https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

BENEFITS, HOLISM, AND THE AGGREGATION OF VALUE*

By David McNaughton and Piers Rawling

I. Introduction

An aggregationist about value, as we shall think of her, claims that the value of the whole (perhaps a state of affairs or a world) is equal to the sum of the values of its parts (the constituents of the state or world). Consider a simple form of hedonism according to which the value of a world is based solely on the amount of pleasure therein, where it is supposed that sentient beings are the repositories of pleasure in the same way that bottles (or cartons) are the repositories of milk. On this view, we might measure the value of a being by measuring how pleasant her life is, and determine the value of the world by summing the values of the beings. This, then, is a simple aggregationist theory. By contrast, in *Principia Ethica* G. E. Moore famously claims that there are "organic wholes," the intrinsic values of which differ "from the sum[s] of the values of [their] parts."[1] This is his "principle of organic unities."[2]

What does a holist about value claim? Perhaps only that the value of a whole cannot be determined by looking merely at the values of its parts—in which case Moore is a holist. But there is another possible account: the holist might claim that, given a whole, the value of its parts cannot be determined in isolation from one another. The simple hedonism outlined above is not holistic in either sense. The value of each being can be determined by looking only at that being, so the theory is not holistic in the second sense. And the value of the whole can be determined by looking only at the values of its parts, so it is not holistic in the first sense either.

These two forms of holism are, however, independent of one another. Moore is a holist in the first sense but not the second. We are holists in the second sense but not the first—or, at least, we favor aggregation for the case in which the bearers of value are states of affairs.

Finally, where do benefits fit in? Our simple hedonist might think that an increase in pleasure for an individual is of benefit to her—indeed, the

* We thank the other contributors to this volume, its editors, and participants at the 2006 conference on Moral Particularism held in Bled, Slovenia, for helpful comments.

[1] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1966), 36. (Originally published in 1903.)

[2] Ibid., 84.

doi:10.1017/S0265052509090141

© 2008 Social Philosophy & Policy Foundation. Printed in the USA.

hedonist would presumably hold that the only kind of benefit is an increase in pleasure. Suppose we combine this hedonism with a simple form of consequentialism according to which our duty at every point is to maximize the good. Then we have a simple form of utilitarianism according to which we should maximize pleasure—or, given the foregoing, maximize the net benefit. On this account, however, a benefit for someone is simply an increase in the general good that happens to take place "in" her. And when comparing two states of affairs, all that is of interest, vis-à-vis benefits, is the question: In which state is the sum of the benefits the greatest? The issue of who gets the benefits is irrelevant. But on an alternative view, which we favor, the notion of benefit is genuinely relational: in assessing a state of affairs, we are interested not only in the sum of the benefits but also their distribution.

Section II of our essay is devoted to discussing these issues in a fairly abstract way; in Section III, we turn to a diverse pair of applications: Derek Parfit's "Repugnant Conclusion" and the role of Pareto-optimality.

II. Value and Benefits

A. Moore's holism

In *Principia Ethica,* Moore claims that "[t]he value of a whole must not be assumed to be the same as the sum of the values of its parts."[3] Given some whole to evaluate, one should, according to Moore, consider the parts in isolation (as if they were the only occupant of the universe) and determine their intrinsic values. (The intrinsic value of an entity is the value it has "'in itself', as opposed to any value it may have as a means to achieving something of intrinsic value.")[4] But the sum of these intrinsic values of the parts need not equal the value of the whole. Rather, it may well be the case, says Moore, that in addition to this sum, there is also a value to their combination. The intrinsic value of the whole is the sum of the intrinsic values of its parts plus the value of their combination.

Moore's general method is laid out on page 214 and the pages that follow, where he considers, among other things, the "theory of vindictive punishment"[5]—perhaps more commonly known now as "retributivism." On a simple utilitarian account, punishment is justified because, although its immediate pain is bad, it is good in the long run due to its tendency to reduce future pain by keeping offenders off the streets and discouraging others from committing crimes. But Moore seeks to give a value-based *retributivist* account of punishment—that is, a value-based justification of punishment that does not appeal to its effects on the future crime rate. Suppose we knew that punishing A for his crime

[3] Ibid., 28.
[4] Ibid., 21.
[5] Ibid., 214.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

would have no deterrent effect, and that he would not offend again regardless of how we treat him; then a simple utilitarian might well see the punishment as unjustified. But a retributivist would disagree: A should be punished for his crime regardless of whether the punishment is effective in reducing future crime.

Here is Moore's account. The wickedness of the crime is bad, but so is the pain of the punishment ("we have a whole . . . exclusively composed of two great positive evils—wickedness and pain").[6] So how are matters improved by the imposition of punishment? In Moore's terminology, the improvement is due to the fact that the "value as a whole" ("which arises solely *from the combination* of two or more things")[7] of crime followed by punishment is positive, where the value as a whole is to be distinguished from the value "on the whole" (it is this latter value that is the intrinsic value of the whole). The latter is the sum of the following: the value of the crime plus the value of the punishment plus their value as a whole.

Here is an overly simple numerical illustration. Suppose the intrinsic values of the parts are:

V(crime) = −8
V(pain inflicted as punishment) = −5
(V(x) is the intrinsic value of x)

Moore's idea is that the value of the combination might be positive—say:

V(combination) = +6 = value as a whole

The value of the crime and punishment *as a whole,* then, is +6. Its value *on the whole* is given by the sum of the three values:

V(on the whole) = −8 + −5 + 6 = −7

Note that −7 > −8, so the situation is better on the whole as a result of the addition of the negatively valued punishment.

Moore counts only the crime and the punishment as parts; hence, the intrinsic value of the whole is not equal to the sum of the values of its parts: $-7 \neq -8 + -5$.

