MUTUAL BENEVOLENCE AND THE THEORY OF HAPPINESS*

What do you want to do?" she said.
"Whatever you want to do."
"I want whatever's best for you."
"What's best for me is to please you," I said.
"I want to make you happy, Jack."
"I'm happy when I'm pleasing you."

White Noise, Don Delillo[1]

IMAGINE just two people, each of whom has no desire other than for the satisfaction of the other's desire. The object of the one's desire is the satisfaction of the other's, but the satisfaction of the other's desire would be constituted only by the satisfaction of the one's, and so on in a loop. The structure of this case is important for moral theory, by way of the havoc it wreaks on the notions of benevolence and happiness.[2] The rough structure of the argument I shall offer in support of this thesis may be summarized as follows. It may seem that the loop depends on the absence of any independent, first-order desires. Although the presence of independent desires supplies some content, it is not a general way out of the loop, however, since there is no reason to think that altruistic or benevolent desires will, in general, conveniently refrain from aiming at the satisfaction of the other's benevolent desires. As an example, this loop will be shown to occur under certain assumptions of Joseph Butler's moral psychology. The occurrence of the loop under the Butlerian assumptions shows that the loop can occur in realistic situations, but, moreover, it suggests that, without deep changes in the ordinary view of happiness, both benevolence and happiness may be empty in a disturbing way. The loop problem depends on, and so puts a certain pressure on, a satisfaction conception of happiness, according to which happiness is, in one way or another, the satisfaction of passions or desires. Butler himself may have held an interesting alternative

* Earlier versions of this paper were presented at various conferences, where I benefitted from discussions with John Broome, Nicholas Denyer, James Griffin, Shelley Kagan, and Patricia Werhane. I am grateful to the University of California/Irvine for a summer research fellowship in 1988. I also received helpful comments from Rachel Cohon, Tyler Cowen, Alon Harel, Michael Hudson-Smedley, Greg Kavka, Penny Maddy, Gary Watson, Joshua Weinstein, and Peter Woodruff.

[1] New York: Viking, 1985. Thanks to Andrew Cutrofello for pointing out this text.

[2] The notions of a person's good, happiness, and well-being will be used interchangeably.

0022-362X/90/8704/187–204

conception of happiness which is not subject to the loop and its associated problems.

I. THE PRESENCE OF MORE PEOPLE IS NO WAY OUT

First, an obvious point. In the interest of realism, consider adding people to the two in the original example. The presence of, for example, a third person will obviously be irrelevant if the first two remain as originally described: interested only in each other. The point of introducing the third person could only be that it may provide a way out of that two-person loop. So suppose that at least one of the original two people is concerned for the third person. We should suppose this third person is like the other two, however, in desiring nothing but the satisfaction of some other person's desires. But the only others, the original two, desire only the satisfaction of one of the two others as well: either the satisfaction of the other member of the original pair, or of the new third person. In either case, every individual's desire is involved in a loop. Clearly, the same holds for four people, or any number.[3]

If the number of individuals is infinite, the problem is, in a way, even simpler. What is puzzling about the loop, after all, is that tracing it is an infinite process even with only a finite number of people. If there are infinitely many people, and the desires are set up so that a loop never forms, the process of tracing the desires to their objects is still infinite.[4]

Obviously, then, the emptiness of the original loop can be present in cases involving any number of people.

II. DOES BENEVOLENCE FADE OUT?

The thought of such a desire chain involving a large number of people naturally prompts the observation that we typically care less about the loved ones of our loved ones than we do about our loved ones themselves. I may want your mother's surgery to be successful because of my concern for you, but if that is the ground of my concern for her, it will be, in some sense, weaker than my concern for you. If chain-linked desires of this sort do diminish in strength, then, if the chain is long enough (e.g., your mother's friend's surgery), the desire strength might be thought to fade out. If this happened, it

[3] Consider person *A*'s desire for the satisfaction of some other person's desire, in a group with *n* members. Even if the chain of desires gets to the *n*th person without looping, if *n* is a finite number the *n*th person's desire will have to be for the satisfaction of the desire of one of the others (*A* through *n*), and that will close the loop.

[4] We cannot just ignore this possibility on the grounds that there are never infinitely many people alive. That would be to assume without argument that the desires in question cannot concern future persons, of whom there may or may not be infinitely many.

could be argued that there is no real infinity involved in the original desire. A closer look reveals, however, that this sort of fade-out is irrelevant to the occurrence of a loop.

It is the strength of the desires that might be thought to fade out, but strength is beside the point. To illustrate, it will be useful to start with a case where the desires do not form the sort of loop in question; so suppose that Joe wants only that Sam get whatever he wants, and Sam wants only to get a job. If we say that this shows that Joe wants Sam to get a job, then the question might arise whether Joe wants this less strongly than he wants Sam to get whatever he wants. But the problem concerning us does not require us to state things this way. Rather than ask what further desires Joe has as a result of the chain, we need only ask: What state of affairs would satisfy the original desire of Joe's, that Sam get whatever he wants? The answer in this case is that Sam's getting a job would satisfy it. But now how can the fade-out thesis be stated? Could it be held that "Sam's getting the job" would satisfy Joe's benevolent desire less than "Sam's getting what he wants"? Of course not; either one satisfies it completely.

