11

An Introduction to Ill-Being

SHELLY KAGAN

I

A person can be better or worse off. Lives can go better or worse.

How does this happen? The obvious and natural suggestion is that lives vary in terms of how much good they contain, and how much bad. That is to say, the overall level of one's well-being is a function of both the good contained within the life and the bad contained within it. Somewhat more precisely, there are elements which themselves directly contribute to a life going badly or well.[1]

By talk of "direct contribution" I mean that these are the very constituents of well-being. Life going well, or better, *consists* in the presence of the various goods; its going badly, or less well, consists in the presence of the various bads. Of course, different theories of well-being disagree concerning what the positive and negative constituents are. But I take it that any reasonable theory of well-being will include not only intrinsically positive elements but also intrinsically negative ones as well.

The point I am making is easily illustrated by thinking about hedonism. According to the hedonist, well-being consists in the presence of pleasure and the absence of pain. The more pleasure, the better the life.

Obviously enough, on a view like this, we can also say that the *less* pleasure the life contains, the *worse* it is. But the crucial point for our purposes is that when thinking about the ways in which a life can go less well, hedonism doesn't limit itself to talking about the mere absence of pleasure. On the contrary, according to hedonism not only is there a

---

[1] I have argued elsewhere that we should in fact distinguish between questions about well-being and questions about what makes a life go best. (See "Me and My Life," *Proceedings of the Aristotelian Society*, 94 (1994): 309–24.) But since the present paper is intended only as an *introduction* to ill-being, I will disregard that complication in what follows.

262   *Shelly Kagan*

positive element—pleasure—and the possible failure to have pleasure, or as much pleasure as one would have under some alternative; there is also a negative element—pain—such that the presence of this negative element directly constitutes a reduction in well-being.

To be sure, the absence of pleasure is bad. But it is bad merely in the sense that it is the lack of something (pleasure) that would be good. However, as I have just pointed out, hedonism also goes beyond this. It posits the existence of something (pain) that is bad in a more robust sense. Just as pleasure is an intrinsic good, robustly good in its own right (and not to be confused with the "mere" absence of bad), pain is an intrinsic bad, robustly bad in its own right (and not to be confused with the "mere" absence of good).

In short, hedonism contains a theory of *ill-being*. That's not a familiar term, but I use it to mark out the part of the more general theory of well-being that identifies and describes the negative elements, the robust bads—the elements that directly constitute a life's going badly (at least, in those regards).

2

I take it to be relatively uncontroversial that any adequate theory of well-being must also contain a theory of ill-being. That is to say, it isn't sufficient to merely spell out and explore the positive elements, the part of the theory that might be said to concern genuine *well*-being, narrowly construed. On the contrary, an adequate theory of well-being (broadly construed) must also spell out and explore the negative elements, the elements that directly constitute a life's going badly.

Admittedly, there have been philosophical views according to which there *are* no robust bads. There are just goods, and the possible lack of those goods. The fewer the goods, the worse the life. Perhaps, in the limiting case, a life might lack any goods whatsoever. Obviously, such a life would be an extremely bad one. But for all that, according to such theories there are no intrinsic bads, and so, in my sense of the term, there is no need for a theory of ill-being. (Or perhaps we should say, instead, that according to such views the theory of ill-being is an extremely simple one, consisting entirely in the claim that there are no intrinsic bads.)

But I take it that such theories are incredible, implausible. They seem to me to be clearly false, though I won't try to defend that claim here. Most of us, at any rate, will readily agree—indeed, insist—that we do need a theory of ill-being (or a fuller theory of ill-being), one that recognizes the existence of intrinsic bads, tells us what those bads are, and says how much weight they have (that is, indicates how much and in what ways the various bads lower one's level of well-being).

And yet it seems to me that in typical discussions of well-being, ill-being is largely neglected. It receives very little attention.

Think about the way we normally introduce our students to the leading theories of well-being. We might, for example, start with hedonism, and then move on to preference theories and objective list approaches. When discussing hedonism, of course, we do point out that on this view pain is indeed intrinsically bad, so that the more pain in a life, the worse off the person is. But then, when we turn to the main rivals to hedonism, we typically stop addressing the question of what the intrinsic bads might come to on the relevant alternatives. We restrict our attention to the positive elements—"for simplicity," as we might say.

The thought, presumably, is that once one understands what to say about the positive half of a given theory of well-being, it is all fairly straightforward how to extend the theory into the negative half. Accordingly, there is no need to explicitly spell out the various theories of ill-being; it is implied by what we have already said.

This is, I imagine, a familiar enough thought. I've certainly had it myself in the past, countless times. Or perhaps it would be more accurate to say that I would have had a thought something like this, if only I had ever bothered to give the matter any thought at all. I suspect that most of us have simply assumed or presupposed something like this idea, while rarely entertaining it all that explicitly.

But be that all as it may, it now seems to me that we are making a mistake when we neglect the theory of ill-being. I've come to appreciate the fact that it is not at all obvious how to extend the relevant theories of well-being so as to cover ill-being as well. Accordingly, we would do well to give the topic considerably more attention than it normally receives in the philosophical literature.

In saying this, I do not at all mean to suggest that what I think we would discover, if we were to do this, is that the standard theories of well-being *cannot* be adequately extended to cover ill-being. I only want

to suggest that this part of the subject is in fact interesting in its own right, and that it certainly isn't obvious what to say about it.

In short, the nature of ill-being is unduly neglected. Or so I hope to convince you. Of course, I won't attempt to undertake anything like a systematic examination of the subject. This paper will take a quick look at a few aspects of the subject, nothing more. Accordingly, you should think of this paper as an introduction to ill-being, not really a full-blown study. But I hope to say enough to convince you that a longer look would be worth the trouble. (If it didn't seem slightly unseemly to put it this way, I would be tempted to describe what I am about to offer as an *invitation* to ill-being.)

3

Let me start by returning to the hedonist's account of well-being. As I have already noted, hedonists should at least be given credit for explicitly recognizing that there are indeed intrinsic bads and not only intrinsic goods. They claim that pain is intrinsically bad, and indeed that it is the only such intrinsic bad.

Of course, not all pains are equally bad. Just as we need to take into account the intensity and duration of a given pleasure, so as to arrive at a measure of the *quantity* of pleasure that the experience contains, we also need to take into account the duration and intensity of a given *pain*. And the greater the quantity of pain contained in a given painful experience, the worse it is for a person to experience that pain.