If we take zero as the point of indifference, Moore preserves the intuition that although the addition of the punishment does improve matters (from −8 to −7), the overall situation is still bad (−7 < 0). He also preserves

[6] Ibid., 216.
[7] Ibid., 215; emphasis in the original. Moore offers two different interpretations of this phrase, and labels this sense (1).

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

the intuition that the punishment must be *bad* in order for it to be *punishment*.

However, we are not convinced by Moore's position. We will give our preferred alternative, and then compare the two.

B. An alternative holism

Moore defines an organic unity as a whole the intrinsic value of which does not equal the sum of the intrinsic values of its parts.[8] He takes recourse to organic unities lest the intrinsic value of a whole be mistakenly calculated by looking only to the isolated intrinsic values of its parts. However, we are suspicious of the isolation test for reasons discussed below, and, once it is abandoned, the notion of an organic unity in Moore's sense becomes subject to challenge. Why be committed to the existence of Moorean organic unities unless one is committed to the claim that the parts of a whole can be evaluated independently of that whole? There may be other potential methods for such independent evaluation apart from the isolation test, but we see little prospect for their success.

Moore avoids the mistake of calculating the value of a whole by summing the values of its parts in isolation. But there is an alternative picture that also avoids this error. We contend that there are indeed wholes that have more to them than is evident from their parts taken by themselves,[9] but we do not conclude from this that the value of such a whole cannot equal the sum of the values of its parts. Rather, we reject both Moore's isolation test and his account of organic unities. We claim, instead, that parts can only be evaluated in a way relevant to the whole of which they are a part if they are evaluated *in situ* (thus, one cannot evaluate a whole by looking to its parts in isolation; this is one reason why we consider ourselves holists); and the value of the whole does equal the sum of the values of its parts.

Or, at least, we hold this view when it comes to the value of states of affairs. We are not opposed to evaluating other kinds of entities, but we shall confine most of our discussion to states. Why? Because, ultimately, our interest is in practical reasons for and against actions, and actions produce states of affairs (which include the acts themselves). When, for example, a sculptor produces a sculpture, or a composer writes a score, we do not deny that the sculpture and the score might be valuable objects. But the sculptor and the composer, in producing these objects, also do a good deal more. The sculptor produces sounds and stone chippings; the composer uses ink and paper. And their actions have many other impacts. They both produce states of affairs. Part of the state produced by the

[8] Ibid., 36.

[9] These parts do not include their combination as an additional part. To do so would not only be misleading, but probably lead to regress: would we include the combination of [the basic parts and their combination] as itself a further part?

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

sculptor is the existence of the sculpture; part of the state produced by the composer is the existence of the score. These two existences are themselves states that are part of the larger states produced by their authors.

When it comes to practical reasons, then, the evaluative comparison of interest is the comparison between states of affairs (we do not draw a distinction here between states of affairs and events)—these are the relevant wholes. And we shall take it that, for purposes of evaluation, their parts are also states of affairs. Consider, for instance, the state of affairs, S, of A committing a crime and subsequently being tried and imprisoned. S has as some of its parts: A committing the crime; A being tried; A being imprisoned; and A believing that he committed the crime, believing that he has been imprisoned because of this, and feeling remorse. There could have been a state of affairs, T, like S in that A is tried, imprisoned, has the same beliefs, and feels the same remorse, but in which he did not actually commit the crime. T and S have some similar parts, then, but, on our account of matters, *pace* Moore, the values of these parts may be influenced by their surroundings.

For example, take the imprisonment of A. On Moore's account, this imprisonment has the same (negative) value in both S and T. But in S the addition of the negatively valued punishment increases the value on the whole (as in the numerical illustration above), whereas in T it does not. On our account, by contrast, the imprisonment in S has positive value—assuming it is a fair punishment for the crime committed. But in T the imprisonment is certainly not just, and hence it is of negative value. Or take A's feelings of remorse. Moore, applying his isolation test, would presumably assign them the same value in both S and T. We do not. In S we take them to be of positive value, but in T this is reversed.

What is the extent of the holistic dependence of one part upon another vis-à-vis their respective values?[10] We believe it is sometimes significant; sometimes nonexistent. For example, in the case of a crime, it might be that its negative value is not dependent upon whether it is later punished. So the disvalue of the crime may be independent of its later punishment. But the value of an episode of incarceration under the auspices of a legal system certainly depends upon whether it is a case of justified punishment for an earlier crime. The value of the later state (the incarceration) depends upon the existence of the earlier one (the crime), but not vice versa. Can the existence of a later state ever impact the value of an earlier one? We certainly do not want to rule this out. It might be, for example, that the disvalue of a tough early episode in a life might depend upon what happens later in that life.

To return to our theme, every action results in a state of affairs, which may be broken down into parts, each of which is another state

[10] Thanks to Joe Long and the other members of our Florida State University graduate seminar in spring 2008 for raising this question.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

of affairs. Care must be taken when evaluating these parts, since one and the same part can be picked out by differing descriptions[11]—and these descriptions can make reference (either implicitly or explicitly) to other parts of the same whole: A's imprisonment in the case above can be redescribed as just punishment in state S but not in state T. Also, when summing the values of parts, the catalogue of parts must be mutually exclusive and exhaustive lest the whole be overvalued or undervalued. (There will, in general, be more than one way to carve up the whole into such a catalogue.)

The descriptions true of a state of affairs may, as it were, pull in different directions vis-à-vis its value. Consider again A's just imprisonment. On some views, the pain of the imprisonment is a bad to be weighed against the good of the justice. But there are not two different states of affairs here, one bad and one good. There is one state of affairs with two different descriptions (the intentional infliction of pain, the imposition of justice), and in determining whether it is better to imprison A rather than not, the good must be weighed against the bad—all relevant descriptions of the state of affairs must be taken into account. We are supposing that the good outweighs the bad here, so we have more reason to impose the punishment than not. (The prisoner, on our view, may have more reason to escape rather than not, but we will discuss this point later, when we consider the notion of a benefit.)

