., 4). However, the first individual has half of what she merits (4/8 = ½) whereas the second individual has one quarter of what she merits (¼). Hence, F2-3 implies that we maximize merit value and thus intrinsic value if we give it to the second individual to reach the distribution ⟨(4, 8), (2, 4)⟩.

So perhaps we should go for an interpretation of the second central fit-idea along the lines of the proportionality view? There are, however, well-known problems with this approach. Firstly, consider who to give one unit of welfare in the distributions ⟨(1, 100), (0, 1)⟩. The first individual has 1/100 of what she merits whereas the second individual has zero (0/1) of what she merits. Thus we should give the second individual one unit of welfare to maximize desert value according to F2-3. This seems quite counterintuitive given the first individual's great gap between receipt and merit. Although this is not a decisive argument against F2-3, it shows that it at least has to be complemented with some principle regarding the gap between receipt and merit.

Now consider the distribution 〈(-1, 100), (0, 1)〉. Here the first individual has -1/100 of what she merits whereas the second individual has zero of what she merits. Does that mean, according to F2-3, that we should give one unit to (-1, 100) to maximize desert value since -1/100 < 0? That seems intuitively correct, since (0, 1) has almost a perfect match between receipt and merit whereas there is a great discrepancy between receipt and merit for (-1, 100). Yet, what about 〈(-0.9, 0.1), (1, 100)〉? Given our reasoning in the preceding case, we have to claim that we should give one unit to (-0.9, 0.1) since -9 < 1/100. However, (-0.9, 0.1) has almost a perfect match between receipt and merit whereas this is definitely not true for (1, 100). These two judgments do not fit together very well.

It gets worse, however. The proportionality view runs into a host of problem in case where people merit zero units of welfare. Consider the distribution 〈(-100, 0), (-1, 0)〉. Who has the smallest proportion of receipt relative her merit in this case? Since division by zero is undefined, it seems that the proportional view is not even applicable to these cases. Or consider the distribution 〈(-100, 0), (0, 0)〉. Here we would like to say, intuitively, that the second individual has a perfect match between receipt and merit and that we maximize desert value if we give one unit of welfare to the first individual. This is not possible to say on the proportionality view.

To the best of my knowledge, no one has given a good answer to the above objections to F2-3 and the proportionality view. As I said above, many intuitions regarding desert and merit seem to be about proportional justice, so it is a big loss to throw them out. It might be that one could develop some interpretation of the second central fit idea that incorporated elements from both F2-2 and F2-3 and avoided the problems of F2-3 and muted the ignorance of proportional concerns in F2-2. I am sorry to say that I have no such “compromise” theory to provide, and I suspect that it would inherit the problems of both interpretations. Now, given the problems with F2-2 and F2-3, I think it is clear from the above discussion that we should stick with F2-1. So, I suggest that we tentatively define Additively Separable Fit Meritarianism, or ASFM for short, as the conjunction of V1, V2, WV1, F1-2 and F2-1. I consider this a “default” theory, that is, a comparatively simple theory that can serve as the starting point for more complicated theories.[24] One such complication that we have not yet considered is whether we should add considerations of comparative desert to our theory, to which we shall now turn.

5. Comparative and Noncomparative Desert

Assume that Wlodek and Ewa both deserve ten apples but there are only ten apples available. Should we then give 10 apples to Wlodek and zero to Ewa, or vice versa, or perhaps five each? One might think that from the perspective of what we could call noncomparative desert, that is, when we only focus on the relation between an individual’s desert and receipt, it is better to give either ten apples to Wlodek or Ewa, since then at least one person gets exactly what he or she deserves. However, from the perspective of what we could call comparative desert, the best distribution is to give Wlodek and Ewa five apples each, or perhaps no apples. The reasoning would be that they have the same desert and thus no one should be better or worse of than the other. Hence, there seems to be two kinds of desert

[24] Cf. Broome (2005), p. 128.

claims that also might be in conflict: comparative and noncomparative desert. As Shelly Kagan puts it: “It also matters how I am doing compared to you, in light of how (noncomparatively) deserving we are. That is the basic idea of comparative desert. Thus … if I am just as deserving as you are … then I should be doing as well as you…”.[25] Both Kagan and Tom Hurka argues that an “adequate theory of desert will need to include not only noncomparative principles, but comparative ones as well”.[26]