That, at least, is the view of quantitative hedonists, who do, in this way, offer us an account of ill-being. But what about *qualitative* hedonists? Here, of course, the idea is that it doesn't suffice to take into account the mere quantity of, say, a given pleasure; one must also take the quality of the pleasure into account as well. Some pleasures are of a "higher quality," as we say, and for such high quality pleasures the contribution made to one's well-being is significantly greater than the contribution that would be made by a comparable amount of the ordinary, "lower" pleasures.

Presumably something similar holds for pains as well. That is to say, just as we need to distinguish between pleasures in terms of their quality, we need to distinguish among pains in terms of quality as well. Those pains that are more significant in terms of quality will be such

that the contribution that they make to lowering one's well-being will be significantly greater than the negative contribution that would be made by a comparable amount of ordinary pain.

But immediately, once we make this point, a question arises. Just what are we to *call* the qualitatively more significant pains, anyway? Presumably not "high quality pains"! That makes them sound positively desirable—whereas the idea, of course, is that the pains in question are ones that it is especially important to avoid. Nor, it might be noted, does it seem plausible to think that they should be called "low quality" pains. For that makes it sound as though these pains are sort of second-rate, whereas something closer to the opposite is the case. So just what is the standard philosophical name for this sort of pain? As far as I can tell, there simply is no standard name at all.

This is, of course, a relatively trivial point, but it illustrates in its small way the utter neglect of ill-being that is, I think, fairly typical of discussions of well-being. Even among the hedonists, who do at least take some steps to gesture in the direction of a theory of ill-being, that gesture is so perfunctory that we don't even have anything like a standard label for qualitatively significant pains. When discussing qualitative hedonism, it seems, we are so blithely confident that the negative part of the theory is—mutatis mutandis—just like the positive part, that we haven't even noticed that the labels that we have produced for ourselves to talk about qualitatively significant *pleasures* cannot be sensibly applied to qualitatively significant *pains* as well.

It is not obvious to me what a better set of labels might be. But it does seem reasonable to suggest that an adequate set of labels would refer to corresponding pleasures and pains in the same way. And while the term I have been using above—where I have talked of "qualitatively significant" pleasures and pains—is a somewhat clumsy one (and the implied contrasting label, "qualitatively insignificant" pleasures and pains, is even worse), it does at least have the virtue of talking about pleasure and pain in a symmetrical manner. So perhaps, until a better proposal has been made, it should be adopted.[2]

---

[2] It is possible that a better proposal has already been made. In discussion Irena Cronin suggested that perhaps we should talk about *deep* (vs. *shallow*) pleasures and pains. That's really quite lovely, though it does lose the explicit connection to the concept of quality.

266   *Shelly Kagan*

There is a further question about the qualitative hedonist's approach to ill-being that I would like to consider as well. If, as seems initially likely, there are qualitatively significant pains, what *are* they? Here too, the common failure to so much as mention ill-being is striking. We can all rattle off the usual suspects when it comes to listing plausible candidates for qualitatively significant pleasures: these are typically thought to be things like the pleasures of friendship and love, of art, music and poetry, the pleasures of creativity, philosophical contemplation and scientific insight, and other spiritual matters. Very well, what belongs on the list of qualitatively significant pains?

Here, there certainly isn't anything like a standard list. Indeed, so far as I know, there isn't even a single proposal that would come readily and instantly to mind among contemporary philosophers. But that's not to say that there aren't any plausible candidates for qualitatively significant pains. Let me briefly consider two interesting proposals.

First, it might plausibly be suggested that the *worst* possible pain is that of being tortured. However, even if this is right, that doesn't show that this is indeed a pain with greater *qualitative* significance. It might simply be that torture involves a horrific *quantity* of pain (in large part, of course, because of the overwhelming intensity of the pain). If so, then we don't yet have an example of a qualitatively significant pain. On the other hand, it seems possible to me that an appeal to the sheer quantity of pain involved in torture may not provide an adequate account of the badness of torture. It might be, instead, that if we are going to grasp the particular way in which torture lowers one's well-being we have to take into account the *social* nature of torture—the fact that another human being is deliberately trying to break you, to destroy your will, to demean and humiliate you. If, as certainly seems possible, it would be significantly worse to experience torture than to undergo, say, excruciating back pain (of comparable intensity and duration), then perhaps torture really is a qualitatively more significant form of pain after all. Indeed, we would have a very interesting symmetry here with at least some of the qualitatively more significant pleasures, insofar as many of the latter are of greater qualitative significance precisely because of the way in which they incorporate "spiritual" and social elements. If torture is indeed a qualitatively more significant pain, it may be so precisely

because of the way in which it too incorporates spiritual and social elements, albeit in a horrendously perverse and abhorrent manner.

Second, consider the grief one feels at the loss of a loved one. This too, it might plausibly be suggested, is a qualitatively significant pain—certainly the worst pain that many of us will ever experience. And if it *is* a qualitatively significant pain, then it is a striking fact that it too incorporates the social and spiritual elements that seem to be the mark of qualitatively significant pleasures. On the other hand, it seems to me that there is something rather puzzling about the suggestion that grief is a qualitatively significant pain. Such pains should be ones that we would be particularly eager to avoid. But is that really true for the grief one feels when a loved one dies? To be sure, normally you would rather not have lost your loved one in the first place. But given that you have, is it really the case that you would prefer not to experience any grief at all? That seems wrong; when you are aware of the death of someone you love, it hardly seems better for you to be indifferent to that fact. On the contrary, it seems better for you to be *pained* by the loss. So it isn't altogether clear to me whether this sort of pain actually lowers your well-being at all; perhaps your life would indeed be *worse* if you didn't feel grief in response to the loss of someone you love. And even if the grief itself does lower well-being, at the very least it is far from clear that it lowers your well-being by a greater amount—let alone a *significantly* greater amount—than would be the case with a comparable amount of ordinary pain. So perhaps grief isn't an example of a qualitatively significant pain after all.

I am not sure what to say about this example. Perhaps the questions I've just been raising about grief are better understood not so much as introducing a puzzle about how best to classify grief, but rather as revealing a more general problem with hedonism altogether (including qualitative hedonism), insofar as grief is a form of pain that doesn't easily fit into the more familiar patterns whereby pain straightforwardly lowers well-being. But however we resolve these issues, thinking in this way about possible examples of qualitatively significant pains should serve to reinforce my more general point, that ill-being is an important and interesting topic in its own right, one that theories of well-being would do well to attend to with greater care.

