

## THE LIMITS OF WELL-BEING

BY SHELLY KAGAN

### I. THE DIALECTIC

What are the limits of well-being? This question nicely captures one of the central debates concerning the nature of the individual human good. For rival theories differ as to what sort of facts directly constitute a person's being well-off. On some views, well-being is limited to the presence of pleasure and the absence of pain.<sup>1</sup> But other views push the boundaries of well-being beyond this, so that it encompasses a variety of mental states, not merely pleasure alone. Some theories then draw the line here, limiting well-being to the presence of the appropriately broadened set of mental states. But still others extend the limits of well-being even further, so that it is constituted in part by facts that are not themselves mental states at all; on such views, well-being is partly constituted by states of affairs that are "external" to the individual's experiences.

In this essay, I want to explore some of this debate by focusing on a particular stretch of the dialectic. That is, I want to think hard about a particular connected series of arguments and counterarguments. These arguments—or, at least, the concerns they seek to express—emerge naturally in the give and take of philosophical discussion.<sup>2</sup> Together they make up a rather simple story, whose plot, in very rough terms, is this: first there is an attempt to push the limits of well-being outward, moving from a narrow to a broader conception; then comes the claim that the resulting notion is too broad, and so we must retreat to a narrower conception after all.

Giving the plot with greater care requires first a bit of machinery. In what is rapidly becoming the traditional classification, there are three basic types of theories of well-being. First, there are *mental state* theories, which hold that an individual's well-being consists solely in the presence of the relevant kinds of mental states. Hedonism is, of course, the most familiar theory of this sort, claiming that well-being consists in the presence of pleasure. But as already noted, one might accept a broader mental state theory according to which, even though being well-off was a

<sup>1</sup> Hereafter I will focus only on the *goods* of a given theory of well-being (e.g., pleasure); for simplicity, I will not discuss the *bads* (e.g., pain).

<sup>2</sup> Examples are to be found in James Griffin, *Well-Being* (Oxford: Oxford University Press, 1986), chapter 1, and in Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), appendix I. Although I will not be concerned with the specifics of either of these discussions, I have obviously learned—and taken—a great deal from both of them.

matter of having the right mental states, the direct relevance of a given mental state need not be exhausted by its pleasantness.

Second, there are *desire* or preference theories, which hold that being well-off is a matter of having one's (intrinsic) desires satisfied. What is intended here, of course, is "satisfaction" in the logician's sense: the question is simply whether or not the states of affairs that are the objects of one's various desires obtain; it is irrelevant whether or not one realizes it, or whether one gets some psychological feeling of satisfaction. (Some desire theorists restrict well-being to the satisfaction of the desires one would have if fully informed and rational, but for our purposes these refinements will not matter.)

Finally, there are *objective* or objective list theories, which hold that various things are objectively good for a person to have, whether or not he realizes it, and whether or not he desires it. Being well-off is simply a matter of one's having the various objective goods. These might include not only pleasure, but also, for example, friendship, fame, knowledge, or wealth. The list of objective goods is, of course, a matter of dispute, but there is no obvious reason to think it would be restricted to kinds of mental states.

One effect of the discussion to follow will be to call into question whether this classificatory scheme is in fact the most illuminating way to divide the terrain. But it will do to get us started.

Here, then, is the stretch of the dialectic to which I want to draw our attention. It begins with hedonism. And it challenges the hedonist to offer an account of what all pleasant mental states have in common. After all, it is not as though "pleasure" designates a single *kind* of mental state. The experiences of taking a hike, thinking about moral philosophy, and eating chocolate can all be pleasant ones, yet they are obviously quite different in their content. Nor does it seem as though there is an identifiable ingredient or "component" that they share that we are prepared to label as the pleasure. But if pleasure is neither a single kind of mental state, nor a single component of mental states, what then is it for a mental state to be a pleasant one?

The answer that gets offered to the hedonist is this: a mental state is pleasant if it is desired. And to say that one mental state is more pleasant than another is to say that the former is preferred to the latter.

Perhaps this is not, on reflection, an altogether satisfactory analysis of pleasure. But it doesn't really matter, for no better analysis is forthcoming. It is the best the hedonist can do, and if she is not to abandon her position altogether, she will have to make do with it. Given this account of pleasure, then, hedonism becomes a theory of well-being that might with equal justice be called *preference mental statism*: I am well-off to the extent that I have the various mental states that I desire.

But once we have gone this far—the argument goes—it is hard to avoid going further. Recasting hedonism as preference mental statism allows us

to see that it is really *preference* that is doing the work. One mental state is more valuable than another—makes a greater contribution to our well-being—by virtue of the fact that we prefer it. But it is obvious enough that our desires are not *limited* to matters involving our mental states. We want various “external” states of affairs to obtain as well. Indeed, sometimes we prefer that some such external state of affairs obtain even if that means that we will be subject to some undesirable mental state. Yet if satisfaction of preference is sufficient to ground well-being when it is a matter of preferences *between* mental states, why shouldn’t it be sufficient to ground well-being *regardless* of the object of the desire?

In this way we seem pushed toward a pure or unrestricted desire theory. What makes one well-off is the satisfaction of one’s desires, period. The restriction to desires concerning one’s mental states falls away as unmotivated. No doubt people typically do have desires concerning their own mental states, and these are often among their most central desires. But well-being is simply a matter of the satisfaction of one’s desires, and these will usually include desires concerning external states of affairs as well.