It is common practice to suppose that our attitudes toward certain events can vary with the description of the event. For example, there may be one event that is both the setting down of a piano one is helping to move, and the crushing of the neighbor's cat. The event may be desired under one description, and abhorred under the other description. The attitudes need not vary so greatly with the different descriptions; one may care greatly about an event under one description, but hardly at all under another. So consider an event that is both the satisfaction of Sam's desire, and the getting of a job by Sam. The fact that Joe may care less about the event under the latter description than under the former does not impugn the strength of Joe's benevolent desire for that event. If we change Sam's desire for a job to a desire for the satisfaction of the desire of some third person, say, his mother, this only lengthens the chain that determines what event would satisfy Joe's desire. The fact that a new description comes to apply to the event (such as, 'the satisfaction of Sam's mother's desire'), a description under which Joe cares little about it, does not diminish Joe's concern about this event, since, whoever else it may satisfy, it satisfies his friend Sam.[5]

[5] Joe will typically care less about the event that would satisfy Sam where it is described in any of the ways Sam cares about it. And if that second description itself involves the satisfaction of a third person (and Joe had at least some concern for it under this description), he will care even less about it under the description that matters to the third person, and so on. (Of course, Joe may coincidentally have some separate strong concern for the event under one of these further descriptions, e.g., for the welfare of some person whom these others happen to care about, too.)

If Joe's concern does not fade out in the relevant sense where Sam's only desire is for his mother's satisfaction, then, for exactly the same reason, it will not fade out where Sam's only desire is for Joe's getting what he wants, a desire which gives rise to a loop.

III. INDEPENDENT DESIRES

Suppose there are not only the desires about each other, but also other "independent" desires whose object is not the satisfaction of anyone's desires. Consider the case of just two people again. Each still has the desire for the satisfaction of the other's desire (call this the benevolent desire), but now each also has another desire as well. Suppose *A* desires to eat an apple and *B* desires to eat a banana, while each still wants the satisfaction of the other's desire. If the benevolent desires in question are just for the satisfaction of *any* desire of the other, then the puzzle does not arise. For example, *A*'s benevolent desire for the satisfaction of any desire of *B*'s is satisfied by the satisfaction of *B*'s banana desire. It does not require the satisfaction of *B*'s benevolent desire, and so there is no loop.[6] If *A* gets an apple and *B* gets a banana, then all the desires in the example are satisfied, including the benevolent ones.

But here the benevolent desires are conveniently undemanding. The presence of independent desires is not a general way out of the loop at all. Consider, for example, a benevolent desire of *A*'s that aims solely and specifically at the satisfaction of *B*'s benevolent desire, and vice versa. Clearly the mere presence of the other, independent desires does not prevent the loop. Admittedly, the example where *A*'s benevolence aims solely and specifically at the satisfaction of *B*'s benevolence is contrived, but then so is any example in which the object of *A*'s benevolent desire is held specifically to *exclude* the benevolent desires of *B*—those desires which can result in a loop. The introduction of independent desires can only avoid the loop in contrived examples of this latter sort, since as long as the satisfaction of another's benevolence is part of the object of everyone's benevolence the interminable satisfaction loop will remain.

Here is the case that should concern us. Suppose *A* has a benevolent desire that aims at some level of satisfaction of all of *B*'s desires, and *B* has the same with regard to *A*. This is not just an example

[6] John Rawls suggests this sort of approach, but it undermines his own use of the loop case, which is to defend mutual disinterestedness among the parties in the original position. Since the parties will have some independent desires whether or not they have benevolent ones, the fact that mutual "*perfect* altruism" (nothing but "dependent" desires) is empty is no grounds for assuming the absence of dependent (including benevolent) desires altogether. The criticism of this solution below would seem better to support that Rawlsian move. *A Theory of Justice* (Cambridge: Harvard, 1971), p. 189.

concocted to yield a loop; it is, in effect,[7] Butler's view of benevolence and happiness. He never discussed the puzzle to which this appears to give rise, namely, that the satisfaction of *A*'s benevolent desire would require the satisfaction of *B*'s, which in turn requires the satisfaction of *A*'s, and so on. But if Butler's account is correct in certain respects, then there are real-world loops, and the emptiness inherent in them is more than a curiosity (but more on that below).

IV. BUTLER ON BENEVOLENCE AND WELL-BEING

On Butler's view, the psyche is populated by multiple passions. A passion is satisfied when its object is obtained.[8] The object of the passion of hunger, for example, is food. One of the passions that all people have by nature is benevolence, which has as its object the happiness of others. Another is self-love, which has as its object one's own happiness.[9] Happiness does not require the full satisfaction of

[7] Substituting passions for desires, as explained below, yields Butler's apparent view. Note, however, the interpretive questions raised in section VII.