4

Turn next to preference or desire-based theories of well-being. The basic idea here, of course, is that one is well off to the extent that one's desires are satisfied. To be sure, preference theories may differ from one another with regard to whether *all* desires are such that their satisfaction increases well-being, or whether there is some special class of *relevant* desires. But for our purposes these differences are unimportant; what is important, rather, is the shared idea that well-being consists in the obtaining of the various objects of one's (relevant) desires. Roughly speaking, then, if I want X, and X obtains, then to that extent I am better off.

Well, that tells us about the positive elements of well-being. Well-being is increased by the satisfaction of one's desires. But what about the negative elements? What is it that *lowers* one's well-being?

One possible answer, I suppose, is this: well-being is lowered by the *frustration* of one's desires. That is, if I want X, and X does not obtain, then to that extent I am worse off.

This is, as I say, a possible answer. But it seems to me unsatisfying in an important way. According to preference theories of well-being, if I want X and X *does* obtain, then I am to that extent made better off. Obviously enough, then, if I want X, and X does *not* obtain, I fail to get the improvement in well-being that I would have gotten if X *had* obtained. But that is, after all, only the absence of a robust good. It seems to me that this doesn't yet introduce any sort of robust bad. (It is important to remember that in the relevant sense of "frustration," one's desire can be frustrated—it can be the case that the object of one's desire, X, does not obtain—without your being aware of this fact. In and of itself, the situation we are considering needn't involve any negative experience at all; there need be no *feeling* of frustration.)

After all, consider the case where I don't have any desire with regard to X at all. Then even if X does obtain, here too I will fail to get an improvement in well-being; I will fail to get the improvement to well-being that I would have gotten if I *had* wanted X. But it seems pretty clear that this does not yet introduce any sort of robust bad—it is only the absence of a robust good. Similarly, then, it seems to me that the most plausible thing to say about the situation in which I *do*

want X, and X fails to obtain, is that I have once again failed to gain the improvement in well-being that I might conceivably have gained. But that doesn't yet introduce anything that constitutes a robust bad in its own right.

Nor does it seem to me that we make any progress if we suppose, instead, that ill-being consists in the situation where what I want is that X *not* obtain, and yet, despite this, X *does* obtain. At first glance it might seem that here we have managed to introduce a robust bad, rather than the mere absence of a good. But at second glance, I think, this appearance proves to be misleading, for all we have actually done is redescribe the proposal we were just considering, where failure to have what you desire is bad for you. Suppose, after all, that I want X to not obtain: then it remains the case that there is something (namely, not X) that I want. So if I am fortunate enough to have it be the case that X does not obtain, then of course it trivially follows that my desire (for not X) is satisfied, in which case I am to that extent better off. Accordingly, if—despite my preference—X *does* obtain, then I fail to get what I wanted (namely, not X), and so there is a failure to improve my well-being in that regard. But this is still, after all, a mere failure to receive a potential boost to well-being. It doesn't yet seem to introduce any sort of intrinsic bad. There is nothing that in and of itself constitutes a reduction in well-being.

I can imagine that some friends of preference theory might be willing to bite this bullet. Perhaps we should insist that well-being simply consists in the satisfaction of one's preferences, and that this can happen to greater or lesser degrees. The more preferences satisfied (weighted, no doubt, to take into account such things as the strength of the desire), the better off one is; the fewer satisfied, the worse. But there are no intrinsic bads. There is simply the failure to achieve some of the potential goods.

That too, I suppose, is a possible position. But it strikes me as incredible, unacceptable, just as it did when we first considered a view of this sort in section 2. I take it to be clearly the case that there are indeed elements that *lower* one's well-being—there are robust, intrinsic bads—and if the preference theorist is forced to deny this, then I think the theory is rendered implausible on account of doing so.

270   *Shelly Kagan*

Here is a different proposal that might be suggested. Perhaps what the preference theory needs is to introduce a second psychological attitude, one that corresponds, in a negative way, to the positive attitude that preference theories normally describe. That is, just as there is a positive attitude—desire or preference—that we can have toward certain objects (or states of affairs), perhaps there is a quite distinct negative attitude—call it *aversion*—that we can also take toward various objects (or states of affairs). And just as preference theory holds that when I want X and X obtains (so my desire is satisfied) this *improves* my level of well-being, so too it should hold that when I have an aversion to X, and yet X obtains nonetheless (so that my aversion is frustrated) this *lowers* my level of well-being.

On a proposal like this, preference theory (which might now need a new name, so as to mark the fact that it incorporates aversion, and not merely preference) includes robust bads, and not only robust goods (or their absence). Admittedly, if I simply want X and X does not obtain, then even though my desire is frustrated this points to nothing more than the absence of a potential good. Nonetheless, on this revised view, there are indeed robust bads: for if I have an *aversion* to something, and yet that thing obtains nonetheless, then my aversion is frustrated, and that will constitute a robust bad. (On the other hand, if I simply have an aversion to something, and that thing does not in fact obtain, then although I have avoided the robust bad, this doesn't yet introduce a robust good.)

So adding "aversion theory" to their account might seem to offer a promising way for the preference theorist to extend their approach to cover ill-being. Unfortunately, it too faces a difficulty. For if preference and aversion are indeed logically distinct psychological attitudes, then as far as I can see, nothing rules out the possibility that one might have both a preference for X and an *aversion* to X—indeed both a preference and an aversion to the very same *feature* of X—at one and the same time. And if one did, then if it should turn out that X does obtain, then that very fact will simultaneously be both intrinsically good for you and intrinsically bad for you.

That seems to me implausible. (I am not prepared to say it is unacceptable; but it does seem to me implausible.) To be sure, we are used to the idea that some generally described object or state of affairs might be good for you in one way and bad for you in another. But in such cases,

I think, we normally point to different features of the object (different aspects of the state of affairs), precisely so as to be able to say that the one feature of the object is good for you, while another feature of the object is bad. What seems troubling is the idea that a *single* feature of a single object could be both intrinsically good and intrinsically bad for you simultaneously. Yet as far as I can see, nothing in the current view rules this out. Once we have added the logically distinct second attitude—aversion—at best it seems to be a merely empirical question whether we ever do happen to want something at the very same time that we also have an aversion to that very same thing.

Of course, it could turn out to be some sort of law of human psychology that no one ever does both desire and have an aversion to the very same feature of a given thing at the very same moment. That would mean that our problem won't ever, in fact, arise. But that seems to me a less than fully satisfactory solution to the problem, for it would still be perfectly coherent to imagine an agent for whom the laws of psychology are different. There is nothing in our revision of preference theory per se that rules out the possibility that one and the same feature could directly constitute both an increase and a lowering of well-being. And that, as I say, strikes me as implausible.