So much for the bearers of value. But what kind of value are we speaking of? Certainly not instrumental value (value as means). Yet Moore restricts value to only two varieties: intrinsic and instrumental. And we have rejected Moore's isolation test, and with that rejection goes, of course, rejection of his account of intrinsic value. Intrinsic value is, for Moore, identical to all of the following: value in itself, value as an end, and (as far as we can tell) value for its own sake.[12] Also, as we have seen, intrinsic value is, on Moore's account, held essentially. So we need some different account of noninstrumental value.

Is there, for example, an alternative account of intrinsic value that does not appeal to the isolation test? Part of the intuitive appeal of the isolation test rests on the thought that intrinsic value should not depend on extrinsic features. For example, suppose that a state in which a beautiful object exists is increased in value by the arrival of an appreciative observer. Is the *intrinsic* value of the former state increased by this arrival? It seems not. To take another example, a red helleborine may be made more valuable by its rarity, but its rarity (since this is dependent upon the number of *others*) is not relevant to its *intrinsic* value. On this approach, the intrinsic value of an entity is dependent upon that entity's intrinsic features—

[11] Cf., e.g., Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 59, on the redescription of actions.

[12] Moore, *Principia Ethica,* 21, 24, 87, respectively.

but work would need to be done to spell out both the nature of an intrinsic feature and the nature of the dependence.[13]

We do not propose to do this work here. Rather, the important category for our present purposes is that of value as an end. And, *pace* Moore, two states of affairs might differ in their value as ends even though they have the same intrinsic value (because they are intrinsically similar). For example, suppose that in one state of affairs red helleborines are rare, but in another they are not. Each of these states of affairs has, as a part, the existence of some red helleborine. The intrinsic values of these two parts are identical, but their values as ends differ: the value as an end of the existence of the plant is greater in the state of affairs in which it is rare.

Or so it seems to us. Intuitions here vary, but we take it that the destruction of a rare plant is, ceteris paribus, worse than the destruction of a more common one—and worse in some more basic sense. It is this basic sense that we are trying to get at with the conception of value as an end. Perhaps it might help to see some further distinctions, particularly in light of the fact that there is a distinction between value as an end and fundamental value, as we shall now see.

Value as an end is, of course, to be contrasted with value as a means. But in addition to this distinction, there is also a distinction between fundamental and derivative value. Instrumental value is clearly derivative: the value of a means derives from the value of the end at which it aims. But there also seem to be cases of derivative value as an end, such as certain cases of constitutive value. For example, innocent enjoyment is valuable as an end, and watching cricket constitutes such a form of innocent enjoyment for David. It is not that watching cricket is a means to enjoyment in the way that driving to the cricket match is such a means, or in the way that surgery is a means to health: David's watching cricket *is* a form of innocent enjoyment. Watching cricket in this case is valuable as an end. It is not fundamentally valuable, however: its value, on this account, derives from the fact that it is a form of innocent enjoyment. One way to distinguish the fundamental from the derivative is to consider the question, "Why is that valuable?" In the case of David's watching cricket, one answer is: "Because he enjoys it." But if we ask, "And why is enjoyment valuable?" we seem to hit bedrock. Although enjoyment may not be valuable (if it is sadistic, for example), when it is valuable, there seems to be no further feature that confers its value. Thus, in this sense, watching cricket is derivatively valuable; it is the enjoyment that is fundamentally valuable, although both the watching and the enjoyment are valuable as ends.[14]

[13] See Jonathan Dancy, *Ethics without Principles* (Oxford: Oxford University Press, 2004), chap. 10.

[14] Setting aside the isolation test, this distinction between fundamental and derivative value, if adopted by a Moorean, might help in her defense against certain complaints. For example, T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

The objects of value, when it comes to actions, at least, are states of affairs,[15] where the value under consideration is value as an end. There will, in general, be more than one way to break a state into a mutually exclusive and exhaustive set of component states, but the sum of the values of the members of any of these sets will equal the sum of the values of the members of any of the others. And this sum is the value of the original state of affairs. This account of matters is not holistic in Moore's sense, then, because the value of a whole is equal to the sum of the values of its parts; but it is holistic in another sense: the value of a part is dependent upon the whole in which it appears.

An analogy might help in drawing the distinction between our position and Moore's. One might think of a state of affairs as in a some ways like a work of art—say, Michelangelo's *David.* (Moore discusses the value of a human arm,[16] and our discussion here will draw on this to some extent.) On both our account and Moore's, the value of *David* is significant. For Moore, however, it might well be that its entire value is its value as a whole. On this account of matters, any part taken in isolation (*David*'s nose, say) has zero value. We agree that any part taken in isolation has zero value—but we contend that this way of valuing the parts is simply irrelevant to the evaluation of the statue. Rather, the relevant value of *David*'s nose is the value of its contribution to the statue. Perhaps *David*'s hand contributes more than his nose, in which case the value of the former is more than the value of the latter. Overall, it might be said by an art critic, it's a superb statue: even though the nose is nothing special, the hands are a perfect "fit." The statue can be (conceptually) carved in multiply different and arbitrary ways. On our account, the evaluation of the

Press, 1998), 88–89, claims that because Moore adopts a purely "teleological" account of value, he (Moore) cannot give a satisfactory account of his own claim that friendship is a good. Scanlon implies that Moore's teleological account of value commits him to the view that "the primary reason to be loyal to one's friends is . . . that this is necessary in order for the friendship to continue to exist" (89). We take Scanlon to be attributing to Moore the following line of reasoning: loyalty is not fundamentally valuable; hence, it must be instrumentally valuable—valuable as a means to continuing friendships. But a modified Moorean position has available the following possibility: loyalty is not a means to continued friendship; rather, it is partly constitutive of friendship. On this account, loyalty is both valuable as an end and derivatively valuable—its value derives from the value of the friendship. We do not think that this will save the Moorean account of friendship, but we do not share Scanlon's account of why it should be rejected. Scanlon rejects a teleological account of value; we accept it (at least as far as acts are concerned). We reject the Moorean account of friendship not because of objections to its account of value, but because we take it that a Moorean account of reasons would entail that the strength of your reasons to act always varies only with the value of your so acting. And we deny this; we claim that positional facts can also play an ineliminable value-independent role in certain reasons, such as reasons of friendship. See, e.g., David McNaughton and Piers Rawling, "Deontology," in David Copp, ed., *Oxford Handbook of Ethical Theory* (Oxford: Oxford University Press, 2006), 424–58.