[8] Joseph Butler, "Fifteen Sermons Preached at Rolls Chapel," (excerpts) in *British Moralists,* D. D. Raphael, ed. (New York: Oxford, 1969). The term 'passion' in Butler is more complex than the notion of desire as commonly understood today. For example, we all have the passion of hunger as a part of our psyche, yet we are only hungry intermittently. Roughly, the passion for food is a category of potential desire for food. As such, it seems wrong to say that passions actually have objects even when they involve no actual desire, but a few brief distinctions will show how Butlerian passions do have objects.

The relation of passion to desire is similar to the relation between having a concept and having a thought or belief. The concept is a category of potential thought or belief, and endures between episodes of these. Satisfaction of a passion is really satisfaction of a desire which is the actualization of a certain passion. We may speak of Butlerian passions as being in one of two states: *latent* or *actual.* A passion is actual rather than latent by virtue of the existence of a desire of a certain kind, i.e., one having an object that falls within the category defined by the passion. It is that category that gives the object of the passion itself.

We can also speak of two different phases of desire: *background desire,* and *effective desire.* This is a different distinction from that between latent and actual passions. Effective desires, let us say, are desires that are part of the rational explanation of some actual behavior. We have desires even when they are not effective in this sense. For example, I want to have dinner sometime tonight, but my wanting this does not figure in explaining any of my behavior during most of the day. The desire is present, but not effective. It is not irrelevant since attributing it supports certain counterfactual propositions in an economical way (e.g., if I were given the choice now of later having dinner or no dinner, I would choose dinner unless paid in some way to forego it, etc.). Both background and effective desires have determinate objects, unlike latent passions. The relevant latent passion in the dinner example is hunger. It is the category of potential desire for food. The passion of hunger is nonlatent whenever one has a desire, either background or effective, for food of some specific kind, or at some specific time. That is, a passion becomes nonlatent by acquiring a determinate object, and both background and effective desires have determinate objects. A passion or desire has a determinate object when there is some state of affairs that would satisfy it. Latent passions, then, have objects in a certain sense, but, unlike both background and effective desires, they do not have determinate objects.

[9] The puzzle might seem to appear here in even a more pure form. Self-love is a passion, however, that has as its object a balanced satisfaction of one's own *particu-*

all passions; it is rather a balanced satisfaction of all of the passions, with the requisite balance determined by nature.[10]

Of course, there would clearly be a loop if everyone's only passion were benevolence, but is the loop avoided by the presence of the other passions? We have already seen (sect. III) that, unless benevolence is specified as excluding from its object the benevolent passions of others, the addition of independent passions will not avoid the loop.

It is contrived and implausible to stipulate, however, that benevolence excludes from its object the benevolent concerns of the cared-for others without any reason to think they are any less important to that person's well-being. Butler indirectly gives good reasons for thinking that such a restricted benevolent passion could not be regarded as having as its object the happiness of others; its object would no longer count as happiness. One of Butler's recurrent lessons is that attention to one's benevolent passions is integral to happiness; it is an ineliminable part of the balanced pattern of satisfactions which is, according to nature, human happiness. It is precisely the neglect of one's own benevolent passions which is the premier failing of self-love, and the result of the failure is selfishness and lost happiness.[11] So, unless the satisfaction of your benevolence

lar passions. The particular passions are just all those except self-love, and if self-love were not excluded in this way the puzzle would indeed arise. Self-love would have as part of its object its own satisfaction. It is an interesting question whether it is legitimate to specify the object in this way (as Butler does), while doing so is here criticized as contrived in the case of benevolence. I do not believe the same arguments are available in this same-person case. Whereas the satisfaction of a loved one's benevolence could naturally be thought to be a crucial component in their well-being, I do not see how the same could be said about the satisfaction of one's own self-love, and, if not, then it is appropriate to exclude it from its own concern. One cares about the objects of one's passions, and one cares about one's own happiness (which involves the satisfaction of those passions), but why say one *also* cares about the satisfaction of one's own concern for one's own happiness?

[10] I offer the following textual composite, from Butler, as some support for this interpretation:

> Every bias, instinct, propension within, is a real part of our nature but not the whole (§404).
>
> . . . Neither can any human creature be said to act conformably to his constitution of nature unless he allows to [conscience] the absolute authority which is due to it (§379).
>
> Happiness or satisfaction consists only in the enjoyment of these objects which are by nature suited to our several particular appetites, passions, and affections (§417).

[11] ". . . the greatest satisfactions to ourselves depend upon our having benevolence in a due degree." Butler, Sermon I, §388 (references are to sections rather than pages, since these are used in the table of contents and index of that edition).

is part of the object of my passion, my passion does not have your happiness as its object. But then it would not be the passion in question—benevolence.