Unfortunately, the move to the desire theory brings its own difficulties. For in pushing outward from hedonism to the desire theory, we seem to have pushed too far. There are many cases where it seems altogether implausible to suggest that the satisfaction of the relevant desire affects the person’s level of well-being. Suppose I meet a stranger on a train.<sup>3</sup> He tells me his story, and I form the desire that he succeed in his projects. We then part, and I never hear of him or even think of him again. If he does in fact succeed, then my desire has been satisfied. According to the desire theory, then, this makes me better off. But this is intuitively an absurd claim. Obviously my level of well-being is not affected *at all* by the success of the stranger. The success of the stranger has nothing to do with *me*.

This suggests that the unrestricted desire theory is hopelessly broad. A theory of well-being must explain which facts constitute my being better off. So they must be facts about *me*. Since my desires can range over facts that have nothing whatsoever to do with me, the satisfaction of such desires cannot constitute my well-being. If we are to preserve a desire theory at all, we will have to move to some sort of restricted desire theory, according to which my well-being consists in the satisfaction of the appropriately restricted subset of my desires. The trick, of course, is to provide some plausible specification of the restriction.

From this perspective, the position of mental statism no longer seems so arbitrary. At least it seems to keep the content of well-being within reasonable bounds, for facts about my mental states are certainly facts about me. In contrast, it is far from clear whether anything external to my mind

<sup>3</sup> The example is Parfit’s, from *Reasons and Persons*, appendix I.

can count as well, that is, can count as—in the relevant sense—facts about me. If not, then the limits of well-being must be drawn at the limits of our minds.

This is the stretch of argument and counterargument which I propose to examine more carefully. Please note that I have not yet endorsed any of it. Now in point of fact, I am inclined to think that much of the argument is mistaken, although I also suspect that some of the conclusions can be supported for all that. But with regard to other portions of the argument I am rather uncertain as to where the truth lies. So the discussion that follows is indeed a genuine exploration: everything is rough, and everything is tentative.

## II. HEDONISM

Let us begin again at the start of the argument, with the challenge to the hedonist to disclose what it is that pleasant mental states have in common. It does seem correct to me to note that 'pleasure' does not pick out a single kind of mental state or experience, or a single shared component of all pleasant experiences, or even a kind of component. But it is, I think, too quick to conclude from this that the best that can be said by the hedonist is that a mental state is pleasant if it is desired.

An alternative move is to identify pleasantness not as a component of experiences, but rather as a *dimension* along which experiences can vary.<sup>4</sup> As an analogy, consider the loudness of auditory experiences—that is, sounds. It is obvious that loudness or volume is not a *kind* of sound. And it seems plausible to insist that loudness is not a single kind of component of auditory experiences. Rather, volume is a dimension along which sounds can vary. It is an aspect of sounds, with regard to which they can be ranked. Recognition of the qualitative differences between the sounds of a symphony, rain falling, and a bird chirping, does nothing at all to call into question our ability to identify a single dimension—volume—with regard to which these and other sounds can be ranked.

Similarly, then, pleasure might well be a distinct dimension of mental states, with regard to which they can be ranked as well. Recognition of the qualitative differences between the experiences of hiking, listening to music, and reading philosophy, need not call into question our ability to identify a single dimension—pleasure—along which they vary in magnitude.

Once we have a picture like this in mind, we might in fact be prepared to insist that there is a sense in which pleasure is an ingredient common to all pleasant experiences. For it seems to me that there is a sense in

<sup>4</sup>I am here influenced by unpublished work by Leonard Katz. See his forthcoming book, *Hedonism: A View of Mind and Value*.

which a specific volume is indeed an ingredient of a given sound, along with a particular pitch, and so forth. (Similarly, intensity or saturation is an ingredient of colors, along with hue.) Thus, pleasantness might well be considered an ingredient of (conscious) mental states in general, albeit an ingredient that we will only notice if we "chop up" experiences in some nonstandard ways. But whether or not pleasure can be helpfully viewed in this way as an ingredient of experiences, the possibility remains that it is a single, specific dimension along which experiences vary.

Of course, even if this is granted, the challenge might still be raised to identify the particular dimension which pleasure is. Mental states can presumably vary along a number of dimensions, and we would like to have some illuminating way to characterize the particular dimension with which discussion of pleasure is concerned.

It is not at all clear to me whether it would be objectionable if this new challenge could not be met. Suppose there were no interesting way to identify the pleasure dimension except ostensively—drawing someone's attention to different experiences which varied significantly and saliently in terms of this particular dimension. I do not see how this would threaten the intelligibility of the various claims the hedonist wants to make concerning pleasure.

Of course, this is not to say that there is no more theoretically satisfying way to identify the pleasure dimension. After all, the connection between pleasure and desire is a striking one. We may well be able to use the latter notion to help fix the referent of the former. Indeed, the original proposal to analyze pleasure as desired mental states can be seen as a friendly suggestion along these lines.

But we can do better. More sophisticated accounts have been offered. No doubt, pleasant experiences are desired, but they seem to be desired in a particular way and for particular reasons. A more plausible account might look something like this:

An experience *E* that occurs at time *t* to a person *P* is pleasant if and only if:

- (1) *P* has a desire at *t* that *E* occur at *t*
- (2) *P*'s desire is an immediate response to *E*'s occurrent phenomenal qualities (i.e., its qualia).

Clause (1) brings out the fact that one likes pleasant experiences while they are occurring. Clause (2) brings out the fact that one likes pleasant experiences because of how they *feel*. This means, roughly, that one's desire is a direct, immediate response to the phenomenal qualities of the experience—and only to those properties of the experience. It is not mediated by further beliefs. In particular, the desire cannot be a response to recognition of the instrumental (or evidential, or moral) properties of

the experience. We often are glad to have a given experience when we know it will lead to certain other experiences, but that doesn't make the first experience pleasant in itself.

I certainly do not mean to suggest that this improved account will do as it stands. Clause (2) in particular needs considerable refinement. But if an account anything at all like this is correct, it allows us to raise certain questions.

