Chapter 3

Indeterminate Desert

Shelly Kagan

1. Many of us are attracted to the idea that a plausible moral theory will be sensitive to considerations of moral desert. While there is room for a great deal of disagreement concerning the details, the basic idea, I suppose, goes something like this:

People differ in terms of their moral worth. Some people are saints, others are vile sinners, and most of us presumably fall various places in between. Furthermore, by virtue of these differences in moral worth people differ in terms of what they deserve. Intuitively, the morally better individuals deserve something "better" than those who are morally worse. Of course, just what it is that the more deserving deserve more of is one of the important topics of disagreement. It might be, for example, that the morally better individuals deserve higher levels of well-being, or it might be that they deserve more praise, or love, or admiration. In any event, it matters morally whether (or to what extent) people are in fact receiving what they deserve. This is reflected by the existence of some sort of obligation (perhaps one among others) to help give people what they deserve, or to treat them as they deserve.

Call a theory that accepts this basic picture, and thus incorporates some obligation along these lines, a *desert sensitive moral theory.* The most familiar examples of desert sensitive moral theories are doubtless deontological theories; but it may be worth noting, in passing, that a consequentialist theory could be desert sensitive as well (provided that it appeals to a theory of the good according to which the goodness of an outcome depends, at least in part, on the extent to which people receive what they deserve).[1]

Much turns, obviously enough, on what it is that makes one individual more or less morally deserving than another. But a natural enough place to start, I suppose, is with the suggestion that someone's moral worth depends, at least in part, on whether the given individual acts morally or not.

The particular issue that I want to explore about such desert sensitive moral theories is the worry that they may be logically unfit, by virtue of their very sensitivity to considerations of desert. The basic worry is easy enough to see: if morality is a matter of giving people what they deserve, and what they deserve depends on their moral worth, and their moral worth depends on whether or not they have acted morally, then morality is a matter of giving "more" (of whatever the appropriate good is) to people who have acted morally; which is to say,

morality is a matter of giving more to people who have given more to people who have given more to people who have given more – and so forth, and so on. There is clearly *some* kind of regress going on here, and the question is whether or not it is a philosophically unacceptable regress.[2]

2. But what, exactly, is the problem supposed to be? Even if desert sensitive moral theories face a regress, not all regresses are vicious. Is this one? And if it is, is there a solution available to desert sensitive theories? Or does the possibility of regress here show us that desert sensitive theories must be abandoned?

Actually, it seems to me that there are several regress problems that face desert sensitive moral theories. I want to identify and distinguish two of them in particular, though only one of the two will be our concern for most of the paper.

We can get a better sense of the problem if we start with a fairly simple model of what a desert sensitive theory might look like. Suppose, first, that all individuals can be classified as either being good or bad. We then need to indicate just what it is that people deserve. A fairly natural proposal would be to suggest that good people deserve to be happy, while bad people deserve to be unhappy. (For those who reject this kind of retributivism, nothing significant would change if we stipulated instead that those who are good deserve a higher level of happiness than that deserved by those who are bad.) But to avoid irrelevant complications arising from the difficulties inherent in distributing happiness let us suppose instead that what good people deserve is to be loved (or approved of), and what bad people deserve is to be hated (or disapproved of). In what follows, I will restrict myself to theories with this feature – where the relevant form of deserved treatment is being loved or hated – but this shouldn't affect any of the essential logical points.

Next, let us assume that our moral theory contains exactly *one* moral obligation (or its equivalent), namely, an obligation to give people what they deserve.[3]

But now a further complication needs to be put aside, since it seems possible that in some cases how I treat a given person – whether I love or hate him – might have a causal influence on whether *other* individuals give people what they deserve. If so, then it might turn out that the only way to bring it about that most people receive the treatment that they do deserve is for me to treat some particular individual in a way that he doesn't deserve. It's not uncontroversial how the obligation to give people what they deserve should tell me to act in such cases. But although this may be important for a complete specification of a desert sensitive theory, such questions are unimportant for our present purposes. So let us suppose – if only for simplicity of discussion – that this sort of influence doesn't arise. That is, let us suppose that whether a given person loves or hates someone has no causal influence on other cases of loving and hating. Thus, in assessing whether someone meets the obligation to give people what they deserve, we can simply look to see whether the person himself loves those who are good and hates those who are bad.