What does ASFM say about a case like the above? It is a paradigmatically noncomparative theory. All merit claims are individual in the sense that it assigns value only to features of individual lives and there is nothing in the theory that directly addresses comparisons of desert or distributional patterns. However, ASFM implies that the best distribution in the case above is that the apples are shared equally since ASFM implies D4, that is, that an equal distribution is the best distribution of any given amount of welfare to any given number of people with the same merit level. Hence, we do not need any further comparative desert principle to get the desired distributional implications.

Tom Hurka, however, argues that there are some gaps in “individualistic principles” (i.e., noncomparative desert principles) such as ASFM that need to be filled with comparative desert principles or, as he calls it, “holistic principles”. He writes:

> Imagine that A and B are equally virtuous and both enjoy less pleasure than they ideally deserve … but B enjoys considerably more pleasure than A. There is a disproportion in this situation, and the individualistic principle says that it

[25] Kagan (2003).

[26] Kagan (2004), p. 93. See also Hurka (2003). pp. 49-50.

> would be better if this were removed by increasing A's pleasure to the level of B's. But what if the disproportion is removed in the opposite way, by reducing B's pleasure to A's? The principle says that this 'levelling down' only makes the situation less good, by replacing a greater desert-good with a lesser one. But I think many of us will see it as in one respect an improvement. The holistic principle captures this view, saying that the reduction in B's pleasure removes a holistic evil of disproportion.[27]

One might think that examples such as this shows that the value of a population cannot be captured by an additively separable function of the value of the lives in the population since comparative desert depends on relations between lives, namely the differences between some lives' desert levels and receipts and others'. Hence, if Hurka is right, then we seem to have an argument against the approach exploited in this paper.

I do not think, however, that Hurka's example shows that we have to jettison separability. I shall first describe a way of accommodating Hurka's intuition that preserves separability. I shall then turn to my main objection against his example and argue that it is not enough to show that an adequate theory of desert has to include comparative desert principles.

Even if we grant Hurka's intuition above, it does not show that separability is false since we could disperse the value of comparative desert over individual lives. Instead of V1, the value of a life would be determined by

V1′: $IV(w, m) = WV(w, m) + FV(w, m) + CV(w, m)$.

[27] Hurka (2003), p. 50.

where *CV* represent the individual's share of the comparative desert value. Of course, for this being more than just a technical manoeuvre, we need a justification for dispersing the value of comparative desert in this manner. Here is a suggestion. The bad thing with the distribution in Hurka's example is that A has less pleasure than B although A deserves (comparatively) to have the same amount of pleasure as B. This is arguably a bad feature of A's life. Thus, the badness of the disproportionate distribution of pleasure in Hurka's example can be located in the value of A's life, and the value of a population will still be an additively separable function of the values of the lives in the population.[28]

At any rate, I am not convinced by Hurka's argument for the need of comparative desert principles in the first place. Firstly, one might doubt that "many of us" will consider levelling down in Hurka's example above an improvement from the perspective of desert, and Hurka's lack of reference to defenders of this view is telling.[29] In the general debate about equality, a common view is that levelling down is not an improvement in

---

[28] Broome (2005), pp. 110-2, and Broome (1991) use this strategy for defending dispersion of the value of equality. Some might consider this defence of separability incoherent since it makes the intrinsic value of a life dependent on relational properties, such as whether there are other lives in the population with the same merit level that enjoys more pleasure. This is, however, no problem if we adopt a conditionalist view on intrinsic value (or final value as it is often called in this context and which might be a more apt term for what I have called intrinsic value in this paper). See Olson (2004). Thanks to Jonas Olson for pointing this out to me.