Suppose that instead of wholly disregarding the benevolence of others, *A*'s benevolence ignores only the *A*-regarding desires of others, thereby still heading off any possible loop. The restriction is ambiguous. Does it mean that *A*'s benevolence does not seek the satisfaction of (1) only *directly A*-regarding concerns, or (2) any concern which directly *or indirectly* is *A*-regarding, as in *B*'s concern for *C*'s concern for *A*? If (1), then only two-person versions of the loop are precluded, and so the puzzle is not avoided. Cases like the following would still be allowed: *A*'s concern for *B*'s concern for *C*'s concern for *A*'s benevolent passion. If the proposal means to rule out directly *and* indirectly *A*-regarding passions as in (2), then the proposal must be rejected as *ad hoc* and unrelated to the real nature of benevolence.[12] After all, there is no distinctive content of a desire that will make it *A*-regarding. Any other-regarding desire may be indirectly *A*-regarding if the facts about other people's desires turn out a certain way. It is implausible to build into the passion of benevolence a clause which excludes any desires of others which, in the facts of the case, turn out to be *A*-regarding. This would mean taking *A*'s concern for the satisfaction of *B*'s benevolent passions to be automatically cancelled if some direct or indirect object of *B*'s benevolence happens to care for *A*. Such factors are irrelevant to the nature of *A*'s concern for *B*.

It is unacceptable, then, to try to avoid the loop by conceiving the benevolent passions as excluding from their objects either the benevolent passions of others in general or those particular benevolent desires of others which concern oneself.

V. WHAT IS THE PROBLEM?

If Butler's moral psychology is correct in key respects, then desire loops are a serious problem. It would, of course, follow that such loops can occur in realistic situations rather than merely in philosophers' concocted examples, and though that by itself would be of

This passage is ambiguous, since it attends to both the having and the satisfying of benevolence. The choice between these is important to our topic and will be discussed below.

[12] It is *ad hoc* if it is ruled out just because of the loop. It would be less *ad hoc* if the objection is that such a desire turns out not to be benevolent but self-interested. The fact that its satisfaction would benefit oneself, however, is not a sufficient ground for denying that it is benevolent. It benefits oneself only because the happiness of the cared-for person depends on one's own happiness, not because of any concern for oneself.

some importance, it would not obviously be a problem. We could simply allow that such cases can occur in the real world and that, where they do, there are no conditions that would satisfy the passions involved. This would seem to shrink the significance of the loop problem, by analogy with loops in other contexts.

For example, it is well-known that similar loops can emerge in certain sets of statements, such as the following: *A* says only:

> "*B*'s statements are always correct."

And *B* says only:

> "*A*'s statements are always correct."

This statement of *A*'s is correct if and only if it is correct, and incorrect if and only if incorrect, and nothing else bears on its correctness. Such loops undeniably can occur in the real world, and so there are sentences for which there are no independent truth conditions. Assuming that some precise account can be given of this condition of "ungroundedness," as it has come to be called, the fact that it can occur may not present any special philosophical problems.[13] Similarly, the fact, important though it may be, that desires can be ungrounded in the real world, would not by itself present serious philosophical difficulties so far as I can see. At least that is not the sort of problem I want to allege.

There is more at stake in the Butlerian account of benevolence, however, than just the question whether desires *can* be ungrounded in real-world cases. If certain features of that view are correct, then benevolence is *always* ungrounded—nothing would count as satisfying anyone's benevolent desires. Most problematically of all, if Butler is correct about the relation between benevolence and happiness, then the general ungroundedness of benevolence renders human happiness utterly undefined. The conclusion is not just that happiness is difficult or even impossible to obtain; it is rather that there is, in principle, no condition that would count.[14] Now these conclusions

[13] For recent advances in the provision of such an account, see Robert Martin and Peter Woodruff, "On Representing 'True-in-*L*' in L," *Philosophia,* v (1975): 213–7; and Saul Kripke, "Outline of a Theory of Truth," this JOURNAL, LXXII, 19 (November 6, 1975): 690–716. Both are reprinted in Robert Martin, ed., *Recent Essays on Truth and the Liar Paradox* (New York: Oxford, 1984).

[14] On the Butlerian assumptions, self-love and benevolence are not just risky as Kripke says we must allow sentences involving truth to be. ". . . an adequate theory must allow our statements involving the notion of truth to be *risky:* they risk being paradoxical [or ungrounded] if the empirical facts are extremely (and unexpectedly) unfavorable." Kripke, *op. cit.* The present passage occurs on p. 55 in the Martin volume. In the Butlerian case, there is not just the risk but the guarantee of ungroundedness.

do not render the theory incoherent; as we have said, there is no incoherence or contradiction in supposing that the emptiness associated with ungrounded desires can actually obtain. The conclusion that the very notions of benevolence and happiness are empty in this way is, however, surely radical enough to warrant a closer examination of the Butlerian claims that would jointly establish it. In what follows, the argument will be that, barring a radical change in our common sense views about happiness, the features of Butler's theory that conspire to enervate benevolence and happiness are well founded. The philosophical difficulty this presents is acute, but it can be made clearer only after a closer critical look at the key Butlerian positions.