Finally, let us suppose that someone is good if and only if he completely meets his moral obligations. Of course, since we are assuming that our moral theory includes only the single obligation to give people what they deserve, this means that someone is good if and only if he loves those who are themselves good, and hates those who are bad. If someone does this perfectly, he is himself good. And if he fails to do this perfectly, he is bad. This last implication is perhaps worth emphasizing. Presumably, someone might give some but not all individuals the love or hate that they deserve, but on the model we are beginning with anyone who fails to conform to the theory perfectly is himself classified as bad, and as such, deserving of hate.

Now obviously enough, this simple model of a desert sensitive theory has a variety of features that more plausible versions would lack. If we find that it is plagued by one or another regress problem, then we can hold out the hope that the problem is in point of fact due to one of the simplifying features, and that a more sophisticated theory would eliminate the feature, and thus avoid the problem. That's a possibility that we will be exploring. But at any rate, let's begin by demonstrating that this admittedly overly simple model does indeed face some regress problems.

3. Consider, then, the following very simple case. Suppose there is exactly one individual, Jones, and that Jones hates himself. What should we say about this case? Is Jones good or bad?

Well, if we suppose that Jones is good, then he actually deserves to be loved; and since he does not love himself it follows that he doesn't give himself what he deserves. Thus Jones does not conform to the requirements of morality, and this means that he is bad. That is, if Jones is good, he is bad.

Should we conclude, then, that Jones is bad? But if Jones is bad, then he deserves to be hated, and since he does hate himself, he is giving himself what he deserves. Furthermore, since Jones is the only person in this world, Jones is obeying morality perfectly: he is giving everyone – that is, himself! – exactly what they deserve. Since Jones conforms to morality perfectly, he is good. Thus, if Jones is bad, he is good.

Obviously this is a kind of regress problem. If we assign one "status" to Jones, then we are led, under the terms of the model, to assigning the opposite status, which in turn leads us to assigning the initial status after all, which in turn leads us to assign the opposite status yet again, and so on and so forth, interminably.

Put a slightly different way, there is no way to assign a status – that is, to classify Jones as being good or bad – in a way that is *stable.* Any given assignment undercuts itself. There are only two possible statuses – being good or being bad – yet we cannot stably assign either status to Jones.

I take it that this is indeed a problem with this initial version of a desert sensitive moral theory. It is unacceptable for a theory to insist that everyone is either good or bad, and yet to lack the very possibility, in certain cases, of assigning any stable status at all. Or so it seems to me.[4]

4. Although the problem just noted seems to me a genuine one, it is not the problem on which I want to focus. In the particular case just discussed, of course, the problem arises from the impossibility of finding a stable assignment of status.

Typically, however, this won't be especially difficult. Consider a different case. Again, let us suppose that Jones is the only individual. But this time let us suppose that Jones *loves* himself. Here it is easy to assign a stable status. We need only assume that Jones is good. If he is good, then he deserves to be loved. Since he deserves to be loved, Jones treats himself in accordance with what he deserves, and thus conforms to morality perfectly. And by virtue of his perfect obedience to morality, Jones is indeed good. Here, then, we have no inner incoherence, no interminable oscillation, no inability to stably assign a particular moral status. Assigning a stable status is, as we see, easily done.

The difficulty, rather, is that it is too easily done. It is certainly true that we can stably assign Jones the status of good. But it seems that we can just as easily assign him the status of bad.