[15] Scanlon, *What We Owe to Each Other,* chap. 2, objects to this, but we answer him in David McNaughton and Piers Rawling, "Holism about Value," in M. Lance, M. Potrc, and V. Strahovnik, eds., *Challenging Moral Particularism* (London: Routledge, 2008), 116–32.

[16] Moore, *Principia Ethica,* 34–35.

carvings in isolation is irrelevant to the evaluation of the statue. But when evaluated *in situ*, the sum of their values is the value of the statue.

We turn now to the question of whether one account has advantages over the other.

C. Comparison of the two holisms

Our central complaint against the Moorean approach is its reliance upon the isolation test. When, for instance, we look to the isolation test to determine the intrinsic value of a punishment, Moore asks us to consider its pain in isolation. But in isolation from the crime, the "punishment" would be the infliction of, say, incarceration, for no reason—it would not be punishment.[17] Presumably, Moore assumes that the prisoner will have all the same beliefs in the isolated case as in the real one, such as the belief that she has been imprisoned after a trial; otherwise, the subjectively experienced pain of the incarceration would be significantly different. However, these beliefs are, of course, all false in the isolated case. Yet in other places,[18] Moore takes true belief itself to be of value. Hence, the false beliefs themselves detract from the value of the incarceration in the isolated case.

The punishment case is but one instance of a general problem for the isolation test, namely, that it is often true that part of what makes a state the state it is, is its relation to other states. Many states are like memories: a memory is not a memory unless it recalls a previous occurrence. Or, to take a forward-looking case, one state cannot be the cause of another unless the latter comes to pass. The value of the striking of a match might depend on whether it causes an explosion, but whether it is the cause of the explosion depends on the existence of the explosion.

It has been suggested that we might be being uncharitable to Moore here.[19] Perhaps what he intends to be evaluated in isolation in (say) the case of punishment is not the incarceration by itself, but the punishment together with all that is necessary in order for it to be (genuine) punishment, which includes the preceding crime. But we doubt this is Moore's intent since, if so, it would seem that we must evaluate the crime and the punishment without considering them in isolation from one another after all. But now we no longer have an isolation test.

Not all states of affairs have the degree of interdependence that we have just been discussing. But even in the absence of such interdependence, isolated evaluation seldom if ever makes much sense. How can one evaluate (say) the contemplation of a beautiful rose in a universe where it and its observer are the sole existents? To begin with, is the life

[17] Thanks to Anthony Price for this point.
[18] See, e.g., Moore, *Principia Ethica*, 197.
[19] The suggestion comes from Brad Hooker.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

of the observer short or long? Perhaps a longer life is better, but what about a long life spent only in the contemplation of a rose?

Given that the isolation test has (at the least) these problems, is it helpful? We doubt it. In order to calculate the value of a combination as a whole, it seems one must consider the whole in isolation, and its parts in isolation, and then calculate the value of the whole minus the sum of the values of its parts. But why bother determining the values of the parts in isolation? Doing so does not help in working out the value of the whole "on the whole," since this is the sum of the values of the parts *plus* the value of the whole "as a whole"—and calculating the latter requires, as we have just seen, prior knowledge of the value of the whole "on the whole." Moore's evaluative procedure is, then, at best unhelpful.

Is ours any more helpful? As on Moore's account, on ours the value of a state of affairs cannot be evaluated by considering the parts in isolation. But at least we do not have an extra component in the form of "value as a whole." Since, of course, we see our view as accurate, we see no alternative to evaluating one part of a state of affairs in light of its relationship to all the others, so that any holistic interactions present are not ignored. But this may not be as difficult as it might appear. For example, the state of affairs of a sentient being experiencing pain is typically bad and detracts from the value of any state of which it is a component (hence, a retributivist view of punishment presents an odd case).

In practice, of course, the evaluation of a state of affairs is very tricky—on our account or on Moore's. But any view that concerns itself with the evaluation of actions, which is our ultimate interest, is going to face tremendous epistemic difficulties given how far-reaching our acts can be.[20]

So much for our discussion of value. We move now to our other central topic: benefits.

D. Benefits and harms

In addition to asking about the value of a state of affairs that might be produced, one can also ask whom its production would benefit. Indeed, on some accounts (ours among them), this latter question is prior to the former.

This is so even on the simple form of hedonism that we considered at the outset, according to which the value of a world is based solely on the amount of pleasure therein. Assuming that pleasure is a benefit to the being that experiences it, in determining the value of a state of affairs we would first work out how much it would benefit each sentient being (that is, on this account, how much pleasure it would bring each of them), and then sum these benefits.

[20] For more on our view of Moore's theory of organic unities, see McNaughton and Rawling, "Holism about Value."

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

However, as we noted, one of the things that makes this view so "simple" is that the distribution of the benefits within a state of affairs is irrelevant to its value. On any view that takes distribution to impact value, the prior appeal to benefits is, obviously, even more crucial. In comparing two states, you need to know not only what the sum of the benefits is in each, but also what the distribution of those benefits is. A standard criticism of utilitarianism is that it fails to take account of benefit distribution. According to utilitarianism's opponents, a state in which the sum of the benefits is less may well be a better state because these benefits are more evenly distributed.