So suppose that something like this improved account is correct. How does this connect to our earlier suggestion that pleasure is a particular dimension along which mental states vary? One possibility is this: A mental state is pleasant if it is the object of the appropriate sort of immediate desire. Mental states can be ranked according to whether or not they generate this kind of desire, and if so, the strength of the desire. The stronger the desire, the more pleasant the mental state.

In effect, we have identified a dimension along which mental states can vary—strength of the generated immediate desire. And if something like the improved account is correct, then this dimension varies in lock step with the dimension of pleasure. But this is still subject to two possible interpretations. One could hold the "reductionist" view, according to which the two dimensions are not genuinely distinct ontologically, but are one and the same. Pleasantness simply *is* strength of immediate desire. Alternatively, one could hold the "nonreductionist" view, according to which the two dimensions remain ontologically distinct. We may use the desire dimension to help us identify the pleasure dimension, but pleasure remains a psychological reality distinct from immediate desire. Presumably on this nonreductionist view, it is the presence of this distinct psychological magnitude that helps to *explain* the presence of the corresponding desire.

I will not attempt to adjudicate between these two interpretations here. Let us consider the implications of each. Suppose, first, that we adopt the reductionist approach. Then for a mental state to be pleasant simply is for it to be the object of the appropriate sort of immediate desire. Now when the hedonist claims that well-being consists in the presence of pleasant mental states, she is presumably claiming that it is the pleasantness of the pleasant mental states that makes me better off. So on the reductionist approach the hedonist is claiming that when I am well-off, I am well-off by virtue of the fact that various mental states for which I have the right sort of immediate desires do in fact obtain.

This does seem to be a version of the desire theory. Perhaps not all desires are such that their satisfaction contributes to my well-being, but for an important subset of my desires—the immediate desires—their satisfaction is constitutive of my being well-off. I am well-off by virtue of the fact that desires of the relevant kind are satisfied. I belabor this point because I am not completely convinced of it, although it seems correct. If I am well-off this is because I have the right kinds of mental states, namely,

pleasant ones. But these states only count as pleasant by virtue of the fact that I have the appropriately immediate desires for their occurrence. If the mental states themselves occurred in the absence of the immediate desires (admittedly, something that might not be causally possible), they would not be pleasant states, and I would not be the better off for their occurrence. So if I am better off, this is only because the immediate desires do in fact exist and are satisfied (that is, the immediately desired mental states do obtain). So it is the satisfaction of the right kind of desires that grounds my well-being. This, as I say, seems to be the implication of adopting the reductionist stance.

In contrast, if the hedonist is a nonreductionist, she is not similarly forced to adopt a version of the desire theory. According to the nonreductionist approach, pleasure is a specific psychological magnitude that we can identify through attending to appropriately immediate desires, but the pleasantness is nonetheless distinct from the desiredness. So when the hedonist claims that being well-off is a matter of having pleasant mental states, she need not claim that when one is well-off this is by virtue of the fact that various desires are satisfied. Of course, since pleasant mental states are immediately desired, it is in fact the case that when one is well-off one has immediate desires and they are indeed satisfied. Nonetheless, the hedonist can insist that it is not by virtue of *this* fact that one is well-off. Rather it is by virtue of the fact that one's mental states exhibit the specific and distinct psychological magnitude—pleasantness—to a significant degree. Even if one did *not* have any desire at all for the occurrence of the pleasant mental states (again, something that admittedly might not be causally possible), one would still be well-off to the extent that one had the pleasant mental states themselves.

On this approach, what the hedonist is endorsing appears to be a version of an objective theory. There is a particular kind of good, namely pleasure, that it is objectively good for a person to have. Its goodness is not at all founded in the fact that the person happens to desire it. In effect, the hedonist is offering an objective list theory with a very short list. Pleasure is an objective good, and it is the *only* such good.

It seems then, that the nonreductionist hedonist need not embrace any version of the desire theory. She can instead adopt an objective approach. But this is not to say that the nonreductionist hedonist *cannot* endorse a desire approach. As far as I can see, it is completely open to the nonreductionist to insist that even though pleasantness is a distinct psychological dimension from that of strength of generated immediate desire, pleasant mental states are indeed valuable by virtue of the fact that they are the object of this sort of desire. That is, even the nonreductionist can hold that when I am well-off, this is by virtue of the fact that the right sorts of desires are satisfied.

In short, it seems that the hedonist must choose between accepting a version of the desire theory (an option open to both the reductionist and

the nonreductionist) and accepting a version of the objective list theory (an option that seems to be open only to the nonreductionist). Insofar as the original dialectic with which this essay began assumed that the hedonist must appeal to some version of the desire theory, it seems to be in error. There is also the possibility of objective hedonism.

Nonetheless, some hedonists will indeed prefer to embrace the desire theory sketched above, according to which well-being consists in the satisfaction of the appropriate sort of immediate desires. And for these hedonists, at the very least, it does seem that we can push a line of attack similar to that given in the original dialectic. For what these hedonists do is identify a particular kind of desire—immediate desires of the right sort—and hold that for this kind of desire alone is its satisfaction constitutive of my well-being. This is doubtless a consistent position to hold, but it is difficult to see how it is to be motivated. Why shouldn't there be other desires whose satisfaction contributes to my well-being as well—desires which admittedly fail to be immediate in the relevant sense?

Consider the first of the two conditions used to characterize pleasant experiences: an experience is pleasant only if the individual has a desire *at the time of its occurrence* for the occurrence of that experience. This seems a reasonable condition to place on what is to count as pleasant: one likes pleasant experiences while they are occurring. But what could possibly explain why only desires that meet this condition are such that their satisfaction leaves me better off?