VI. THE SUFFICIENT BUTLERIAN POSITIONS

The two features of Butler's theory that would void benevolence of any satisfaction conditions are:

1. Benevolence aims at the good (happiness) of others.
2. A person's good (happiness) necessarily includes some degree of satisfaction of their benevolent passions.

It follows from these that each person's good depends on someone else's good, giving rise to an infinite chain of dependence with either infinitely many individuals as links, or (more likely) a loop.[15] Benevolence is, therefore, ungrounded (in the logicians' term), leaving happiness unachievable in principle. It is not just perfect happiness that is affected, but happiness of any degree, as will be discussed below.

The first claim is stipulative. There could certainly be disagreements about whether real benevolence, or any passion, has the good of others as its object, but these will be beside the point as long as (2) can be defended while understanding benevolence in the way set out in (1).

Consider next claim (2). It says that a person's good involves some degree of satisfaction of the benevolent passions. Now, since the benevolent passions have as their object the good of some others, then it follows that each person's good depends on the good of some others. But this could easily seem too strong. Perhaps *A*'s benevolence can be satisfied, at least to some extent [and no more is explic-

[15] Without also assuming that *everyone* has benevolent passions it might seem that only those who have benevolent passions will be affected by the loop. (2) says, however, that everyone's happiness depends on benevolent satisfactions, and it does not restrict this to those who have benevolence. Therefore, (1) and (2) are sufficient to implicate everyone in benevolence's ungroundedness, since there is nothing that would count as the satisfaction of benevolence for anyone—whether they happen to be benevolent or not. Still, the universal presence of benevolent passions is one natural way of supporting the connection between benevolence and happiness in (2).

itly required by (2)], by something less than *B*'s actual happiness, or good life, such as, for example, some contribution to it.[16] In that case, one person's happiness would not in general depend on another person's happiness, and still benevolent satisfactions (of this partial kind at least) would be acknowledged as necessary for happiness. Whereas benevolence would still have the *actual* happiness of others as its object (unlike the earlier contrivance of excluding benevolence satisfaction of others from benevolence's object), happiness would not require that one's benevolence be satisfied in this "whole" way; mere contribution satisfaction would be sufficient for happiness. This would avoid any loop that would render benevolence ungrounded. Even on this view, *A*'s benevolence could not be satisfied in a *whole* way without bringing some *B*'s benevolence into the picture (though, as just noted, *A*'s happiness would not require this). But no one else's benevolence (neither *A*'s nor some other *C*'s) need enter since *B*'s happiness (which is the "whole" object of *A*'s benevolence) requires only a contribution satisfaction of *B*'s benevolence. That could be achieved without any benevolent satisfactions of any others, and so the chain stops short of a loop. This account of the relation between benevolence and happiness purports to ground benevolence, and save happiness from meaninglessness.

The above account, however, oddly leaves benevolence's primary object, the happiness of others, out of the pattern that constitutes the benevolent person's happiness. All that matter are contributions to that condition. It is important to notice that the proposed account does not just say that happiness does not require complete satisfaction of benevolence; it says that happiness does not require even a single instance of the object of benevolence, the happiness of others. All that is said to be required is something else: contributions to happiness. Mere contributions to the object of a passion are not generally satisfaction of that passion at all. For example, the passion for shelter is not at all satisfied by the possession of some nails.

The proposal can be clarified by giving a separate status to the passion that has as its object contributions to the happiness of others. Call it *contribution benevolence*. Let us grant that its satisfaction is a necessary component of happiness. But certainly contributions are not the only objects of benevolence; there is also the happiness of others. That object is sufficiently different from mere contributions to happiness to individuate a separate passion—call it *happiness benevolence*. We must now ask, however, whether satisfaction of *this*

[16] I do not mean an act of contributing to it, but a contributing part or component itself.

passion is a necessary component of happiness.[17] If not, then the proposal succeeds at avoiding the loop. If it is, then happiness benevolence is ungrounded, and happiness is empty.

One could try to deny that the actual happiness of others is a necessary component of happiness, while still retaining the general view of happiness as the balanced satisfaction of the passions, by denying the existence of any fundamental human passion that aims specifically at the happiness of others rather than aim merely at contributions to it. That is, it could be argued that contribution benevolence is, at most, the only benevolence that is central and universal enough to be regarded as a necessary component of human happiness. And, as we have seen, that would avoid the loop. This move is inadequate, however, especially in the context of benevolence among loved ones. The benevolence of a parent for a child, or of one spouse for another, is more than a passion for contributions to their well-being; it is also a passion precisely for that well-being. Of course, one also desires lesser things such as benefits, and contributions to the well-being of loved ones. But nothing is more natural than to desire the well-being of loved ones itself; it is one of the fundamental human passions.[18]

In a cautious mood, we might wish to avoid such a broad appeal to human nature. Perhaps the claim does not hold for individuals in all cultures, or in all historical and material circumstances. Still, as long as there are some significant cultural or other domains in which the generalization appears to hold, then in that context the loop is guaranteed.