Admittedly, it isn't clear to me whether this problem arises only when we introduce the second attitude, aversion. After all, what should a fan of *traditional* preference theory say about the possibility of a case in which someone simultaneously wants both X and not X (by virtue of the very same feature)? Won't this also be a situation in which the obtaining of X is both intrinsically good and intrinsically bad for that person?

This does seem to me to be troubling. Perhaps the preference theorist will have to content herself with the suggestion that there is something *irrational* about having a pair of preferences of this sort. To be sure, if you do have such a pair of preferences, then that will indeed be both intrinsically good for you and intrinsically bad for you. Nonetheless, a rational agent will never find herself in this situation.

Is a similar move plausible with regard to aversion and preference? Can we plausibly claim that a rational agent will never simultaneously have both a preference for X and an *aversion* to X (by virtue of the very same feature)? I don't think the answer is clear. If aversion is genuinely a distinct psychological attitude, logically *independent* of preference,

then it is considerably less obvious why a rational agent couldn't have both of these attitudes at the very same time (with regard to the very same feature).

Obviously, more needs to be said about this issue, and others may not share my own judgments about these matters; but I am tentatively inclined to think that preference theory does best by reverting to the first proposal I considered, where we say that if I want X, and X does not obtain, this is indeed robustly bad for me and directly constitutes a lowering of my well-being. This contrasts, then, with the situation in which I have no desire with regard to X at all. In the latter case, even if X does obtain, this neither raises nor lowers my well-being; it is nothing more than the absence of a possible good. But if I *do* want X to obtain, and yet it does not, this is more than the absence of a good; it is indeed a robust bad.

There is, as it happens, an additional potential problem for preference theories (whichever version we accept) that arises once we start thinking more carefully about the nature of ill-being. But we are not yet in a position to appreciate it; we'll come back to it later (in section 9).

5

Let us now consider ill-being from the perspective of objective list theories. According to such accounts there are various goods—such as love, achievement, or knowledge—with objective value, and one's well-being is directly increased when one possesses the goods in question. (On the simplest views of this sort, the objective goods are indeed simply listed, with no attempt to provide any kind of systematic or unifying account. In more ambitious versions—such as perfectionism—a more general account is provided concerning what it is that makes the objective goods valuable. But for our purposes these differences won't concern us.)

Presumably, then, on such a view, just as there are objective goods there are objective bads as well, and one's well-being is directly lowered when one possesses the given bads. What should appear on the list of bads? This is, as we shall see, not altogether obvious; and I think it fair to say that many of those drawn to objective list accounts spend rather

more time discussing the objective goods than they do discussing the objective bads.

But this is not to say that when such lists are provided they overlook the existence of objective bads altogether. Here, for example, is one such list:

The good things might include moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty. The bad things might include being betrayed, manipulated, slandered, deceived, being deprived of liberty or dignity, and enjoying either sadistic pleasure, or aesthetic pleasure in what is in fact ugly.[3]

The first thing that jumps out at me when I consider this list is that it isn't particularly systematic. That isn't necessarily any sort of objection, of course, since the author of the list offered it only as an illustration of what such a list might potentially look like, nothing more. But suppose we try to impose at least a little order on it, by pairing the various items on the list where we can, so as to show a given objective good alongside its corresponding objective bad (and vice versa). If we do try to do that, we are immediately struck by the fact that not everything on the list also has such an "opposite" explicitly listed, so we may need to provide some of the missing elements ourselves, so as to fill out the table. Furthermore, once we start doing that, it isn't always clear whether there always is an opposite, and if so, what it might be.

Here, at any rate, is one attempt to pair up the elements on the list with their opposites, explicitly naming the elements that are missing. I've placed my additions in square brackets. (Note that I am simply trying to identify the "partner" for each element that appears on the original list. I am not here trying to ask what pairs, if any, might be missing altogether, with neither element being represented in that list.)

Once you pair the various elements in this way, several related points jump out at you. One such point is this: some of the pairs would need to be restated somewhat, if they were really going to serve as opposites in the way I have in mind, that is, as corresponding goods and bads of the same sort. For example, while moral goodness

---

[3] Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), appendix I, p. 499.

Good	Bad
Moral goodness [Virtue?]	Sadistic pleasure [Moral badness? Vice?]
Rational activity	[Irrational activity?]
Developing abilities	[Stunting abilities?]
Having children	[??]
Being a good parent	[Being a bad parent?]
Knowledge	[Ignorance?]
[Being treated honestly?]	Being deceived
[Proper relations?]	Being betrayed and manipulated
Awareness of beauty	Aesthetic pleasure in the ugly
[??]	Being slandered
[Being free?]	Lacking liberty
[Having dignity?]	Lacking dignity

and sadistic pleasure are presumably opposites of some sort, it seems plausible to think that sadistic pleasure is really only an instance of the relevant more general kind, moral badness or vice. Perhaps then the opposing pairs are really moral goodness (or virtue) and moral badness (or vice). Similarly, while it seems clear that "awareness of beauty" and "aesthetic pleasure in the ugly" are in some sense relevant opposites, as stated they aren't quite parallel, since the former doesn't explicitly involve taking *aesthetic pleasure* in the beauty of which one is aware, while the latter does explicitly involve such aesthetic pleasure. (But if we do revise the element involving beauty so as to explicitly require aesthetic pleasure, this in turn raises a further question: is it indeed the case that the opposite of aesthetic pleasure in the beautiful is aesthetic *pleasure* in the ugly? Or might the opposite instead be aesthetic *pain* in the ugly? This is a fascinating question, but I won't pursue it here; this is, after all, only an *introduction* to ill-being.)

A second point, already mentioned, is that at least some of the elements in the table don't seem to have any opposites at all. Take, for example, having children. Supposing that this is indeed an objective good, what might its opposite be? Is it, say, lacking children? But that seems to be an extremely clear example of the mere absence of a robust good; it isn't something that's a genuine intrinsic bad in its own right. Yet if lacking children is not the opposite (in our sense) of having children, what else could possibly play that role? It is hard to think of any

plausible candidate. So should we conclude, accordingly, that having children is an objective good with no corresponding objective bad?

I do think that this is a possibility that in principle, at least, objective list theorists should be prepared to countenance. For we should not assume, in the absence of further argument, that every objective good or every objective bad *has* an opposite.