Suppose, after all, that Jones is bad. If he is bad, then he deserves to be *hated.* And since he actually loves himself, it follows that he is not giving himself what he deserves. And this means, of course, that he is violating morality. Indeed, it is by virtue of that very violation that Jones is bad. That is, Jones is bad, and his badness is grounded in the very fact that he loves himself – in violation of morality – despite the fact that he is bad. Thus, if we assign Jones the status of bad, this is stable as well.

In short, both possible assignments – Jones as good, and Jones as bad – are *stable.* Nothing in the theory forces us to rescind the assignment – either assignment – as soon as we make it.

Very well, then, which assignment is *correct?* Which is it? Is Jones good? Or is Jones bad?

But of course to ask the question is to see that we cannot possibly answer it. There is nothing *more* in the theory that we have failed to take into account. Jones's status is a simple matter of whether or not he perfectly obeys the moral theory. But whether he does obey – given that he loves himself – is itself a simple matter of what his status is. Thus: he is good if and only if he is good, and he is bad if and only if he is bad. Of course, there is nothing wrong with a theory saying *that.* What *is* a problem is when a theory has nothing *more* to say about it, and that, in effect, seems to be the situation here.

So what is Jones's status according to our theory? I think we will have to say: there is no fact of the matter; it is indeterminate.

And this, it seems to me, is indeed a problem. It is, I suppose, a kind of regress problem as well: to settle Jones's status it must first be settled whether or not he obeys morality, but to settle that we must settle his status. There is nothing in the theory that forces a particular assignment upon us – and thus nothing in the theory to allow the attempt to determine that status to terminate. But in any event, whether or not this second problem is indeed helpfully viewed as a regress problem, the very fact of indeterminacy does seem to me problematic.

It is this second problem upon which I want to focus. And it is worth noting explicitly how this second problem differs from the first one I identified. In a sense, of course, we might say that both problems are problems of *indeterminacy,* for in neither case are we able to assign a single, determinate status. But the problems are interestingly different for all that. In the first case, after all, the problem arises from the fact that *no* stable status can be assigned. Thus, we could say, there is a "determinate" fact of the matter after all, which is that Jones's status can be *neither* good nor bad (even though it must be one or the other). I will call this the problem of *instability.* In the second case, in contrast, the problem arises from the fact that *more than one* stable status can be assigned. Thus, there is no determinate fact of the matter as to *which* status Jones has (even though it must be one or the other). It is this that I will refer to as the problem of *indeterminacy.*

And the question immediately facing us, accordingly, is this: can desert sensitive moral theories avoid this problem?

5. The first possible "solution" to the problem, I suppose, would be to deny that it is a genuine problem after all. That is, it might be suggested that indeterminacy is not a problematic feature for a moral theory to have. Obviously, if indeterminacy is not troublesome, there is no need to ask whether desert sensitive theories can avoid it.

I find this first proposal difficult to take seriously, however, for it seems to me clearly problematic for the theory to hold that there is no fact of the matter as to whether Jones is good or bad. After all, to state the obvious, Jones's status itself determines how he is to be treated. In the simple model we have started with, it determines whether he is to be loved or hated. How can it be acceptable to suggest that there is simply no fact of the matter as to which of these is appropriate? And of course, in other desert sensitive theories still other forms of treatment will be called for, depending on what particular status Jones has. How can it be unproblematic to have a situation where, say, there is simply no fact of the matter as to whether Jones is to be rewarded or punished?

It is worth bearing in mind in this regard that the problem of indeterminacy is not an epistemological one. The situation might be tolerable if we could say that there *is* a fact of the matter as to what Jones's status is, but for one or another reason we are simply unable to *tell* which it is. It would hardly surprise us to discover that under various conditions we are unable to determine how someone is to be treated. In point of fact, however, the indeterminacy we have uncovered is a metaphysical one. There is no fact of the matter concerning Jones's status, and thus no fact of the matter concerning how he is to be treated. And this, it seems to me, is unacceptable.