As another form of caution, we might note that, if the postulation of benevolence holds for almost all, but *not all,* individuals, even in a specific limited domain, that is enough to remove the guarantee of a loop. This is correct, but the situation would still be qualitatively different from the mere risk of a loop. Each person's happiness and benevolence would probably, even if not certainly, be rendered empty by loops. Whether these more cautious estimations are required or not, the scope of the problem remains considerable. For

[17] Treating it as a separate passion is only heuristic. We could just as well keep it as one passion and ask whether something more than contribution-satisfaction of this passion is required for happiness, e.g., "whole" satisfaction: the actual happiness of some others.

[18] If happiness were *defined* as requiring, among other things, the happiness of some other people, then the definitions would apparently be viciously circular. But this is avoided if happiness is defined simply as a balanced satisfaction of the passions, and each passion's role in happiness is regarded not as definitionally necessary, but as contingent on its actually being one of the human passions.

simplicity I shall go on to state the case in the less cautious way, as though all humans had such benevolent passions.

While accepting that everyone has such a fundamental passion for the happiness of some others, one might deny that happiness requires its satisfaction. Suppose someone had everything else that matters, even the near happiness of those about whom one especially cared. If the other satisfactions were had in sufficient abundance, could the person not be happy? The question is whether this absence of benevolence satisfaction could be compensated by extra satisfactions in other areas. If it could, then the charge that the loop is guaranteed to occur for all people should be dismissed. One's happiness would not in general depend on the happiness of someone else; nor, therefore, would the satisfaction of one's benevolence depend on that of another. It is a third condition of the ubiquity of benevolence loops that satisfactions of other passions cannot compensate for the nonsatisfaction of happiness benevolence, that different satisfactions are not interchangeable in this way.

Allowing such interchangeability is, in principle, a way of avoiding the loop. Should it be allowed? First, probably few would believe (these days) that happiness is sufficiently homogeneous for there to be *no* category of satisfaction that is indispensable. Most have rejected the view of happiness as a single psychic quantity that is produced by such different activities as pushpin and poetry. Therefore, the lack of one kind of satisfaction may be irremediable via other kinds. If the rejection of interchangeability of satisfactions is based on the difficulty in specifying some single feeling or psychic state that is common to all satisfactions, however, then the point does not carry over to the sort of nonpsychological satisfaction in question here (where a desire for x is satisfied if and only if x). But there is an influential and structurally similar move that is made in the latter case. The view is that happiness is not some separate state (psychological or not) that is produced by each of the multiplicity of categories of satisfaction. It is argued to be, in John Rawls's phrase, an *inclusive* rather than a *dominant* end. It consists of those other satisfactions (*op. cit.,* §83). Such a view must deny general interchangeability in order to avoid positing a single dominant end after all, namely, *satisfaction.*[19]

[19] Views that allow general interchangeability of satisfactions could avoid the loop problem, but I believe such views are seriously flawed. That cannot be argued here, but if they are then it matters much less whether they can avoid the benevolence loop or not.

It would not be surprising if some satisfactions necessary for happiness were indispensable in this way. But the indispensability may be limited to a few certain passions. We need to ask whether the set of indispensable satisfactions includes those of happiness benevolence.

The denial of complete interchangeability of different kinds of satisfactions is powerful; we no longer expect any satisfactions to be interchangeable. Perhaps some are, for special reasons or in certain circumstances. For example, it may well be that satisfactions can be interchanged in the context of a mediocre life. A mediocre life without achievement can be just as good as one with achievement if the former has some satisfactions that the other lacks. Mediocre lives are fundamentally flawed, and the flaws might as well be in one category as another as far as the degree of well-being is concerned. But a truly happy life, human flourishing, is not fundamentally flawed. That does not mean that only a perfect life is a truly happy life; even the happy life could always be better, and so there is no perfect life. Rather, what must be meant, at least in part, by a life that is not fundamentally flawed is one that is not deficient (at the very least, not wholly lacking) in respect of any fundamental component of well-being. No catalogue of fundamental components is needed here. Only one candidate matters for our purposes. Even if one has more of some other satisfaction than usual, a person whose passion for the happiness of others (one's happiness benevolence) is not satisfied, at least in the case of some loved ones, is deprived in a fundamental way. Of course, such a life could still be valuable, but a truly happy life would not have this fundamental flaw (or so a satisfaction account of happiness seems to be compelled to say.)[20] I conclude that satisfactions are not interchangeable in the way that would be needed to avoid the loop.

The importance of the loop problem would be vastly greater if it affected not only satisfaction views that are concerned with *actual* desires or passions, but also with the currently more popular satisfaction views that deal with desires that are, in certain ways, *ideal*. The latter sort of satisfaction view is widely preferred for avoiding the problem of rash, ill-considered desires. Still, such views may well be subject to the same kind of benevolence loop as actual desire theories, and so to whatever problems this might involve. The following considerations lend some support to this possibility.

[20] Rawls apparently sees benevolence as an indispensable part of a person's good, §75). This does not yet make benevolent *satisfactions* indispensable, which is the issue. There seems no reason to doubt it, however, as will be suggested below.