There is conceptual space for distributional concerns only if there are benefits (and harms) that can go to individuals. That is, in addition to the value of a state, there must also be a relational notion of benefit in play if we are to make sense of distributional concerns. (Benefit is, as it were, a relation between a state and an individual.) Furthermore, we must have the information about benefits prior to assessing the value of a state if we want to take account of the distribution of those benefits in assessing that value.

What is the relationship between benefits and value? Even if the value of a state is not a straightforward sum of the benefits that go to individuals in that state, perhaps at least the receipt of a benefit by an individual is always good. We think not. Here are a couple of examples.

One of the most contentious issues that arises in running an academic department is the distribution of merit pay raises. Some members of the department will receive benefits that others do not. If the distribution of these benefits is fair and just, then it may well be that the provision of each benefit contributes to the good, and perhaps does so doubly: first, there is the good of the deserved pleasure to the recipients; second, there is the good of a just act being performed. But what if the distribution is unjust? The pleasure of at least some of the recipients is undeserved. Do we want to say that this undeserved pleasure is good? We think not. But do these undeserving recipients not receive a benefit? We think they do. So we have a case in which someone receives a benefit, but this does not increase the good.

Or take the case of punishment. Recall Moore's treatment: both the wickedness of the crime and the pain of the punishment are bad, but the value "as a whole" of the two taken together is positive; thus, the addition of the (bad of the) punishment actually increases the good. We noted that Moore preserves the intuition that the punishment must be *bad* in order for it to be *punishment*. But how do we preserve this intuition? In our treatment, we have no "value as a whole" to which to appeal; rather, the addition of the just punishment straightforwardly increases the good. However, just as we admit the relational notion of a benefit, so we admit its counterpart: a relational notion of harm, where harms can be inflicted on particular individuals without decreasing the good. Just imprisonment

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

is such a case: the prisoner is harmed, but the good is increased because the harm is deserved. (Or, at least, our position allows room for this claim—we shall not here discuss whether it is actually so.) This might help explain some people's conflicting intuitions about just imprisonment. It appears to be good because it is the just infliction of pain, but bad because it is the infliction of pain. Some might see this apparent conflict resolved by weighing the good of the justice versus the bad of the pain. But we now have an alternative: just imprisonment is better than unpunished crime, but it does harm the prisoner.

Rather than the term 'benefit' some might prefer the explicitly relational 'good for'. We are not wedded to any particular terminology, but it is important to keep this latter usage of 'good' distinct from the monadic use, as when we say, 'This is a good state of affairs'. As we have just noted, a state of affairs might be good for a person and yet bad; or it might be bad for the person and yet good.

We suspect that Moore (or, at least, the Moore of *Principia Ethica*) does not countenance the possibility of relational benefits: at least, that is our conclusion from our attempt at interpreting his famous argument to the effect that egoism is incoherent. (He does later recant this claim of incoherence,[21] but analysis of the original argument is instructive.) In order to proceed with the discussion of egoism, it will be helpful to say something first about practical reasons.

Donald Davidson is justly famous for his account of practical reasons, according to which a reason is a belief-desire pair: the desire, say, for a cup of coffee, combined with the belief that one is to be found down the hall, constitutes your reason for setting off down the hall.[22] Michael Smith refers to these states as "motivating reasons."[23] There is, however, an alternative notion of a reason according to which a reason is not a combination of psychological states but, rather, a fact—a consideration—that counts in favor of (or against) something.[24]

Theoretical reasons count in favor of believing some proposition; practical reasons count in favor of acting in a certain way. Facts can be normative or nonnormative. Normative facts are, roughly, ones that can only be expressed using normative terms, such as 'reason', 'ought', 'should', etc. Reasons are facts that have, to borrow Derek Parfit's phrase, normative significance; but they need not themselves be (and perhaps usually are not) normative.[25] That Eve is in pain is a nonnormative fact, and it is a reason for you to do something—namely, give her a pain-killer. There

[21] G. E. Moore, *Ethics* (London: Williams and Norgate, 1912), 99–100.

[22] Donald Davidson, "Actions, Reasons, and Causes," in his *Essays on Actions and Events,* 3–19.

[23] Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), 96.

[24] See, for example, Derek Parfit, "Reasons and Motivation," *Proceedings of the Aristotelian Society,* supplementary volume 71 (1997): 99–130.

[25] Ibid., 124.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

are, then, two facts here: (1) Eve is in pain; and (2) this first fact is a reason to give her a pain-killer. Fact (1) is not normative, but fact (2) is: it is a normative fact that Eve's being in pain is a reason to give her a pain-killer.

Practical reasons, then, as we shall think of them, are not psychological states. Furthermore, although it is not inconsistent with our usage to maintain that someone's being in a certain psychological state (as opposed to the state itself), such as their having a desire, is a reason, we are doubtful that many such facts are reasons. For example, we tend to agree with those who deny that an agent's having a desire to X is itself a reason for that agent to X—although the fact that she has the desire may be *evidence* that she has such a reason.[26] For instance, the fact that someone wants to X can be evidence that she would enjoy X-ing; and the fact that she would enjoy X-ing can, in an appropriate context, be a reason for her to X.

According to some forms of consequentialism, if some potential act of ours will contribute to making the world better, then we have some reason to perform it. We agree. But we disagree with the consequentialist if her view is that the strength of one's reasons to perform an act varies with, and only with, the amount of good the act will produce (where the act itself is included in its own product). Let's call this view *simple consequentialism.* On our view, by contrast, if an act of yours will increase value, then you have reason to do it, and we all have some reason, say, to encourage you to do it. But if the act will also bring a benefit to you, it might well be that you have some *extra* reason to perform the act beyond its contribution to the impersonal good.