[29] Hurka (2003), p. 50, only refers to a levelling down objector who calls it "plainly unacceptable".

any respect.[30] Secondly, and more importantly, even if we grant that there is an intuition to the effect that levelling down is in one respect an improvement in such situations (as I am inclined to do), it is not enough to justify the inclusion of comparative desert principles in our theory of desert since there might be countervailing reasons. One such theoretical reason is whether a stipulated value will make any difference in some situation to the overall ranking of outcomes. If a stipulated value never makes a difference to the overall ranking, then we have a theoretical reason to omit it from our theory. Parsimony is a virtue of theories and Ockham's razor is as applicable to scientific theories as it is to axiological theories. Applied to scientific theories, Ockham's razor instructs us to eliminate those assumptions that make no difference in the observable predictions of the theory. Applied to axiological theories, we could render this injunction as telling us to eliminate those assumptions that, for any set of logically possible outcomes, make no difference for the all things considered ranking of the outcomes. In other words, if the value of the factor $CV(w, m)$ in V1′ never makes a difference to the overall ranking of outcomes, then we should eliminate it from the theory.

To the best of my knowledge, Hurka never explicitly says whether he thinks comparative desert consideration ever makes a difference for the all things considered ranking of outcomes. There are, however, strong reasons against that being the case in the situation described by Hurka since it also involves two significant worsenings. Levelling down in the above case makes the situation worse in respect to welfare --- B's welfare is decreased and no one's welfare is increased --- and there is an increased mismatch between receipt and desert --- B is getting even less than she deserves.

[30] See e.g., Parfit (1993).

Thus, the intuition that levelling down in the above situation clearly makes the situation worse all things considered is quite robust, I surmise, and Hurka's holistic principle will have no impact on the overall axiological evaluation of the situation.

Hurka has a second argument in favour of holistic principles:

> …[I]magine that A and B both enjoy more pleasure than they deserve … but B again has more. The individualistic principle says it would only make things worse if A were raised to B's level, but the holistic principle, again in my view more plausibly, says this levelling up would in one respect improve them.[31]

Again, I am unsure whether the addition of a holistic principle will add much of interest to our thinking about cases like this. ASFM agrees with Hurka that in this case there is both an improvement and a worsening. There is an improvement in respect to welfare and a decline in respect to the match between receipt and desert. This seems enough for judging the all things considered desirability of the two possible outcomes.

It is fair to ask, however, how these two respects should be weighed against each other. The principles that we have so far used to define ASFM are silent on this issue. Hence, ASFM is compatible with the view that a decline in fit value sometimes, always, or never, outweighs an increase in welfare value. So far, we have not taken a stand on this issue. If a decline in fit sometimes can outweigh an increase in welfare, then there is a conceptual space for comparative desert consideration to affect the overall ranking of outcomes. It could then be the case that although the increase in

---

[31] Hurka (2003), p. 50. Kagan (2003), p. 97, uses a similar case as an argument in favour of comparative desert principles.

welfare cannot outweigh the decrease in fit, the combined value of the increase in welfare and a better satisfaction of comparative desert considerations can outweigh the decrease in fit value. I would suggest, however, that a reasonable Meritarian axiology should satisfy the following dominance condition:

P: For any m, w, and $e > 0$, $IV(w + e, m) > IV(w, m)$.

In words: An improvement in welfare always increases the intrinsic value of a life. This is a much stronger claim than the one invoked in the argument against Hurka's first example above. It implies that a decline in the fit value because of an increase in welfare (a person getting more than she merits) cannot outweigh the increase in welfare value. Moreover, it implies that levelling down always makes an outcome worse overall. I guess that this principle will be attractive to those of us that are not prepared to sacrifice wellbeing on the altar of justice, or to, as it were, throw wellbeing in the sea when there are people who could otherwise enjoy it.

If we add P to the defining features of ASFM, then it will imply that it would be an overall improvement if A were raised to B's level in Hurka's second case. Not only will there be no need for a supplementary holistic principle to reach this result, there will also be no space for such principle to affect the overall ranking of the outcomes in cases of this kind.

Similarly for a case described by Kagan:

> …[I]magine that A is more deserving than B, but B has more than A. Even if A is already at his peak [has exactly as much as he deserves], isn't there

something to be said in favor of improving his lot even more, so that he has more than B?

ASFM concurs since an improvement in welfare is always a consideration in favour of an outcome, and, in conjunction with P, it yields that it is an improvement in the outcome's intrinsic value. Again, there will be no space for comparative desert principles to affect the overall ranking in situations of this type.