Many adopt some version of the view that what is good for a person is what they would desire if they were fully informed, and I shall concentrate on these.[21] Suppose that

1. One thing anyone would want if fully informed is the good of some others.
2. A (nonhypothetical) person's good requires that their hypothetical benevolent desires [from (1)] be satisfied.

Clearly then, one person's good would depend on another's, in a loop.

Do the "ideal theorists" need to accept (1)? The fully informed person would surely want a life with deep personal relations of love and friendship.[22] These in turn would involve desires for the good of others. But so far all this shows is that *having these desires* is part of the good life. It does not yet show that the good of others (the object of the desires) is part of the good life as (1) implies.

The only way to avoid the conclusion, however, is to suppose that the fully informed person would *want to want* the good of others—would want to have such concerns, but would not necessarily *want the good of others.* There is no contradiction in this, but nor is there any good reason to believe it. Without some special explanation, it is natural to assume (in the context of a satisfaction account of happiness) that, if having the passion is part of one's good, then, for that reason, so is the object of the passion. This follows from the normal efficacy of passions. But even apart from that point we would expect the fully informed to want the good of at least some others in the more direct way.

Clause (2) prompts a distinction between a person's benefit, and their happiness or good, since no particular hypothetical-desire satisfaction is necessary for a person's benefit. But are there any that are necessary for one's happiness as (2) states? The same arguments used above (for the actual-desire version) for the necessity of some satis-

[21] Consider Rawls: "Thus the best plan for an individual is the one that he would adopt if he possessed full information. It is the objectively rational plan for him and determines his real good," p. 417). John Harsanyi's view is that "true" preferences are those one "*would* have if he had all the relevant factual information" and was rational. Social utility should be based on these hypothetical "true" preferences, though the account of individual utility should also consult actual "manifest" preferences "in a suitable way"; see "Morality and the Theory of Rational Behaviour," in *Utilitarianism and Beyond,* Amartya Sen and Bernard Williams, eds. (New York: Cambridge, 1982), pp. 39–62, esp. pp. 55–6. James Griffin has a similar view: " 'utility' is the fulfillment of informed desires," "an 'informed' desire is one formed by appreciation of the nature of the object"; cf. *Well-Being: Its Meaning, Measurement, and Moral Importance* (New York: Oxford, 1986), p. 14.

[22] Rawls, §75, and Griffin, pp. 67–8.

faction of happiness benevolence can apparently be applied in the ideal-desire theories as well. The ideal-desire theories should not diverge from actual-desire theories in the case of actual desires that represent deep, central human passions.

Ideal-desire satisfaction theories of happiness seem compelled to grant (1) and (2), and that would give rise to the same loop that attaches to actual-desire satisfaction theories. Therefore, there is reason to doubt that moving from actual- to ideal-desire accounts of well-being will give a decisive advantage in avoiding the loop.

If our theory is that a person's happiness depends on satisfactions of desires, passions, etc., then it would be hard not to allow that the happiness of at least some people could, in principle, depend on the happiness of some others. But if this is allowed in principle, then all that is needed for the problem of the loop to emerge is a certain empirical situation, namely, that this situation which is allowed by the theory to hold for at least some people happens to hold for all people. And from within the view that happiness is constituted by satisfactions of (real or ideal) desires, it would apparently be arbitrary not to include as necessary the satisfaction, to some degree or other, of as central a passion as that for the happiness of some others. The problem is that including these passions will subject happiness to the consequences of a satisfaction loop, consequences which are severe: the ungroundedness of (happiness) benevolence, and the emptiness of happiness. This outcome is not incoherent, but still, it is such a radical conclusion (and such an undesirable one) that the assumptions that lead to it deserve scrutiny, even skepticism. They are, at the very least, put under pressure.

VII. IMPLICATIONS

If the relevance to one person's happiness of the happiness of others cannot be denied within a satisfaction model of happiness, we must ask on what deeper assumptions the radical conclusion depends, and whether they should be retained under this pressure. The satisfaction model of happiness should itself be scrutinized in this context. Now, although the radical conclusion can be easily shown to depend on a satisfaction model, that is insufficient reason for abandoning the model. We might, on balance, have more reason for retaining the satisfaction model than for fleeing the radical conclusions about benevolence and happiness. But that is what needs to be determined.

No full evaluation of the satisfaction model is possible here, but the importance and interest of such a project are heightened by two considerations which can be briefly discussed. First, and most importantly, the loop problem can be shown to depend on a satisfaction model of happiness. Second, since the problem has emerged out of a consideration of certain views of Butler's, it is of some interest to note that the issue of whether happiness is a matter of satisfaction or not is present, though not explicit, in Butler's own thought.