To be sure, the thought that each element on the objective list does have an opposite—call this the *symmetry* thesis—is an attractive one. It might well be true. But we shouldn't simply assume its truth, especially without recognizing that it is indeed a substantive assumption.

In my own case, I suspect that at least some of the pull toward the symmetry thesis is due to the fact that when thinking about theories of well-being, I typically start (as I have here) with hedonism. For something like the symmetry thesis certainly does seem to hold as far as hedonism is concerned: there is a single objective good, pleasure, and it does indeed have an opposite—a corresponding objective bad—pain. I suspect that in the past I have unwittingly generalized from this theory, blindly assuming (without even realizing that I was doing this) that all objective list theories of well-being are like hedonism in this regard, with each objective good or objective bad having an opposite. Conceivably, then, there is no reason at all to accept the symmetry thesis, and it is a sheer accident, nothing more, that it happens to hold in the case of hedonism.

If the symmetry thesis isn't true, then of course we should not be troubled by the thought that the objective good we are discussing, having children, lacks an opposite. But that is not to say that this is, indeed, the right thing to say about having children. For on reflection it also seems plausible to suggest that having children in and of itself really isn't an objective good at all! On the contrary, it seems likely that what is envisioned here is something more like having children with whom one has appropriately loving and respectful relationships. Things would be quite different if, instead, one's children were rude, inconsiderate, and self-absorbed. So perhaps there is a genuine pair here after all: having children, where one stands in desirable and appropriate relations with those children, and having children where one stands in undesirable and inappropriate relations with them. Of course, once we make this suggestion it seems likely that both of these might in turn be subsumed under even more general categories, where

Shelly Kagan

there is, perhaps, the broad objective good of standing in any of a variety of loving and respectful relationships, and the broad objective bad of standing in any of a variety of disrespectful relationships, relationships marred by animosity and ill will. However, since it is not our goal here to try to find the most general categories in an objective list theory, I will put this last point aside. (Alternatively, when having children is an objective good, perhaps this is partly a matter of having *accomplished* something, namely, having raised a loving, happy, conscientious and autonomous human being. Perhaps, then, the relevant opposite is having children but failing to help them to become such people.)

Similar questions might be raised about another one of the items on our list: being slandered. Does this really have an opposite? If so, what is it? Being spoken truly of? But that simply doesn't seem to be robustly good in the way that being slandered does seem to be robustly bad. Is this then a counterexample to the symmetry thesis? Or have we simply failed to properly identify the relevant opposite? (Might it be, perhaps, being accurately acclaimed?)

6

Let's consider another item on the list in greater detail. Knowledge routinely shows up in objective list theories as one of the objective goods. Let us suppose that it does indeed belong there. If we also suppose that, whether or not the symmetry thesis holds more generally, in this case, at least, there is in fact an opposite—what is it? What exactly is the objective bad that corresponds to the objective good of knowledge? (Of course, a reasonable objective list theory will not treat all instances of knowledge as having equal value. But this complication needn't concern us.)

I suppose that the most natural suggestion to make here is that the opposite of knowledge is ignorance. That's what I wrote in the table above. But a moment's further reflection suggests that this probably isn't right. After all, ignorance is a rather broad category, and it includes the case where you simply have no belief whatsoever about the relevant subject matter. And while it certainly does seem as though ignorance of this sort involves the absence of the objective good (knowledge) that you might conceivably have possessed, it doesn't seem to be any kind

of robust bad. So what, then, is the objective bad that is the opposite of the objective good of knowledge?

Obviously enough, if knowledge is an objective good, then the corresponding bad is going to be some state in which one lacks knowledge (though not, it seems, ignorance per se). So what we need to do is to consider all the various ways in which one can *fail* to have knowledge, and then ask ourselves which of them is objectively bad. And if it should turn out that more than *one* of the ways in which we can lack knowledge is objectively bad, we will also need to ask which of them is, other things being equal, the worst. For if we assume, as seems plausible, that knowledge is the best epistemic state to be in (that is to say, it is the epistemic state which most contributes, directly, to enhancing one's well-being) then it seems that the *worst* epistemic state to be in, whatever that turns out to be (that is to say, the epistemic state which most contributes, directly, to *lowering* one's well-being) will be the relevant opposite.

(Might there be more than one epistemic state that is maximally bad? Certainly at this point in the investigation I don't see any good reason to rule this possibility out. As we will see in a moment, there are many ways in which one can fail to have knowledge. These alternatives will differ from one another in various ways, but there is no obvious reason to assume that exactly one of these ways will be worse than all of the others. Perhaps then there are rival epistemic states such that, other things being equal, they are equally and maximally bad. For simplicity, I won't keep mentioning this possibility, but I do want to note how it presents a somewhat different challenge to the symmetry thesis. Perhaps some goods are such that they have more than one "opposite." If so, that would make for a quite interesting asymmetry. And of course the same might turn out to be true for some objective bads. That is, perhaps there are *bads* that have more than one opposite.)

Now in thinking about this question, I am going to make a simplifying assumption. I am going to pretend that the classical analysis of knowledge is correct, and that knowledge is nothing more than justified true belief. That is pretty unlikely to be so. But as we will see, even if we do help ourselves to this simplifying assumption, determining the opposite of knowledge is a rather surprisingly complicated affair. So this will suffice to establish my point, that it is no straightforward matter to appropriately fill in the account of ill-being. (A more complicated

analysis of knowledge would result in an even more complicated search for the opposite of knowledge, and things are sufficiently complicated even with the simplifying assumption.)

Very well. We want to consider all of the various ways in which one might fail to have knowledge. How many ways does that come to? Consider a given proposition, P. Simplifying somewhat, it seems that we might believe P, disbelieve P, or have no opinion with regard to P at all. So there are three possible belief states. And P itself might be true, or it might be false. So that gives us two possible truth values. Finally—and again, I am obviously simplifying here—the evidence available to us might be such that we would be justified in believing P, justified in disbelieving P (that is, justified in believing not P), or it might be the case that there is no overall justification for having a belief one way or the other with regard to P. So that gives us three possible justificatory states.

Three possible belief states times two possible truth values times three possible justificatory states gives 18 possible epistemic states one might find oneself in with regard to P. One of these is knowledge. That's the state in which I believe P, am justified in believing P, and P is true. That leaves 17 alternative states, states in which something goes wrong and I fail to have knowledge. And our job is to decide which of these is the objectively bad state, the state that is the opposite (as far as well-being is concerned) of the objective good of knowledge.