Nor, I think, would it be plausible to suggest that indeterminacy is not a problem if it does not arise in *actual* cases. That is, it is not plausible to suggest that indeterminacy is acceptable so long as the cases in which it occurs are merely logically possible ones, but never in fact actualized. Obviously, a proposal along these lines is not an incoherent one – it is coherent to hold that indeterminacy would be a problem if it actually occurred, but that it is acceptable so long as it remains a mere theoretical possibility. But for all that I find it implausible. It seems to me that an adequate moral theory should cover all possible cases. Just as, I am assuming, it is unacceptable for a theory to say of a given actual case that there *is* no fact of the matter how one or more individuals should be treated, so too, it

seems to me, it is unacceptable for a theory to say of a merely possible case that *were* the case actual there *would* be no fact of the matter how one or more individuals should be treated. Accordingly, I won't concern myself, in what follows, with the question of whether or not cases of indeterminacy ever actually arise. It is sufficiently problematic if they might. In short: it is unacceptable if there is – or would be – no fact of the matter concerning Jones's status and thus no fact of the matter concerning how he is to be treated.

6. But this immediately suggests a second possible solution. If it is unacceptable for there to be no fact of the matter concerning Jones's status, and if our original theory is itself compatible with either assignment, let us simply modify the theory, expanding it, so as to assign Jones a determinate status in precisely this kind of case. That is, when there is no independent ground for assigning Jones a particular status, let us add a further rule that assigns one. Presumably, which way we go on this is somewhat arbitrary, but perhaps charity might lead us to propose that if there is no independent ground for assigning Jones a particular status, he is to be assigned the status of good.

This obviously solves the indeterminacy problem, at least in the particular case we have been considering. In the case where Jones loves himself, as we know, either assignment – good or bad – is stable. Thus there has been no ground for assigning him a particular status. But now, with our new rule, we can say that, precisely by virtue of that fact, he does have a particular status, namely that of being good. And this assignment remains stable, of course: Jones is good, and thus deserves to be loved; and he does love himself, thus conforms to morality, in keeping with his goodness.

7. Unfortunately, this second solution won't work either, for there are other cases where it yields unacceptable results. To see this, however, we must first note that indeterminacy is not at all limited to cases with only a single individual. More complex cases can face the same problem. For example, consider a case in which Jones and Smith are the only two individuals, each loves himself, and each loves the other as well. Here too we must face the fact that there is more than one stable way of assigning statuses to the relevant parties.

Thus, on the one hand, we can hold that both Jones and Smith are good. Since each is good, each deserves to be loved, by himself and by others, and since each does in fact love himself and the other, each conforms perfectly to morality, and thus each is indeed good. That is to say, assigning the status of good to both Jones and Smith is a stable option.

Similarly, however, if we assign the status of *bad* to both Jones and Smith, this is stable as well. Since each is bad, each deserves to be hated, by himself and by others. And since each nonetheless loves himself as well as the other, each violates morality, and thus each is indeed bad. So here we have a second, incompatible and yet stable, assignment of status.

Now as it happens, the second proposed solution yields an acceptable position here as well. Since there is no independent ground for assigning statuses to Jones and Smith, we take both to be good; and this is, indeed, a stable solution: each is good, and appropriately loves himself as well as the other.

Obviously, however, cases involving two or more parties could easily be multiplied. And in at least some of these cases, the second proposed solution is unworkable.

Let me mention just one such case. In this variant, Jones and Smith each loves himself, but hates the other. What stable assignments are possible here?

Well, if Jones is good, then it is inappropriate for Smith to hate him, and so Smith must be bad. And this assignment is indeed stable: Jones is good, and thus appropriately loves himself, and hates Smith, who is bad; Smith is bad, and this is confirmed by the fact that he hates Jones, who is good, and loves himself, even though he is himself bad.

So the assignment in which Jones is good and Smith is bad is stable. But, of course, the opposite assignment is stable as well, that is, the assignment in which it is Smith who is good and Jones who is bad. So which is it? Is Jones good, or is Jones bad? And what about Smith? Here, too, our initial theory faces a problem of indeterminacy.