There can also be cases where your act will make the world worse, so we have reason to prevent you from performing it, even though it will benefit you. Thus, for example, our position allows for the following thought: the prisoner's escape will make the world worse, but it is of benefit to him, and he has some reason to do it—even though we have reason to prevent him and no reason to aid or abet him. (We do not necessarily endorse this view: our framework is flexible enough to allow also for the Socratic thought that the prisoner's escape actually harms him.)

Moore does not discuss practical reasons per se in *Principia Ethica*—or, at least, not in detail. But his position seems to be consistent with the simple consequentialist view above, according to which the strength of one's reason to perform an act varies with, and only with, the amount of good the act will produce. This is so particularly in light of his argument against egoism.

We are in agreement, in part, with Moore's view of goodness as laid out in *Principia Ethica.* Among other points of agreement, we see goodness as

[26] See, for example, Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000), chap. 2; Scanlon, *What We Owe to Each Other,* 43–49; and Parfit, "Reasons and Motivation," 128.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

nonrelational: goodness and badness are monadic properties of states of affairs. And we agree that goodness is impartial—or at least, as we would prefer to put it, impersonal or agent-neutral. But we disagree with Moore's position if he denies the existence of benefits and harms that are relational in the sense that a state of affairs can be of benefit or harm to something.

Why might one think he denies the existence of benefits and harms (or, at least, does so in *Principia Ethica*)? Consider the following key part of his argument against egoism:

> The only reason I can have for aiming at "my own good," is that it is *good absolutely* that what I so call should belong to me—*good absolutely* that I should *have* something, which, if I have it, others cannot have. But if it is *good absolutely* that I should have it, then everyone else has as much reason for aiming at *my* having it, as I have myself. If, therefore, it is true of *any* single man's "interest" or "happiness" that it ought to be his sole ultimate end, this can only mean that *that* man's "interest" or "happiness" is *the sole good, the* Universal Good, and the only thing that anybody ought to aim at. What Egoism holds, therefore, is that *each* man's happiness is the sole good—that a number of different things are *each* of them the only good thing there is—an absolute contradiction! No more complete and thorough refutation of any theory could be desired.[27]

One ground for rejecting this argument is that Moore does not acknowledge the possibility of benefits that can go to individuals. Thus, for example, he rules out the possibility of asking the question: Is it good that this individual receive this benefit? He does not acknowledge the possibility that although goodness is a nonrelational property, nevertheless there might also be a relational property of benefiting that holds between states of affairs and individuals.

By contrast, we acknowledge the existence of such relationships, and we see the egoist as arguing that we each only have reason to pursue our own benefit, regardless of whether, and to what degree, a state of affairs in which we benefit is good simpliciter. Of course, someone with Moorean sympathies might respond by claiming that only goodness is reason-providing, and we all have equal reason to promote the good and most reason to pursue the best. The egoist must reject at least part of this. Suppose she does not. Then she must argue that only states of affairs in which *she* receives some benefit are good, and that the state in which *she* benefits most is best, so that she only has reason to pursue states in which she benefits, and has most reason to pursue the state in which she benefits most. Her opponent, however, can counter by noting that, since we all have most reason to pursue the best state,

[27] Moore, *Principia Ethica,* 99; emphasis in the original.

we all have most reason to pursue the state in which she benefits most. And now symmetry requires that, for each person, we all have most reason, on every occasion, to pursue the state from which that person benefits most. But this is impossible given the prevalence of occasions on which your benefit is my loss.

In fact, if, with the Moorean sympathizer, we accept that (1) goodness is nonrelational and impersonal, and (2) only goodness provides reasons, then we could establish not merely the falsity of egoism, but also the falsity of the claim that we ever have *any* reason at all to pursue our own benefit beyond its contribution to the general good. And we are back with simple consequentialism.

Moore's argument against egoism in *Principia Ethica,* then, establishes too much. While we disagree with egoism, we do not think it is incoherent in the way that Moore suggests. And we are not simple consequentialists, as we have dubbed them: we think that we do, at least on occasion, have reason to pursue our own benefit beyond its contribution to the general good. (We do not, of course, advocate selfishness, but to be selfish is to go beyond what reason endorses in the pursuit of self-interest.) We depart from consequentialism in many other respects, but that discussion would take us too far afield. What is relevant here is some discussion of consequentialism and distributional concerns.

As we have noted, in order to raise the question of distribution it must be acknowledged that there are benefits and harms that can be distributed differentially. Thus, what we might call the pure Moorean framework, which eschews benefits and harms, does not even allow the issue to be raised. Such a view we find implausible. But what about a consequentialist who claims that, say, a just distribution of benefits and harms makes the world better?[28] Such a consequentialist position is, we think, potentially unstable. Of course, there are many sophisticated varieties of consequentialism, so all we can do here is give the basic idea of what we see as the problem that confronts consequentialism when it comes to the distribution of harms and benefits.

The problem is not about the distribution per se, however. It concerns, rather, whether consequentialism can accommodate the very existence of benefits and harms. First let us consider the question of whether a consequentialist can think that we ever have reasons to pursue our own benefit beyond its contribution to the general good. We think not. Either there is the possibility of such reason's outweighing the reason to maximize the good or, if not, then such reasons would be, as it were, redundant: they would not truly be reasons to act, since they would always be trumped by the reason to maximize the good. But if such a consequen-

[28] Fred Feldman, for example, expounds this sort of position in his "Justice, Desert, and the Repugnant Conclusion," in his *Utilitarianism, Hedonism, and Desert* (Cambridge: Cambridge University Press, 1997), 195–214. We discuss this essay below.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

tialist also accepts that there are benefits and harms that can go to individuals, she is under pressure from two directions.