Happily, we can simplify a bit more. Disbelieving P is the same as believing P to be false, which for our purposes we can take to be the same as believing that not-P. So for our purposes disbelief is just a particular case of belief (for the relevant proposition, not-P). So we don't really need to consider three different possible belief states. Rather there are only two: believing the proposition in question, and failing to have any belief with regard to it. That means that instead of having 18 possible epistemic states, there are only 12 (two possible belief states times two possible truth values times three possible justificatory states). One of these is knowledge. That leaves 11 alternative states in which I fail to have knowledge—and we have to decide which of these is the *opposite* of knowledge.

But of course by now it will be obvious that there is a further question we would like answered as well: what should we say about the *remaining* ten states? By hypothesis, none of them are knowledge, and so none of them are as *valuable* as knowledge. But saying this doesn't tell us a whole lot. Supposing that these remaining ten states are somehow intermediate

in value between the best of the dozen possible epistemic states (knowledge) and the worst (whichever that turns out to be), can we say anything about how they compare to one another? Can we rank them? (Can we at least provide a *partial* ranking?) And even if we can order all of the various alternatives, what should we say about their value in absolute terms? Are some or all of them objectively good? Are some or all of them objectively bad? Might some be objectively neutral—neither good nor bad?

It seems to me likely that even if all of the intermediate ten states are less good than knowledge itself, it is probably still the case for at least some of them that they are, nonetheless, objective goods. But *which* states are the good ones? That's just not obvious. Similarly, it seems to me likely that even if none of the intermediate states are quite as objectively bad as the particular state (whichever it turns out to be!) which is the opposite of knowledge, it still might be the case for at least some of them that they are, nonetheless, objective bads. But which states are the bad ones? The answer isn't remotely clear.

These are, I think, questions of a sort that you might easily overlook if you follow the usual practice and neglect direct consideration of the nature of ill-being. It is, after all, easy to assert, blithely, that knowledge is an objective good. But unless you are also seriously thinking about the nature of *ill-being*, you might not even notice the fact that it is not at all obvious just what the bad is that corresponds to the good of knowledge. And you might also easily overlook the fact that there will be many additional epistemic states—intermediate between knowledge and its opposite—and that they too will need to be ranked and classified as good or bad.
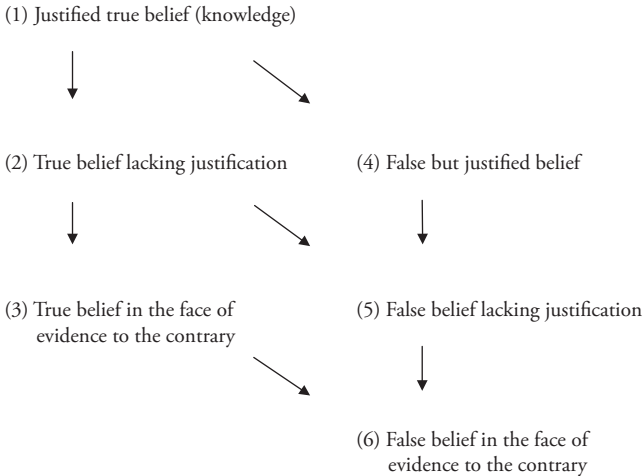
7

Although I don't think it at all obvious how the various epistemic states compare to one another, I do think it likely that if we gave the matter further thought we could probably make some progress on this front. Let me tentatively take a few steps in that direction.

Suppose we start by considering just those cases in which you have a belief. The following two principles seem to me plausible. First, other things being equal, if you believe a proposition it is better to be justified in your belief than to lack justification one way or the other; and it is better to simply lack justification for your belief (one way or the other) than

280 *Shelly Kagan*

to believe in the face of overall evidence to the contrary (to be justified in rejecting the proposition). Second, other things being equal, if you believe a proposition it is better for your belief to be true than for it to be false.

These two principles suggest the following partial ranking (where downward arrows point to inferior states):

(1) Justified true belief (knowledge)

(2) True belief lacking justification   (4) False but justified belief

(3) True belief in the face of   (5) False belief lacking justification
    evidence to the contrary

                                           (6) False belief in the face of
                                               evidence to the contrary

Obviously, this is still rather limited. On the one hand, all I have done is produce a partial ranking, rather than a complete one. (Note, in particular, that the diagram says nothing about how (2) and (3) compare to (4), or how (3) compares to (5).) And on the other hand, even that partial ranking is limited to the subset of epistemic states in which one believes that P. We also would need to produce a ranking (perhaps a partial ranking) of the *remaining* states, states in which one has no belief one way or the other with regard to P. And then we would need to ask how the elements in these two rankings compare to one another. And we would also need to ask of each of these elements whether being in this state is an objective good, an objective bad, or whether, perhaps, it is neither.

I certainly don't think I can get very far in this project right now, though I am tentatively inclined to believe that there is indeed a uniquely worst state in our set of epistemic states, namely, the state in which you believe P even though P is false and you are believing in the

face of overall evidence to the contrary.[4] What's more, it seems to me not merely the case that this is the worst epistemic state to be in, it also seems to me to be the case that being in this state is indeed bad for you. At the very least, it seems to me likely that this is a claim that those attracted to objective list theories will find attractive.

If that is right,[5] then we will at last have identified the objective bad which is the opposite of knowledge. Just as your well-being is directly enhanced when you know something, your well-being is directly *lowered* when you believe something false in the face of evidence to the contrary. Corresponding to the objective good of knowledge is the objective bad of. . . well, of what? What is the *name* for this state? As far as I can tell, there isn't anything like a simple, standard name for it, and that very fact is itself a striking one. We are so used to neglecting the theory of ill-being that as far as I can see we haven't even developed a standard label for the opposite of knowledge.

<div align="center">8</div>

I hope it is clear that the sort of complications I have been discussing with regard to knowledge are not at all unique to that particular objective good. On the contrary, it seems to me quite likely that something similar would be found if we were to look with greater care at most of the objective goods or bads that are going to plausibly appear on our list.

For what causes the discussion of knowledge to be so complicated is the fact that knowledge is what we might call a *structured* good. There are several conditions, all of which have to be in place for us to have

---

[4] My views about the epistemic states in which I *don't* have a belief about P are particularly tentative, so I won't develop them here. But I find myself inclined to think that the best two of these six states are ones in which there is no overall justification for believing one way or the other and I appropriately withhold belief. The four remaining states—where there is justification for having a belief, but I lack one anyway—are worse, I think, though equal in value. Even the latter, however, are better than *having* a belief which is both false and held in the face of evidence to the contrary.