Suppose there were a kind of experience such that while it was occurring, the person had *no* desire that it occur at that time, but immediately after its conclusion the person was spontaneously glad that he had had the experience. His desire or "pro attitude" toward the earlier experience might still be immediate in the sense intended by the second of the two conditions—it is a direct response to the immediate phenomenal qualities of the experience, unmediated by cognitive concerns such as the instrumental value of the experience. But the desire occurs later rather than concurrently with the experience.<sup>5</sup>

If there were such experiences, they would not be labeled as pleasant by the improved account. So the hedonist would not count them as contributing to one's well-being. But why shouldn't they count? What could possibly explain why the satisfaction of a desire should count toward my well-being when the obtaining of its object is concurrent with the existence of the desire, but not count when the obtaining of its object is not concurrent?

Similar questions arise if we focus instead on the second of the two restrictions. Pleasant mental states are desired in an immediate—that is, not cognitively mediated—response to their phenomenal qualities. But sup-

<sup>5</sup> Are there any such desires? Their logical possibility suffices to make my point, but I am intrigued by the empirical question as well. Perhaps certain meditative experiences are too "empty" to allow for the desire at the same time; other experiences might be too "full" or consuming to allow for the concurrent desire, crowding it out.



pose there were some mental states that were intrinsically desired, but in a nonimmediate manner. The occurrence of the given mental state might be desired at the time of its occurrence, but only because the person had one or another belief concerning some of the properties of that state. That is, one might find various mental states intrinsically attractive, but only as a result of reflection.

Once more, if there were such experiences, the hedonist would not view them as pleasant, and thus would not count them toward my well-being. But why shouldn't they count too? There is, of course, a difference between desiring a mental state solely in response to its phenomenal qualities and desiring a mental state at least in part because of one's various other beliefs. But what could possibly explain why it is only the satisfaction of the former sort of desires that counts toward my well-being, and never the satisfaction of the latter sort of desires?

It seems that the hedonist has two options. She might concede the theoretical point, and admit that *if* one had desires of these various kinds, then their satisfaction too would contribute to one's well-being. Or she could simply insist that even if one did have desires of these kinds, one's well-being would be completely unaffected by their satisfaction.

On the first option, the hedonist is conceding that well-being is a matter of the satisfaction of one's various desires concerning one's mental states. She could, of course, go on to argue that, as a matter of empirical fact, one never does have intrinsic desires for mental states except in the case of the kind of immediate desires that ground pleasures. (Or perhaps: one *would* not, if one were fully informed and rational.) But even if such an empirical claim were plausible, the point would still remain that the hedonist would indeed have been forced to accept preference mental statism. I am well-off to the extent that I have the mental states that I desire. If the only states I desire are pleasant ones, so be it. But even if this is so, the differences between pleasantness and other aspects of mental states are of no fundamental significance to the theory of well-being. What would be fundamental would be the theory of preference mental statism.

If the hedonist did take this option, then it also seems right that she could be pushed further still. Why should it be that only preferences concerning my mental states are relevant to my well-being? Up to this stage of the dialectic, it is indeed difficult to see how the hedonist can justify restricting the set of relevant desires. Of course, the hedonist might try to argue that, in fact, I have no intrinsic desires concerning matters other than my mental states (or would have none were I rational and fully informed). But whether or not this empirical claim is true, the underlying position now does seem to be an appeal to an unrestricted desire theory.

This is what happens if the hedonist concedes that *if* someone had desires other than the appropriate kind of immediate desires, then they would indeed be relevant to well-being. But the hedonist might prefer the other option, holding that *only* immediate desires would be relevant to

well-being. As I have already noted, this is a consistent position, but it is an unsatisfactory one philosophically. Why exactly is it that only the one sort of desire contributes to well-being?

The hedonist might, I suppose, attempt to defend her position by appeal to an inspection of cases. Perhaps we will agree that—inexplicable as this may be—well-being is affected in all and only those cases where immediate desires are at stake. But nonhedonists are likely to disagree about the results of this survey, and it leaves the hedonist's position seeming rather vulnerable. And even those who agree with the hedonist about the results of her survey should find it philosophically unsettling to lack any account of why the various distinctions to which she appeals matter.

However, none of this amounts to a proof that the various restrictions and distinctions cannot be defended. Indeed, in the next section of this essay we will turn to the second half of the dialectic and consider an argument for the view that well-being must be solely a matter of having the right sorts of mental states. If such an argument were successful, then the hedonist's rejection of an unrestricted desire theory would not be unmotivated. The restriction to preference *mental statism* would have some foundation. Similarly, then, there might be additional arguments justifying the hedonist's further restriction to *immediate* desires. As I have already indicated, I do not currently have any plausible suggestions for arguments along these lines, and I am strongly inclined to doubt that any such arguments would succeed. But it seems to me premature to conclude that there are no such arguments.

Up to this point we have been considering the implications of the hedonist's adopting a version of the desire theory. But I noted the possibility of the hedonist preferring an objective theory, and I want to consider the implications of this approach as well. Here, however, I can be more brief. For it seems to me that the objective hedonist faces difficulties that closely parallel those that trouble the hedonist who insists upon the fundamental significance of immediate desires.

Of course, if the hedonist does take an objective approach, then to the extent that immediate desires are especially significant, this will simply be because of the special significance of the objects that are uniquely picked out by such desires. That is, the hedonist will no longer need to argue for the intrinsic significance of certain distinctions within a desire theory. Instead, the defense of hedonism will require arguing for the uniquely valuable nature of the mental states such immediate desires pick out. The hedonist must insist that it is simply a fact that certain mental states, namely the pleasant ones, are objectively good for the person to have.