On the one hand, there is pressure to acknowledge special reasons to pursue one's own benefit: how can something be of benefit *to me* without the possibility, at least, of my having a special reason to pursue it for myself? That is (in the case where my receiving the benefit would also increase the general good), how can something be of benefit to me without my having additional reason to pursue it beyond its contribution to the general good? On the other hand, there is pressure from the opposite direction—what we have called the "pure" Moorean position—to deny that benefits can go to individuals.

It may be open to the consequentialist to distinguish between moral and prudential reasons,[29] and to argue that although we have most *moral* reason to maximize the good, we have most *prudential* reason to maximize our own benefit, where moral and prudential reasons are incommensurable (otherwise, we would be back with the problem that we might have most reason *not* to maximize the good). On such an account, however, we would seem to lose the requirement, central to simple consequentialism, that we must maximize the good. We could, within the bounds of reasons, simply choose to ignore morality—consequentialism would draw close to egoism.

E. Summary and conclusions

Before turning to a pair of applications, let us summarize the situation thus far.

(1) We reject Moorean holism in part because of its reliance upon the isolation test. We favor instead an aggregationist holism according to which the value of a state of affairs equals the sum of the values of its constituent states. But these constituents must be evaluated *in situ.*

(2) Unlike Moore, we claim that there are benefits and harms that can go to individuals. To deny this claim is to deny, among other things, the possibility of even raising distributional concerns, or of stating egoism in a coherent fashion; and denying these possibilities is implausible. But once benefits and harms are in the picture, there is pressure to acknowledge that we can have extra reason to pursue or avoid them beyond considerations of the effect of such pursuit or avoidance on the general good.

III. Two Applications

A. Pareto-optimality

A Pareto-optimal state (named for Vilfredo Pareto) is one in which no one can be made better off without someone else being made worse off,

[29] Alastair Norcross suggested this possibility to us.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

where 'better off' and 'worse off' are typically defined in terms of allocation of resources—someone is worse off if and only if he has fewer resources. Of course, one can debate both whether this is the only viable notion of 'worse off', and what comprises 'resources'. For example, one might put matters in terms of preference-satisfaction:[30]

> A state of affairs is (strongly) Pareto-optimal if and only if there is no other accessible state in which someone's preferences are better satisfied and no one's preferences are worse satisfied.

For example, suppose we are in S_1, and it is in the second rank in everyone's preferences. But we could move to S_2, which is in the second rank in everyone's preferences except yours, in which it is in the first rank. And all other accessible states are in the third rank in at least one person's ranking. Then S_2 is Pareto-optimal, but none of the other states are.

At least in some popular accounts, preference-satisfaction is simply assumed to be the measure of how well off one is. Here, for example, is part of the *Wikipedia* entry on the topic:[31]

> Given a set of alternative allocations of, say, goods or income for a set of individuals, a movement from one allocation to another that can make at least one individual better off without making any other individual worse off is called a *Pareto improvement.* An allocation is *Pareto efficient* or *Pareto optimal* when no further Pareto improvements can be made. This is often called a *strong Pareto optimum (SPO).*
>
> A *weak Pareto optimum (WPO)* satisfies a less stringent requirement, in which a new allocation is only considered to be a Pareto improvement if it is strictly preferred by *all* individuals (i.e., *all* must gain with the new allocation). The set of SPO solutions is a subset of the set of WPO solutions, because an SPO satisfies the stronger requirement that there is no allocation that is strictly preferred by one individual and weakly preferred by the rest (i.e., no individual loses out, and at least one individual gains).

Note the move from an allocation's making someone "better off" to that allocation's being preferred by that person. Do we always prefer the allocation that makes us better off? One could simply define 'makes S better off' as 'is preferred by S'. But this is, at the least, a contentious definition, and reflects an extreme subjectivism concerning the issue of 'doing well'. We find it implausible that, say, someone whose self-

[30] See, for example, Derek Parfit, *Reasons and Persons,* rev. ed. (Oxford: Clarendon Press, 1987), 493–501.

[31] Its popularity explains our citing *Wikipedia* in a scholarly essay. See http://en.wikipedia.org/wiki/Pareto_efficiency (accessed July 2007).

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

destructive preferences are satisfied is doing well. There are, no doubt, exceptions to this, but, in general, such a person is doing badly. Let us suppose, then, that we have some more objective notion of what it is to be better off, and define Pareto-optimality as follows:

> A state of affairs is Pareto-optimal if and only if there is no other accessible state in which someone is better off and no one is worse off.

On most accounts, Pareto-optimality is not sufficient for a state to be best. Suppose our current state is one in which a small number of people are doing very well, but the vast majority have undeservedly awful lives. Suppose further that the lives of the members of this majority could be easily improved, but that any such improvements would result in some minor costs to members of the fortunate minority. In fact, no one's life can be improved without a cost to someone else. Then our current state is Pareto-optimal, but it is by no means the best of the accessible states.

However, it may be that a state must be Pareto-optimal if it is to be best—although Pareto-optimality is not sufficient for being best, it is necessary. Let us call the following condition "PON" (for "Pareto-optimality is necessary"):

> *PON:* A state is objectively best only if there is no other accessible state in which some individual would be objectively better off while no individual would be objectively worse off.

Is PON true? First, let's see how a Moorean holist might respond. She claims that the value of the whole need not equal the sum of the values of its parts. So even if she agrees that the parts here are individual lives, and the value of a life is measured by how well it is going, nevertheless there might be two possible states in the better of which the sum of the values of the lives is lower. Consider again the case of just imprisonment, and compare the state of affairs in which the prisoner escapes with the state in which he does not. It could be that in these two states everybody else's lives go equally well, but the prisoner's life goes better in the former state. According to PON, the latter state (in which the prisoner does not escape) could not be the best. But the Moorean can disagree: the addition of the negative "value as a whole" of the evasion of just imprisonment in the latter state might make it worse than the former.