[5] It may not be! In discussion Kieran Setiya suggested a further attractive principle: other things being equal, it is better if the available evidence points towards the truth, rather than being misleading. (Think of this as a kind of epistemic good fortune.) Note that if we accept this further principle, then we cannot be so quick to conclude that (4) is better than (5), or that (5) is better than (6). This in turn threatens the tentative conclusion that (6) is indeed the worst epistemic state to be in. As I say, the topic is a vexed one, worthy of further study.

Shelly Kagan

knowledge. Obviously enough, if we fall short with regard to any one or more of these conditions then we will fail to have knowledge; and as we have seen, there are numerous ways in which this can happen (even given our simplifying assumption that there are only three basic conditions in play). This complicates the discussion of the value of knowledge in at least two ways. On the one hand, it makes it less than obvious which particular alternative to knowledge counts as the *opposite* of knowledge, the robust bad corresponding to the robust good. And on the other hand, even once the opposite of knowledge is identified, we are still left needing to say something about all of the *other* ways in which one might fall short of knowing. These too need to be compared and ranked, and we need to say which (if any) are good (though less good than knowledge) and which (if any) are bad (though less bad than knowledge's opposite).

It is, I think, reasonable to suspect that something similar will be true for many other goods or bads. They too will be structured. That certainly seems a plausible position to take with regard to moral goodness, say, or rational activity (to mention two more goods from our earlier list), or with regard to being deceived, or taking aesthetic pleasure in the ugly (two bads from that same list). In each of these cases, it is plausible to think that attaining the relevant good or bad will require the satisfaction of several conditions, and so there will be various ways in which this might fail to occur. This may make it less than obvious, in at least some of these cases, what the opposite of the relevant good or bad might be, and how the various intermediate possibilities are to be ranked and classified. In effect, providing an adequate account of the value of structured goods and bads will often prove a surprisingly complicated affair. But this very fact may easily escape our attention; or at least, it may escape our attention until we try to work out a theory of well-being that is sufficiently detailed so as to go beyond the merely superficial consideration of ill-being that is more typical.

The existence of intermediate possibilities also complicates consideration of what I earlier called the symmetry thesis. According to that thesis, you will recall, each objective good has an opposite, as does each objective bad. But when we introduced that thesis we had not yet recognized the existence of the various *intermediate* goods and bads that will also fall along a given dimension. So we need to ask whether all of

these can be paired as well, so that each intermediate good or bad has a corresponding intermediate element that is its counterpart or opposite.

In what we can now think of as the *bold* version of the symmetry thesis, we assert that *all* objective goods and bads can indeed be paired. In particular, then, each *intermediate* good or bad has a counterpart as well. (Of course, intrinsically *neutral* intermediates, if there are any, won't have opposing counterparts.) Conceivably, we should accept symmetry in this bold form. But even if we reject it, we might nonetheless want to affirm symmetry in its *modest* version. Here the symmetry thesis merely asserts that for any given dimension along which we find either objective goods or objective bads, there are always both goods *and* bads; that is, there is always at least *something* on the other side. On this more modest view, then, there needn't be a precise counterpart or opposite for each objective good or bad; but where there are goods (or bads) along a given dimension, there will also be bads (or goods) of *some* sort along that same dimension as well.

In fact, however, the issues surrounding symmetry are considerably more complicated than this. Suppose the bold symmetry thesis is true. Then each objective good or bad has an opposing counterpart. But this is merely a metaphysical claim and says nothing at all about how the values of the relevant counterparts compare. To be sure, the counterpart of an objective good will be an objective bad, while the counterpart of an objective bad will be an objective good. But will the goodness of the objective good be exactly as great as the badness of the corresponding objective bad? Or might one of them be more significant than the other?

Suppose, for example, that I am right, and the opposite of knowledge is false belief, held in the face of evidence to the contrary; and take a given proposition, P. According to objective list theories, we are supposing, knowledge that P will directly enhance your level of well-being, while the false belief that P (in the face of overall evidence that not P) will directly lower it. Either of these will affect your level of well-being; but will they do it by the same amount? Is the *size* of the effect on well-being—whether positive or negative—the same in both cases? Or might one size be larger than the other?

In principle, it seems, we will want an adequate account of well-being not only to *identify* the various elements of ill-being; we will also want it to explore whether corresponding elements—counterparts—differ in

Shelly Kagan

their significance. Note, however, that it isn't always easy to keep these metaphysical and value theoretic questions apart. For it won't always be obvious what it means to identify a given element as something else's counterpart or opposite (even when both lie along the same basic dimension), if they differ in terms of their impact on well-being. This is especially likely to be the case with structured goods and bads where, as we have seen, it can be far from clear which alternative constitutes the opposite of the relevant good or bad. And given that the various alternatives need not fall into a natural ordering at all (that is, an ordering that is constructed independently of impact on well-being), identifying counterparts for the intermediate elements may prove especially problematic.

Still, if we *can* identify counterparts, it is natural to wonder whether their potential impact on well-being is the same. So this is a question that will seem particularly important if we accept the bold symmetry thesis. On the other hand, if we *reject* bold symmetry, then the entire issue is likely to seem moot. After all, if some objective goods or bads lack exact counterparts altogether, it isn't even clear what two things we would be trying to compare when we ask whether they have similar impact on well-being. In any event, without the bold symmetry thesis, the thought that for any good or bad there must be *something* else of opposite value but with the exact same impact on well-being will be significantly less appealing, if indeed it retains any intuitive appeal at all.[6]

9

I want to return now to a point I mentioned at the end of section 4 but which I temporarily put aside. I said there that thinking about the nature of ill-being raises an additional problem for preference theories, but one that we were not yet in a position to examine. Given what we

---

[6] Several issues relevant to thinking about symmetry are discussed in Thomas Hurka's fascinating "Asymmetries in Value," *Noûs*, 44 (2010): 199–223, though Hurka isn't concerned there with well-being per se. Interestingly, Hurka suggests that whether knowledge is more significant than its opposite may depend on the *subject matter* of the proposition in question (although it should also be noted that Hurka is working with a different view about the opposite of knowledge than my own).

have since learned about structured goods and bads, however, it should now be easy to understand the further problem that I have in mind.

This problem is essentially the same regardless of whether we are considering the traditional form of preference theory—which makes do with a single attitude, preference—or the alternative version of preference theory I also discussed, where a second attitude—aversion—is introduced as well. But the details differ slightly depending on which version we are considering, so for simplicity let's focus on the traditional account. On this version, of course, if I want X and X obtains, that is intrinsically good for me; but if I want X and yet X *fails* to obtain, then this is intrinsically *bad* for me.