But it is obvious that this leaves the hedonist with a great deal to explain as well. Even if we put aside the difficulty of arguing that pleasure is an objective good, there is the significant difficulty of arguing that it is

the *only* such good. Why is it that of all the various aspects of mental states, it is only this particular dimension that contributes to a person's well-being? What explains the unique significance of *pleasure* as distinct from the various other aspects of mental states for which they might be valued? There is, no doubt, a strong temptation to appeal to the fact that only pleasure involves being immediately desired—but this would push the hedonist back in the direction of the desire theory. Yet what else could explain why this should be the only aspect of mental states that contributes to being well-off?

Once the hedonist is engaged in offering arguments for the objective value of pleasure, the possibility emerges that other aspects of mental states might be shown to have objective value as well. These might well be aspects whose value can only be recognized upon reflection, given the appropriate beliefs. As such, they will differ from the aspect of pleasure, whose value is apparently recognized immediately. But it is difficult to see why the only objective goods should be ones whose value is recognized directly and immediately.

In this way even the objective hedonist can be pushed toward the possibility of a more general mental statism. Even if well-being were limited to the presence of the right mental states, it might well be that the relevance of a given mental state is not limited to its pleasantness. And of course, once we are this far, it seems that we can again push further still. Why should the objective goods that contribute to well-being be limited to the possession of the right sorts of mental states? Once the possibility of objective goods is recognized, it seems an open question whether they include goods external to the individual's mind.

Of course, as should also be obvious, none of this amounts to a refutation of the hedonist's position. We are about to turn to an argument that well-being must be limited to having the proper mental states. Should this argument—or another to the same effect—be successful, then clearly it is not ad hoc for the hedonist to restrict her attention to mental states. And there might well be other, additional arguments that pleasure is the sole aspect of a mental state that can contribute to its value. Once more, I do not currently know what these further arguments might look like, but it seems premature to assume that none could succeed.

To summarize quickly, I think the first half of the dialectic is correct in claiming that there is a certain kind of philosophical pressure that can be put upon the hedonist, pushing outward from hedonism to a more general mental statism, and then further still. But I think it too hasty to conclude that this pressure cannot be resisted. On this point, I think, the verdict may still be out. At any rate, even if the hedonist must eventually succumb to this pressure, the position that emerges need not be the unrestricted desire theory. If the hedonist should start as a proponent of an objective theory, then the position that eventually emerges might well be objective as well, but with a richer and longer list of objective goods.

## III. RESTRICTING THE LIMITS

The second half of the dialectic argued that if we move to an unrestricted desire theory, we have moved too far. For there are desires the satisfaction of which intuitively contributes nothing at all toward my well-being. Recall the example of the stranger on the train. Even though I wish him success, I never hear of him or even think of him again, and so the question of whether or not he does succeed seems simply irrelevant to the level of my well-being. The moral seems to be that the unrestricted desire theory is hopelessly broad. The desire theorist must claim that it is only a certain subset of desires whose satisfaction can contribute to well-being. If a desire theory of well-being is to be plausible at all, it must be a suitably restricted desire theory.

Now I have been at some pains to argue that even if the hedonist must eventually succumb to the pressure to push the limits of well-being outward, this need not result in any version of the desire theory at all. So even if the argument of the second half of the dialectic is correct, and the only plausible version of a desire theory will be a restricted version, the overall significance of this conclusion might still be doubted, since it seems to have no implications for objective theories. I believe this appearance is incorrect, however, for I think that once we move beyond the striking example itself, and attempt to extract a general argument, the result is an argument relevant to objective theories as well as to desire theories.

Why is it that it seems so clear that the success of the stranger does not contribute to my well-being, even though his success satisfies one of my desires? The answer does not seem to turn on the fact that by the time the stranger succeeds, my desire for his success has faded and been forgotten. For even if I continued to wish the stranger success, so long as I did nothing about it and never heard of his success, it *still* seems as though his success contributes nothing at all to my well-being. Why not?

The irresistible answer is that the stranger's success has no effect on *me*. I remain exactly as I was before. I am not altered at all by the fact of the stranger's success. But if something is to make a difference to my level of well-being it must make a difference to *me*. The facts that constitute my being well-off must be facts about me.

Presumably your level of well-being is not a free-floating fact about you; it supervenes on various natural facts. You are well-off by virtue of the fact that the relevant natural facts obtain. But individual well-being is a state of the individual person. So it seems plausible to insist that differences in individual well-being must supervene on things that constitute differences in the individual person.

The point can be summarized this way: changes in well-being must involve changes in the person. It is because the stranger's success does not

involve a change in me, that it cannot involve a change in my level of well-being.

This still leaves open the issue of what exactly does constitute a change in the person. One could, of course, decline to tackle this question. We might simply concede the point that the desire theory must be restricted to desires concerning the state of the person, without attempting to further specify what states are indeed states of the person.

But at this point the mental statism may well want to seize the initiative. Boldly insisting that persons are simply the right sort of collections of mental states—that is, minds—the mental statism is now in a position to argue for the necessity of a mental state theory of well-being.

The argument she puts forward has two premises:

- (1) Changes in well-being must involve changes in the person.
- (2) A person simply is a collection of mental states.

These two premises seem to yield the desired conclusion that changes in well-being of a person must involve changes in the mental states of that person. Nothing can make a difference to my well-being that does not make a difference to my mental states. My being well-off just is a matter of having the right mental states.

What are we to make of this argument? When I first began to think about it, I thought it fairly clear that the problem was with the second premise. No doubt we all have our moments in which we are drawn to the thought that we are indeed simply our minds. But typically this is not our considered view of the matter. Generally, we are quite prepared to insist that we have bodies as well as minds. Surely no one of naturalist or physicalist inclinations should want to endorse (2).