P will not be liked by people with strong retributivist intuitions, however. According to P, if we can increase the welfare of a very undeserving individual that already enjoys more than she deserves, then that will be an overall improvement in intrinsic value, other things being equal. Those with strong retributivist intuitions will not agree but rather claim that such an increase in welfare do not make an outcome better and might even make it worse. Hence, they will reject P and without P there is space for comparative desert consideration to have an impact on the overall ranking of outcomes.

Nevertheless, comparative desert considerations along the lines of Hurka and Kagan are not very appealing for retributivists, I surmise. Consider again Hurka's second example above and assume that the two involved individuals are vicious criminals. One of the criminals got away and is enjoying a luxuries lifestyle on an island in the Caribbean. The other one was caught and imprisoned but a bleeding heart liberal judge gave him a bit too lenient a punishment. This will be upsetting for retributivist. If we then told them that there is a reason from desert to set the other criminal free, buy him a ticket to the Caribbean, and hand over a bag containing a million dollars, I am afraid they will not be very impressed but rather stretch for their gun.

In summary, I propose the following dilemma for proponents of comparative desert: One can either endorse P and reject retributivism or one can reject P and endorse retributivism. If one endorses P, then there is no space for comparative desert principles to affect the all things considered ranking of outcomes and such principles are thus superfluous. If one is a retributivist, then comparative desert principles have strongly counterintuitive implications.

Kagan suggests two constraints on an adequate theory of comparative desert. The first we have already discussed: If two person are equally deserving, then they should all be equally well off. The second constraint says that if one person is more deserving than another, then the former should be better off than the latter. Do we perhaps here have a comparative concern that cannot be captured by a noncomparative principle? Again, I do not think so. As we showed above, ASFM implies J1. It follows that if we are to distribute twelve units of welfare to two persons *x* and *y*, and *x* deserves eight units and *y* deserves four units, then the best distribution according to ASFM is that *x* receives eight units and *y* receives four units. Hence, in this case ASFM yields the desired result by itself.

ASFM, as we have defined it, is so far silent on cases where there is an unequal distribution of merit and too much or too little welfare to distribute. As we saw above, F2-2 implies J2. Recall the case in which we have only six units of welfare to distribute to *x* and *y*. F2-2 implies that the best distribution is that *x* receives five units and *y* receives one unit, that is, in the optimal distribution, Kagan's second constraint is satisfied. F2-3 has analogous implications. We rejected these principles because they had a number of counterintuitive implications. However, ASFM can, I surmise, be complemented with some weaker individualist principle, or some

combination of F2-2 and F2-3, that will imply that if *x* has higher merit than *y*, then for any amount of welfare, *x* will have higher welfare than *y* in the optimal distribution.
It remains to be seen, I surmise, whether there is any real work left for comparative desert principles when we spell out our theory of noncomparative desert along the lines of ASFM.

Appendix: Proof of D3

We shall show that F1-2, F2-1, V1, V2, and WV1 together imply:

D3: For any m, if $w_1 > w_2 > w_3$ and $w_1 + w_3 = 2w_2$, then
$IV\langle(w_2, m), (w_2, m)\rangle > IV\langle(w_1, m), (w_3, m)\rangle$.

We shall divide the proof into three different cases depending on the merit level: $m > w_1 - e/2$, $m < w_3 - e/2$, and $w_1 - e/2 \geq m \geq w_3 + e/2$, where e is the difference in welfare between w_1 and w_2. Let us start with $m > w_1 - e/2$. Let

(1) w_1, w_2, w_3 be three welfare levels, m be a merit level, and e a difference in welfare levels such that $w_1 > w_2 > w_3$, $w_1 + w_3 = 2w_2$, $e = w_1 - w_2$, and $m > w_1 - e/2$.
(2) A and B be two populations such that $A=\langle(w_2, m), (w_2, m)\rangle$ and $B=\langle(w_1, m), (w_3, m)\rangle$.

It follows from (1), (2), and the definition of the welfare value of a population that

(3) $WV(\text{A}) = WV(\text{B})$.