Suppose then that I want knowledge with regard to some matter. If I have it, all is well: my preference is satisfied, and this directly increases my level of well-being. But suppose instead that I do *not* have knowledge of the relevant issue, I fall short in one or another of the ways we have considered. If so, then according to the version of preference theory now being considered this is actually bad for me: the fact that I want something (knowledge) and fail to have it directly lowers my well-being.

But as we can now see, this is too crude an approach to take. As I have already suggested (in section 6), it seems plausible to think that at least some of the ways in which one might fail to have knowledge might yet still be good, even if they are less good than knowledge itself. Presumably, preference theories need to be able to accommodate this idea as well. That is, they need a way to recognize that even if I fail to have knowledge, I might still be in a good enough epistemic state to have something that enhances my level of well-being, rather than lowering it (even if it does not enhance well-being by as much as knowledge itself would do).

But if it is bad for me to want something and fail to have it, then under the scenario we are now considering all we can point to is something that is *bad* for me, not something that is good for me. I want knowledge and by hypothesis I don't have it. So that is simply a robust bad, not something that can be good for me at all. Accordingly, preference theory does not yet accommodate the plausible idea that even if I lack knowledge, I still might have something that enhances my level of well-being. Similarly, even with regard to the various epistemic states that, intuitively, *are* bad for me, preference theory has no obvious way

286 *Shelly Kagan*

of accommodating the thought that they are not all *equally* bad for me. In effect, the worry is that preference theory may be inept when it comes to handling the various *intermediate* epistemic states (those that fall between knowledge and its opposite). Its judgments seem too limited: everything is either pure white or pure black; there is no room for anything in the middle.

And it should be easy to see that the problem is not at all limited to the treatment of knowledge and its alternatives. A similar problem will presumably arise for other structured goods as well. If you want the good, but fail to achieve it, preference theory will insist that this is bad for you. It does not readily accommodate the plausible thought that there may be intermediate states that are *also* good for you, albeit less good than the original structured good itself would have been. Nor does it readily accommodate the further plausible thought that even if some of the intermediate possibilities are indeed bad for you, they need not all be equally bad. When dealing with structured goods, then, preference theory seems too crude. (A similar point arises, mutatis mutandis, with regard to structured *bads*. Even if one wants to avoid a structured bad, and one does in fact avoid it, not all ways of avoiding it are equally good. Indeed, some ways of avoiding it may still be bad, even if less bad than the original structured bad would have been. But it seems as though preference theory will have to judge all of these alternatives as good, and equally so. So here too, the pronouncements of preference theory seem too crude.)

In raising this difficulty for preference theory, however, I do not mean to claim that it cannot be satisfactorily resolved. As usual, my point is only that explicitly thinking about the nature of ill-being raises questions about theories of well-being that have not typically been recognized. Here we have another case in point: how should a preference theorist deal with structured goods and bads and the various intermediate possibilities to which they give rise?

I imagine that the most promising suggestion is this. The preference theorist might point out that in typical cases, at least, our preferences are not limited to those concerning the given structured good (or bad). Rather, we have additional relevant preferences, and it is by virtue of these further preferences that the various intermediates can enhance or lower well-being to differing degrees.

Consider the case of knowledge once again. In typical cases, presumably, we desire knowledge; but this is not the only epistemic condition with regard to which we have a preference. We also want our beliefs to be true, and we want our beliefs to be justified. (And in some cases, I suppose, we also want to *have* a belief with regard to the given subject matter.) That is, we want these things in their *own* right—whether or not we have knowledge.

Suppose then that I fall short of knowledge. Suppose, for example, that my belief is justified but false. Here I want knowledge and fail to have it, so that is bad for me. But it is also the case that I want my belief to be justified, and happily it is, so in that regard at least my situation is actually good for me. On the other hand, I also want my belief to be true, and it isn't, so that is yet another way in which my situation is bad. Thus my situation is a mixed bag; in some ways good for me, in some ways bad. It is less good for me than it would be if I had knowledge (where all three desires would be satisfied), but it is nonetheless better for me than if I believed falsely, in the face of evidence to the contrary (where none of the three desires would be satisfied). In short, the overall impact on my level of well-being of this particular intermediate epistemic state is itself intermediate, just as it should be.

Of course, this doesn't yet show that an approach like this will yield plausible answers across the board, generating a plausible ranking of all the various intermediate alternatives, and appropriately classifying them as good or bad. Nor does it show that this approach will yield plausible answers for other structured goods and bads. That's a matter for further investigation. My point, as usual, is that this is a question we might not have noticed at all, had we not explicitly turned our attention to ill-being.

(Suppose instead that we adopt the alternative version of preference theory, where something is bad for me just in case I have an aversion to X, and X obtains. The problem raised by structured goods and bads is essentially the same as before: if I want knowledge and (as we can suppose) I also have an aversion to the lack of knowledge, then—implausibly—all cases where I fail to have knowledge will be bad for me, and equally so. Here too, the most promising solution posits additional preferences—and now, additional aversions as well. The preference theorist can hope that these will generate appropriate intermediate values for the relevant intermediate states.)

288   *Shelly Kagan*

10

There are, of course, still other theories of well-being besides those that I have discussed here, including hybrids which incorporate elements from more than one of these more familiar accounts. It was, in fact, while thinking about one such theory[7] that I first came to recognize what a puzzling and intriguing topic ill-being really is. I found that it wasn't at all obvious how to extend the particular nonstandard theory I was examining, so as to explicitly cover ill-being as well. Puzzled by my confusion, I began thinking about the nature of ill-being more generally; and I soon realized that the topic was both difficult and significant, and—relatively speaking, at least—neglected.

Nonetheless, despite the origin of my own interest in ill-being, I won't try to discuss these more exotic theories and their possible extensions here. As I have already noted, this paper aims only to be an introduction to the topic of ill-being, not a systematic or complete survey. My goal has been a modest one: to persuade you that the subject is worthy of more careful study than it normally receives. My hope is that what I have already said is enough to help achieve that goal.

11

Most moral philosophers rush past the subject of ill-being as though it has nothing surprising and interesting to teach us. But that's a mistake. We should all stop rushing, and take another look.

---

[7] See my "Well-Being as Enjoying the Good," *Philosophical Perspectives*, 23 (2009): 253–72.