Perhaps I am too hasty in thinking this. It has been suggested to me that certain contemporary theories of personal identity—namely those that appeal to a psychological criterion of identity—imply that persons are minds.<sup>6</sup> Since these theories are available to physicalists, the second premise, or at least something very much like it, may well be a part of a robust physicalist point of view.

I am not yet convinced that this is so. But it now seems to me that the issue is something of a red herring. The really central question is whether the *first* premise is to be accepted. To see this, suppose that we agree to modify the second premise. Whatever it is that people are, presumably a given person is nothing more than a body and a mind. So suppose we insert a suitable revision of the second premise, keeping the first as it is:

<sup>6</sup> The suggestion was made by Tim Snow, with whom I have discussed this argument with benefit; he has in mind theories like that put forward by Parfit in Part III of *Reasons and Persons*.

- (1) Changes in well-being must involve changes in the person.
- (2') A person simply is a body and a mind.

These two premises seem to yield the conclusion that changes in a person's well-being must involve changes in either the person's body or his mind.

This obviously will not satisfy the mental statist, since it allows for the possibility that some changes in my body that make no difference to my mental states might nonetheless affect my level of well-being. But the conclusion seems a strong and surprising one nonetheless. I find myself strongly inclined to think it must be in error. Yet the second premise is now unexceptionable,<sup>7</sup> so the question is whether the first is correct. (This assumes, of course, that the conclusion does follow from the premises.)

This much at least seems clear. If the first premise is true, it is not as an instantiation of the following generalization: Changes in the value of an object must involve changes in that object. For a simple example seems to suffice to disprove this more general claim. A car is presumably just a (properly organized) collection of metal, rubber, and so forth. So one might put forward the thesis that changes in the value of the car must involve changes in the metal, or the rubber, or what have you. But this is obviously incorrect. For example, if other cars of the same model are destroyed, the original car may become rare, and hence more valuable. But there will have been no changes in the metal or rubber.

No doubt one should hesitate before saying this. There are changes and there are changes. In some sense, the metal and the rubber have changed, since they have changed some of their relational properties—in particular, properties having to do with their coexistence in a world with other cars of the same model. One could preserve the general claim that change in value of an object must involve changes in that object if one were prepared to count such merely relational changes.

But the mental statist or the defender of our modified argument had better not be prepared to count such merely relational changes as potentially sufficient to ground a change in well-being. After all, when the stranger on the train later succeeds, this too alters my relational properties. So if something like the argument we are evaluating is to explain our intuition that my well-being cannot be affected by the stranger's success, the first premise must be understood to exclude merely relational changes. It is only nonrelational, or intrinsic, changes that are to count.

<sup>7</sup> In contrast, it is not nearly as plausible to assert that a person's *life* is comprised solely of facts about that person's body and mind. This raises the intriguing—and generally overlooked—possibility that it might be one thing for a person to be well-off and quite another for that person's life to go well. Unfortunately, I cannot explore this fascinating question here; in this essay, I am considering only the nature of the individual *person's* well-being.

Understood in this way, then, the general claim is false. The car undergoes changes in its merely relational properties, but none in its intrinsic properties. Yet its value does increase. So not all changes in the value of an object must involve (intrinsic) changes in that object.

What this means is that if the first premise is true, this must be because of something about the particular type of value that well-being is. Not all values depend solely on intrinsic properties, but some class of values does, and well-being is a member of this class.

Further reflection on the car example may seem to indicate how this more narrow class of values is to be identified. When the value of the car varies, it is its economic or market value that changes. But economic value is clearly an example of *instrumental* value; what varies is the car's usefulness in acquiring various other goods. And it should hardly surprise us that instrumental value will depend in part on merely relational properties.

But even if well-being does sometimes take on instrumental value, this clearly does not exhaust its significance, nor is this even where its primary significance lies. Well-being has intrinsic value; we desire it for its own sake. When my level of well-being varies, something of intrinsic value varies. And unlike instrumental value, intrinsic value can plausibly be thought to be something that must depend solely upon intrinsic properties.

So the claim that changes in well-being must involve intrinsic changes in the person can be defended as an instance of the more general thesis that changes in the intrinsic value of something must involve changes in the intrinsic properties of that thing.

Is this more general thesis true? It has about it the air of an analytic truth. Surely—one might think—intrinsic value must depend solely upon intrinsic properties.

Yet I think this appearance of being a truism is illusory. I think it trades upon two different senses of 'intrinsic value'.<sup>8</sup> One concept of intrinsic value is that of the value an object has independently of all other objects. It is the value an object has "in itself"—the value it would have even if it were the only thing existing in the universe. If anything does indeed have intrinsic value in this sense, then it seems clear that such intrinsic value must depend solely upon the intrinsic properties of the object. After all, since its intrinsic value is had independently of all other objects, it cannot depend upon its various relational properties; so its intrinsic value must depend solely upon its intrinsic properties.

But this first sense of 'intrinsic value' should not be confused with a second sense, according to which an object has intrinsic value when it is desired (or deserves to be desired) "for its own sake." Typically one goes

<sup>8</sup> Compare Christine Korsgaard, "Two Distinctions in Goodness," *Philosophical Review*, vol. 92 (1983), pp. 169-95.

on to contrast intrinsic value in this sense with instrumental value. Although I think this further contrast problematic, its familiarity should help fix this second sense of 'intrinsic value'. In this second sense, an object has intrinsic value to the extent that it is valuable as an end.