First, the loop cannot occur without the assumption that a person's good depends on the satisfaction of passions or desires in some way. Suppose instead that happiness consisted in *having* certain passions (to certain degrees and in certain combinations), apart from whether they were satisfied. One person's benevolence would still have another person's good as its object, and that second person's good would still involve their benevolence in a certain way. It would not be the satisfaction of that benevolence, however, but only the having of it that figures in the person's good. Therefore, having someone's good as the object of my benevolence would not involve also taking on the objects of that other person's benevolence. Their good is comprised of their character, and their having the proper character is largely independent of their getting what they, in virtue of their character, want. The point applies more generally as well. Whatever happiness does depend on (character, pleasure, an objective list of goods, etc.), as long as it does not depend on the satisfaction of passions and desires my benevolence will not automatically take on the object of yours. A nonsatisfaction account of well-being, then, would avoid the puzzle of ungrounded desire as it arises in the case of mutual benevolence.

There is one kind of satisfaction view that is well positioned to avoid the loop. Rather than deny the relevance of the happiness of others in an *ad hoc* way, one might deny the relevance to one's own happiness of any state of affairs that is not actually an effect on oneself. Indeed, I think there is much to be said for such an "effect requirement." If it is admitted, then the only acceptable satisfaction view of happiness will be the view that

> something bears on one's happiness if and only if it both satisfies some (real or ideal) desire, and is an effect or change in oneself—that is, if and only if it satisfies some desire for some effect on oneself.[23]

Like any view that denies that the happiness of one person requires the happiness of some others, this view avoids the loop. I mention it only because the device by which it avoids the loop is well motivated in its own right. But, of course, it is forced to bite the bullet and disqualify many kinds of desire satisfaction that more standard satisfaction views would like to include. One could, of course, accept the effect requirement without trying to reconcile it with a satisfaction

[23] That need not be the description under which the person desires it.

view of happiness at all, three of the salient alternatives being character views, hedonist views, and objective-list views.

Although the language of satisfaction and gratification occurs often in Butler, there is some reason to wonder whether he might have held a character view of happiness rather than a satisfaction view. First, while it may seem that a central Butlerian argument, the argument against psychological egoism, depends on a satisfaction view, this is not actually so. He argued from the assumption that a person's good consists in the balanced satisfaction of one's passions, that therefore, if one's only passion were for one's own good, it would simply have its own satisfaction as its object (*op. cit.*, sect. 383; cf. also fn. 9). This was held to be untenable, and we can see why in the fact that it is an instance of the sort of ungrounded desire that is our central topic; indeed, it is the purest case since it involves only one person and one desire. Butler concluded that there must be passions for things other than our own well-being. It may look, then, as though it is part of the Butlerian rejection of psychological egoism to suppose that our well-being consists in getting what we care about in certain ways. But Butler's deeper point does not depend on this assumption. Butler's lesson is that we cannot understand a person's good in isolation from what matters to the person besides her own good. This, if true, would be enough to refute psychological egoism. I have stated this lesson carefully, however, so as to show that it does not require a satisfaction view of well-being. The fact that an account of well-being is impossible without discussing what else matters to the person does not show that well-being consists in getting what matters to us. It could just as well consist in the mattering itself. Butler's lesson is consistent with holding that well-being is a pattern of concerns rather than a pattern of satisfactions.

There is also more positive evidence that Butler's own views betray at least some ambivalence about the choice between a character view and a satisfaction view of well-being. In one of the more ambivalent formulations, he writes, "the greatest *satisfactions* to ourselves depend upon our *having* benevolence in a due degree" (*op. cit.*, sect. 388; emphasis added). We see elsewhere, however, that Butler sometimes speaks of a kind of satisfaction that is not the obtaining of the object of any passion. For example,

> Happiness consists in the gratification of certain affections, appetites, passions, with objects which are by nature adapted to them. . . . Love of our neighbour is one of those affections. This, considered as a *virtuous principle,* is gratified by a consciousness of *endeavoring* to promote the good of others; but considered as a natural affection, its gratification

> consists in the actual accomplishment of this endeavour (*op. cit.*, sect. 421).

"Love of our neighbor" is, in Butler, another name for benevolence, the passion for the good of others. The good of others is its object, but yet he says that in a certain way it can be satisfied by something else altogether, the very possession and exercise of the passion. This is not satisfaction in the sense that is present in the notion of a satisfaction view of happiness—the obtaining of the object of a passion or desire. If this other sort of satisfaction can be at least part of the balanced satisfaction of the passions that is happiness, then according to Butler at least some measure of well-being can come from certain concerns themselves, apart from whether they are met.[24]

These passages are inconclusive, but suggestive of some measure of agreement with Aristotle that happiness is the exercise of virtue —a character view of happiness, not a satisfaction view. If this is Butler's view, then it is not troubled by the loop we have been discussing. And whatever Butler may have believed, the loop problem appears to put pressure on satisfaction views of well-being, pressure that does not apply to the view that well-being is a matter of character, nor to other nonsatisfaction views, such as hedonist or objective-list views. As for which model of happiness is ultimately superior, this is only one among many important considerations.

DAVID M. ESTLUND

University of California/Irvine

[24] That he still speaks of this case in terms of gratification should make us cautious in our interpretation of his other satisfaction-like language. It does not necessarily indicate the presence of a satisfaction model of happiness.