From (2), (3), and the definition of the intrinsic value of a population, it follows that

(4) $IV(\text{A}) > IV(\text{B})$ iff $FV(\text{A}) > FV(\text{B})$ iff $2FV(w_2, m) > FV(w_1, m) + FV(w_3, m)$.

That is, the ranking of A and B will be decided by their respective fit values. Since $m > w_1 - e/2$ and $e = w_1 - w_2$ (from (1)), we get

(5) $|w_3 - m|>|w_2 - m|$

(6) $|w_2 - m|>|w_2 + e - m|$

(7) $|w_3 - m|>|w_3 + e - m|$.

F2-1 and ((5)-(7)) imply

(8) $FV(w_2 + e, m) - FV(w_2, m) < FV(w_3 + e, m) - FV(w_3, m)$.

Since $w_1 + w_3 = 2w_2$ and $w_1 - w_2 = e$ (from (1)), it follows that

(9) $w_1 - w_2 = w_2 - w_3 = e$

which implies

(10) $w_1 = w_2 + e$ and $w_2 = w_3 + e$

which in turn yields that (8) is equivalent with

(11) $FV(w_1, m) - FV(w_2, m) < FV(w_2, m) - FV(w_3, m)$.

By rearranging the terms in (11) we get

(12) $FV(w_1, m) + FV(w_3, m) < FV(w_2, m) + FV(w_2,m)$

which together with (4) implies that

(13) IV(A) > IV(B). Q.E.D.

The proof for $w_3 > m + e/2$ follows the same pattern as the above proof so I shall not spell it out here. The proof for $w_1 - e/2 \geq m \geq w_3 + e/2$ is quite simple. Assume first that $w_1 - e/2 \geq m > w_3 + e/2$. It follows that $|w_2 - m| \leq |w_1 - m|$ and $|w_2 - m| < |w_3 - m|$ (just look at the maximal and minimal values of m). F1-2 thus implies that $FV(w_2, m) \geq FV(w_1, m)$ and $FV(w_2, m) > FV(w_3, m)$. It follows that $FV(w_2, m) + FV(w_2, m) > FV(w_1, m) + FV(w_3, m)$ which together with (4) above implies that IV(A) > IV(B).
Assume secondly that $w_1 - e/2 > m \geq w_3 + e/2$. It follows that $|w_2 - m| < |w_1 - m|$ and $|w_2 - m| \leq |w_3 - m|$. F1-2 thus implies that $FV(w_2, m) > FV(w_1, m)$ and $FV(w_2, m) \geq FV(w_3, m)$. It follows that $FV(w_2, m) + FV(w_2, m) > FV(w_1, m) + FV(w_3, m)$ which together with (4) above implies that IV(A) > IV(B). Thus, for any m, if $w_1 > w_2 > w_3$ and $w_1 + w_3 = 2w_2$, then F2-1 implies

that $IV\langle(w_2, m), (w_2, m)\rangle > IV\langle(w_1, m), (w_3, m)\rangle$, which is exactly what D3 states. Q.E.D.[32]

---

[32] I would like to thank Ben Bradley, Krister Bykvist, Erik Carlson, Fred Feldman, Marc Fleurbaey, Tom Hurka, Jonas Olson, Wlodek Rabinowicz, Stuart Rachels, and Michael Zimmerman for fruitful discussions and comments. Earlier versions of this paper were presented at the Dept. of Philosophy, Uppsala University, April 06; and at *Syracuse Philosophy Annual Workshop & Network (SPAWN)*, July 2006. I would like to thank the participants at these occasions for their stimulating criticism, and especially Noah Lemos for his commentary at the superb SPAWN Conference. Thanks also to Oxford Uehiro Centre for Practical Ethics and Jesus College, Oxford, for being such generous hosts during some of the time when this paper were written. Financial support from the Bank of Sweden Tercentenary Foundation and the Swedish Collegium for Advanced Study is gratefully acknowledged.

References

Arneson, Richard (1989), “Equality of Opportunity for Welfare”, *Philosophical Studies* 56, pp. 77-93.

Arrhenius, Gustaf (2000), *Future Generations: A Challenge for Moral Theory*, F.D.-Diss., University Printers, Uppsala.

------------ (2003), ‘Feldman’s Desert-Adjusted Utilitarianism and Population Ethics’, *Utilitas*, Vol. **15**, No. 2, pp. 225-236.