On the face of it, there is no reason to assume that what has intrinsic value in this second sense—value as an end—must have this value solely by virtue of its intrinsic properties. What is valued for its own sake might well be valuable in part because of various relational properties. Or so it seems. One might think, for example, that something's uniqueness contributes to its intrinsic value. Yet uniqueness is clearly a relational property; it is not a property that an object has independently of whatever else exists in the world. So here we have at least one possible view according to which something's intrinsic value in the second sense is not a matter of its intrinsic value in the first sense.

More generally, one might consistently hold that absolutely nothing has intrinsic value in the first sense, while still insisting that many things have intrinsic value in the second sense. One might, for example, hold a radically subjectivist conception of value, according to which nothing would be valuable as an end in the absence of there being some creature who values it. Given that there *are* creatures who value some objects as ends, some things do have intrinsic value in the second sense, but since being valuable in this way depends upon the relational property of being valued, nothing has intrinsic value in the first sense.

Thus, as a matter of logic, at least, being intrinsically valuable in the second sense does not entail being intrinsically valuable in the first sense. These points are important for the argument we are evaluating in the following way. We were looking for reason to believe that well-being must depend solely upon the intrinsic properties of the person. This seemed plausible in light of the thought that well-being is intrinsically valuable, and the further thought that intrinsic value must depend solely upon intrinsic properties. The difficulty is that the first of these two thoughts seems obviously correct only if it is the second sort of intrinsic value that we are ascribing to well-being, while the second of these two thoughts seems obviously correct only if it is the first sort of intrinsic value that is at issue.

I am certainly prepared to grant that well-being has intrinsic value as an end. But this does not suffice to show that this value depends solely upon the person's intrinsic properties, since intrinsic value in the second sense does not entail intrinsic value in the first sense. What is still needed, then, is some reason to believe that well-being is an intrinsic value in the *first* of our two senses.

Now one might try to argue that all value as an end must be a matter of intrinsic value in the first sense—value "in itself." As we have seen, this would have to be a substantive claim about the nature of value, rather than some trivial entailment. But if this general claim could be defended, then we would indeed be in a position to conclude that well-being must



depend solely upon intrinsic properties of the person. However, once the distinction between the two senses of 'intrinsic value' is kept firmly in mind, I see no good reason to think that this general claim is true.

This still leaves the possibility of arguing with regard to well-being in particular that it is intrinsically valuable in the first sense. But now it becomes increasingly difficult to see just how this particular claim is to be defended. Admittedly, if well-being is an intrinsic value in the first of our senses, then it must depend solely upon the person's intrinsic properties. But why should we believe that it is, in fact, intrinsically valuable in this sense? The assertion amounts to little more than a begging of the question.

Given that I am hostile to the conclusion of the argument—that is, to the claim that changes in well-being must involve changes in the body or the mind of the person—I wish the matter could simply be left here. But it now seems to me that there is a further move that must be considered.

A theory of well-being attempts to specify in general terms the set of facts that comprise the good for the individual. An adequate theory of well-being would have to meet several conditions. One condition—one that we have, in effect, been exploring—is that the specified facts must be about the person. We might call this the *content* condition. A second condition—the *value* condition—is that there must be a plausible account of why it is good that the specified facts obtain. But there is a third condition—the *benefit* condition—that must be met as well: the specified facts must be good *for* the person who is well-off; the well-off individual must *benefit* from being well-off.

Meeting the first two conditions need not guarantee meeting the third as well. For all that the first two conditions guarantee is that the relevant facts are good *concerning* the given individual; they do not guarantee that they are good *for* him. For example, suppose one held a retributivist view of desert, according to which it is a good thing (other things being equal) if the wicked suffer. Here is Abdul, an unrepentantly evil individual, being forced to undergo misery and pain as punishment for past crimes. If one holds a sufficiently extreme version of the desert view, then one will claim that it is a good thing that Abdul suffers. The fact of Abdul's suffering surely constitutes a fact about Abdul, and on the view being considered it is a good thing that this fact obtains. So both the content and the value condition appear to be met. But for all that, one might well want to insist that Abdul is not well-off: however well deserved his suffering may be, it is not good for *him*.

The point, of course, is not that I want to endorse this view of desert, but that an adequate account of well-being will have to meet the third condition as well as the first two: the specified facts must be good for the person, not merely good concerning him.

Consider then one further claim: What is of benefit to a person must involve the person's intrinsic properties. Now if we include merely instrumental benefits, this is certainly false. But well-being is an intrinsic ben-

efit; it is the payoff itself, and not a mere means to the payoff. So let us restrict our attention to final, or ultimate, benefits, and understand the claim accordingly: If something constitutes an (ultimate) benefit to a person, it must involve the person's intrinsic properties.

If this claim is correct, then the argument we have been considering goes through. Increasing well-being is providing an intrinsic, ultimate benefit to the person; thus, it would have to involve altering the person's intrinsic properties. Since a person just is a body and a mind, changes in well-being would have to involve changes in the person's body or mind.

But is the claim correct? Much to my dismay, I find myself strongly inclined to think that it *is*. If something is to be of genuine (ultimate) benefit to a person, then it must *affect* the person; it must make a difference *in* the person. That is, it must affect the person's intrinsic properties. Changes in merely relational properties cannot be what is of ultimate value *for* the person.

I certainly have no argument for this claim. It is simply that I find it overwhelmingly plausible. What benefits the person must make some intrinsic difference in the person. Otherwise there would be nothing in it *for him*.

Now it is easy enough to reject this claim on the grounds that it must be false or else we are led to the undesirable conclusion that well-being can only be affected by intrinsic changes in the body or the mind. But that doesn't make it any easier for me to maintain the dismissal of the claim when it is considered in its own right.

Similarly, one might point to some intuitively plausible example where one is strongly disposed to claim that something affects well-being even though it does not affect the person's body or mind. Take your favorite example, say, someone who has been deceived into thinking that she is loved and successful. Surely—I want to say—this person is not as well-off as she would have been had she genuinely been loved and successful. So it must be false that what benefits someone must affect her intrinsically.