On our view, as we have emphasized, there are benefits and harms that can go to individuals. And for one individual to be better off than another is for the former to have more benefits. But we deny that the world need always be made better if the net benefits to individuals are increased. Thus, we deny PON: there could be two states A and B such that, in moving from A to B, one person benefits and no one suffers any harm,

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

and yet A is the best state. It might be, for example, that in state B the person who benefits does so undeservedly.

Of course, the defender of PON might respond to such a case by claiming that everyone is harmed by someone's benefiting undeservedly. We have some sympathy with this claim. And, indeed, perhaps our objections to PON, per se, are not that strong. What we do object strongly to, however, is the following thought, which might underlie a commitment to PON:

> *WEL:* The measure of the value of a world is the net sum of the benefits minus the harms enjoyed by its inhabitants.

We call this "WEL" since it is a welfarist principle. If WEL is true then so is the following equivalent of PON:

> *PON_2:* A state, S_1, is objectively best only if there is no other accessible state, S_2, such that some individual would benefit if the world moved from S_1 to S_2, while no individual would be harmed by such a move.

Why do we find WEL so much more objectionable than PON? PON is somewhat credible because it concerns only cases in which the population is constant. In applying PON, we compare states in which the populations are identical, and we ask about the relative benefits and harms accruing to the members of that one population in the various states. But WEL opens up the prospect of varying populations, and that introduces Parfit's notorious Repugnant Conclusion.[32]

B. The Repugnant Conclusion

Parfit's discussion of his Repugnant Conclusion and related topics is long and subtle, and there is much that we shall not address. But here are the basics that are relevant for our present purpose. Consider first Parfit's "Impersonal Total Principle":

> If other things are equal, the best outcome is the one in which there would be the greatest quantity of whatever makes life worth living.[33]

This is a version of WEL. And here is a consequence of it:

> *The Repugnant Conclusion.* For any possible population of at least ten billion people, all with a very high quality of life, there must be some

[32] Parfit, *Reasons and Persons,* 387–90.
[33] Ibid., 387.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

> much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.[34]

Suppose each inhabitant of world W_1 enjoys only minimal benefits, but each still has a life worth living. And suppose, by contrast, that each inhabitant of W_2 has great benefits. But let us further suppose that the population of W_1 is vastly greater than that of W_2—so much greater that the total sum of benefits minus harms in W_1 is greater than that in W_2. Hence, by the lights of WEL, W_1 is the better world.

We have discussed above (in Section II.D) the possibility of incorporating desert into one's account of value, so let us look first at an attempt to avoid the Repugnant Conclusion by appealing to desert. The attempt we have in mind is due to Fred Feldman.[35] Feldman endorses a view he calls "justicism," which he gets by modifying utilitarianism in the following way. Rather than evaluating a world by looking solely at the total amount of happiness in it, look also at the happiness of each individual and ask whether that happiness is more or less than the individual deserves. A mismatch between happiness and desert in a world reduces the value of that world. The value of W_1 in the previous paragraph is then much reduced on the assumption that its inhabitants deserve many more benefits than they are receiving. In W_2, by contrast, let us suppose that the inhabitants deserve their benefits. Then, provided the reduction of the value of W_1 is large enough, W_2 proves to be the better world, and the Repugnant Conclusion is avoided.

However, we are not convinced of the truth of justicism. As Feldman himself points out, its advocate must answer the following objection. What if in a third world, W_3, the people are receiving the same benefits as in W_1, but unlike the occupants of W_1, the people in W_3 are all behaving badly, so that they are getting exactly the meager benefits they deserve?[36] This may, depending upon how one counts the disvalue of the bad deeds, make the value of W_3 greater than that of W_2—an implausible result.

How do we avoid the Repugnant Conclusion? The first point to note is that, strictly speaking, it is not as repugnant to us as it is to certain consequentialists, since we are not committed to the view that we have most reason to bring about the best world. Recall that we disagree with the consequentialist if her view is that the strength of one's reason to perform an act varies with, and only with, the amount of good the act will produce. So if the "repugnant" world is the best, so be it. That does not give us the most reason to bring it about.

[34] Ibid., 388.
[35] Feldman, "Justice, Desert, and the Repugnant Conclusion."
[36] Ibid., 212.

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press

However, there does seem to be something wrong with a view of value that entails the conclusion that the repugnant world is best. Our view of value and benefits avoids this conclusion. How? The short answer is that we reject WEL. Of course, we could be accused of rejecting WEL simply in order to avoid the Repugnant Conclusion, with no further rationale. In a sense, perhaps we come close to this. Our view is that benefits and harms that go to individuals are one thing, and value is another, and one way of seeing this is to consider the Repugnant Conclusion. Why is it repugnant? Well, clearly the repugnant world is worse than the first alternative in Parfit's example (the world with at least ten billion people with a high quality of life), even though the sum total of the benefits within the repugnant world is the larger. (We can do no more than appeal to widely shared intuitions here.) This all goes to show that value is not to be measured by the sum total of benefits. We are aggregationists about value, but benefits are not what we aggregate.

IV. Conclusion

We have given a view of value that is both aggregationist and holist, where our holism avoids the difficulties that beset Moore's version of the doctrine—difficulties associated with his isolationism about the values of parts.

We have further argued that, in addition to nonrelational impersonal value, there are also relational benefits and harms that go to individuals. Such benefits and harms enable us to make some sense of egoism—we think it false, but not incoherent. And Parfit's Repugnant Conclusion works as a straightforward example in our favor, rather than as a difficult case to be overcome.

Philosophy, Florida State University

https://doi.org/10.1017/S0265052509090141 Published online by Cambridge University Press