------------ (2005) ‘Superiority in Value’, *Philosophical Studies*.

------------ (2006) ‘Desert as Fit: An Axiomatic Analysis’, in Richard Feldman, Kris McDaniel, Jason R. Raibley and Michael J. Zimmerman (eds.) *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman,* Ashgate.

Carlson, Erik (1997), ‘Consequentialism, Distribution and Desert’, *Utilitas*, Vol. **9**, No. 3, pp. 307-318.

Broome, John (1991), *Weighing Goods*, Oxford: Basil Blackwell.

——— (2005), *Weighing Lives*, Oxford UP, Oxford.

Cohen, Gerald A (1989), “On the Currency of Egalitarian Justice”, *Ethics* 99.

——— (1993), “Equality of What? On Welfare, Goods and Capabilities”, in M. Nussbaum & A. Sen (eds.), *The Quality of Life*, Oxford: Clarendon Press.

Cupit, Geoffrey (1996), ‘Desert and Responsibility,’ *Canadian Journal of Philosophy* 26: 83-100.

Danielsson, Sven (1974), *Two Papers on Rationality and Group Preferences*, Filosofiska Studier no. 21, Uppsala: Filosofiska föreningen och filosofiska institutionen.

Dworkin, Ronald (2000), *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge, Mass.: Harvard University Press.

——— (1981a), “What is Equality? Part 1: Equality of Welfare”, *Philosophy and Public Affairs* 10, no. 3.

——— (1981b), “What is Equality? Part 2: Equality of Resources”, *Philosophy and Public Affairs* 10, no. 4.

Feldman, Fred (1997), *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy,* Cambridge University Press, Cambridge.

------------ (2000), ‘Basic Intrinsic Value’, *Philosophical Studies*, Vol. **99**, No. 3, pp. 319-346.

Hurka, Thomas (1998), ‘Two Kinds of Organic Unity’, *The Journal of Ethics* 2, pp. 299-320.

------------ (2001a), ‘The Common Structure of Virtue and Desert’, *Ethics* 112, pp. 6-31.
------------ (2001b), *Virtue, Vice, and Value,* Oxford University Press, New York.
------------ (2003), ‘Desert: Individualistic and Holistic’ in Serena Olsaretti (ed.) *Desert and Justice*, Clarendon Press, Oxford.
Kagan, Shelly (2003) ‘Comparative Desert’ in Serena Olsaretti (ed.) *Desert and Justice*, Clarendon Press, Oxford.
McLeod, Owen (2003), ‘Desert’ in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/fall2003/entries/desert/>.
Miller, David (2003), ‘Comparative and Noncomparative Desert’ in Serena Olsaretti (ed.) *Desert and Justice*, Clarendon Press, Oxford.
Olson, Jonas (2004), “Intrinsicalism and Conditionalism about Final Value”, *Ethical Theory and Moral Practice* 7.
Parfit, Derek (1984 [1991]), *Reasons and Persons*, Clarendon Press, Oxford.
------------ (1993) “Does Equality Matter?” in J. R. Richards (ed.), *Philosophical Problems of Equality*, Milton Keynes: The Open University.
Persson, Ingmar (1997), ‘Ambiguities in Feldman’s Desert-adjusted Values’, *Utilitas*, Vol. **9**, No. 3, pp. 319-328.
Rachels, James (1978), ‘What People Deserve’ in Arthur, John and Shaw, William, (eds.) *Justice and Economic Distribution*, , Englewood Cliffs, NJ: Prentice-Hall: pp. 167-196.
Rawls, John (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press
Sadurski, Wojceich (1985), *Giving Desert its Due*, Dordrecht: D. Reidel Publishing.
Sen, Amartya (1980), “Equality of What?” in S. McMurrin (ed.), *The Tanner Lecture on Human Values*, Cambridge: Cambridge University Press.
Vallentyne, Peter (1995), ‘Taking Justice Too Seriously’, *Utilitas*, Vol. 7, No. 2, pp. 207-216.
Zimmerman, Michael J. (2001), *The Nature of Intrinsic Value*, Rowman & Littlefield, Lanham, Md.