But when I reconsider the claim once more I find myself unable to maintain this rejection. How could something be of genuine benefit *to* the person, if it never "touches" her, if it never alters the person at all?

At best, then, I find myself with a set of mutually inconsistent beliefs. (1) What benefits someone must affect her intrinsically. (2) Were the person genuinely loved rather than deceived she would be better off. (3) The deception does not affect the person intrinsically. Presumably, the set can be rendered consistent by rejecting any one of the members. But that gives us no guidance in choosing which member to reject.

In fact, however, the situation in my own case is not nearly so symmetrical. I wish I could comfortably reject the claim that what benefits someone must affect her intrinsically. But before I can do this, I need some sort of account that would precisely locate and diagnose the error in the

thought that benefit must make an intrinsic difference—as opposed to merely rejecting this thought despite its intuitive appeal. In the absence of such an account, I currently find myself more inclined to think that it is the standard examples that are mistaken. They appear to be cases where well-being is affected, but they must not be, for they cannot be.<sup>9</sup>

As I hope I have made clear, I am not at all happy with this result. I remain open to a persuasive diagnosis of the error. But for the time being, it seems to me that the argument for severely restricting the limits of well-being may well be sound. The limits of well-being may be the limits of the person.

#### IV. CLASSIFICATORY CONCERNS

Suppose the argument of the last section were correct. This would not yet establish the truth of mental statism. For the argument would only establish that well-being must be a matter of the appropriate intrinsic states of the body or the mind. As I have already noted, this leaves open the possibility that certain changes in the body might affect well-being even though they involve no changes in one's mental states.

Of course, if it could be shown that a person is indeed just a mind, then the argument would support a mental state theory. But I myself remain convinced that people are bodies as well as minds.

However, this does not yet close off all hope for the mental statist. Even if people are bodies as well as minds, all that the argument shows is the *possibility* that certain nonmental bodily changes might affect well-being. It is still open to the mental statist to argue that, in fact, no nonmental bodily change *does* affect well-being. And this does not seem an altogether implausible claim.

Nonetheless, at this point the possibility still remains that there are certain goods of well-being that are not a matter of one's mental states. Perhaps, for example, certain dispositions or abilities are themselves partly constitutive of well-being independently of their effect on one's mental states. I cannot here explore this possibility.

But one thing that does seem to emerge fairly clearly is the inadequacy of dividing theories of well-being into mental state, desire, or objective theories. This gets things wrong for three reasons.

First, it places mental state theories on the side, as though they were removed from the choice between desire and objective theories. We saw that this was mistaken when we realized that even hedonism could be given either an objective interpretation or a desire theory interpretation. It seems to me more illuminating to say that theories of well-being divide

<sup>9</sup> Alvin Goldman has helpfully characterized my argument as being: "Yes, yes, I've heard that example before."

on the question of the source of the value of well-being. There are subjective theories (most saliently, desire theories) and there are objective theories. And this choice between subjective and objective is one that arises within mental state theories as well as within a second cluster of theories that draw the limits of well-being more broadly.

The second shortcoming is that the trichotomy—mental state, desire, objective—fails to provide any kind of label for this second cluster of theories, that is, theories (whether objective or subjective) that hold that well-being can involve facts that go beyond facts involving mental states. There is a distinction to be drawn that roughly corresponds to that between mental state theories and all other theories, and we need a name for it.

But this points to a third, related shortcoming, which is that to the limited extent that the trichotomy seems to recognize this second distinction at all, it mistakenly suggests that the fundamental divide is between mental state theories and all other theories. But in fact, the more fundamental distinction seems to be between theories that limit well-being to intrinsic facts about the person and theories that allow for relational facts to directly contribute to well-being as well. Mental state theories may be the most well-known or the most plausible examples of the former type of theory, but they do not exhaust the class. Perhaps this second division should be labeled as the division between intrinsic theories and relational theories (or intrinsic theories and extrinsic theories), but I myself am drawn to a slightly different set of labels: some theories restrict well-being to facts *internal* to the person; other theories allow for the direct relevance of facts *external* to the person as well.

These two distinctions—subjective/objective and internal/external—cut across each other. Thus, there are four basic types of theory, not three, and the traditional classification fails to properly demarcate any of them.

Rather than pursue these classificatory concerns any further, let me turn to one final point. Suppose the argument of the previous section were correct. Then a variety of “external” goods that are often taken to be constitutive of well-being are actually irrelevant to it. I am distressed to think that this may be so, but for the time being at least I find myself pulled in this direction. But even if this is so, this does not at all show that the various external goods are not genuine goods. Indeed, they may well be more significant than well-being itself. We will still care deeply about the presence of these external, relational goods. And nothing suggests that we are mistaken to do so.<sup>10</sup>

Even if they are external to well-being, there will remain an important sense in which these external goods are *personal*, in that their value lies

<sup>10</sup> Having a life that goes well may plausibly turn out to be one of these goods (or a function of them)—if the quality of my life is indeed distinct from my level of well-being (see note 7).

in their relation to the given person. From a moral point of view, we may still have weighty reasons to promote the existence of these goods. Admittedly, promotion of these external goods may do nothing at all to benefit the person, but we may still be obliged to promote them out of *respect* for the person.

If this is right, then the importance of well-being may be less than we often take it to be. In many cases, the pursuit of the external personal goods will be far more important than the pursuit of the internal goods that happen to comprise well-being. The more narrowly we understand well-being, the more likely that this is the case.

If well-being is limited in its extent, then it may also be limited in its significance.

*Philosophy, University of Illinois at Chicago*