Suppose, then, that we adopt the proposal according to which when there is no independent ground for assigning status, we are to take the given individual to be good. Then it seems we must hold that Jones is good, and that Smith is good as well.

But this assignment, as is easily seen, is not in fact one of the stable ones. If Jones is good, then it is inappropriate that Smith hates him, and so it follows that Smith violates morality, and thus is bad. Yet this is incompatible with his being good. Thus the newly expanded theory tells us – inconsistently – that Smith is both good and bad. This is obviously an unacceptable result. (Similarly, of course, were we to start with the claim that Smith is good, as per the proposal under consideration, we will be led to the inconsistent position that it is Jones that is both good and bad.)

The problem, of course, is that in the particular case we are now looking at the only stable assignments under the original theory are ones in which exactly one of the two people is good and one is bad. The theory doesn't determine which is which, but it insists that exactly one has each status. Yet the proposal under discussion holds that when there are no independent grounds for assigning status to an individual, he is to be taken to be good. This, as we have seen, leads to unacceptable inconsistencies.

Perhaps then we should adopt instead a variant on the second proposal. Instead of taking everyone to be good, when there is no independent fact concerning their status, we should initially take only one such individual, assign him the status of good, see whether this assignment (in light of other facts of the situation) now fixes the status of one or more other individuals, and if there remain still others with no determinate status yet, assign one more of these others the status of good as well, and so on, repeating the process as needed until all statuses are assigned.

On this variant, instead of incompatibly taking both Jones and Smith to be good, we take only one of them, perhaps Jones, to be good, and see whether this

fixes the status of the others. And as we have seen, if Jones is good then it violates morality for Smith to hate him, and so this does indeed fix Smith's status, as bad. And this is indeed a stable assignment, since if Smith is bad, then Jones is right to hate him, as he does.

The problem, however, with this variant is the very fact that when there is more than one individual with no determinate status it is *arbitrary* which one we initially take to be good. It is an arbitrary choice, yet it can have significant implications concerning the status ultimately assigned to others.

After all, as we have just seen, if we initially pick Jones, and take him to be good, we end up holding that Smith is bad. Yet had we selected Smith instead, and taken him to be good, instead of Jones, then obviously enough Smith would have ended up good, and it is Jones that would have ended up bad. Thus, depending on our initial, utterly arbitrary choice concerning which individual to take to be good, we end up with quite different views concerning who is good and who is bad, who deserves to be hated and who loved. And it strikes me as completely implausible to suggest that whether Smith, say, should be hated or loved (rewarded or punished) is properly viewed as a matter of arbitrary choice. It seems then that we must reject the second proposed solution as well, along with, at least, its most obvious variants.

8. The possibility of cases of indeterminacy involving two or more individuals also shows why a third possible solution won't work either. If indeterminacy arose only in cases like our initial case of Jones loving himself, it might have been thought that the problem was simply due to the duty one has with regard to oneself to treat oneself appropriately. It might have been suggested, then, that we should eliminate this duty. Perhaps one has a duty to treat *others* as they deserve – loving them if they are good, hating them if they are bad – but one has no similar duty toward oneself. And it might have been thought that this would eliminate the problem of indeterminacy.[5]

In fact, however, the proposed solution is inadequate. For as we have now seen, indeterminacy can easily arise in cases that turn on the treatment of *others*. Recall the case where Jones and Smith hate each other. Even if we simply disregard as morally irrelevant the fact that each loves himself, it remains the case that there are two stable ways to assign status: it could be that Jones is good and Smith is bad, or it could be that Jones is bad and Smith is good. If we are to find a solution to the problem of indeterminacy, it won't suffice to insist that it does not matter how one treats oneself.

Instead, any successful solution along these lines would need to be far more radical. It would need to insist that it does not matter how one treats *anyone* – whether oneself or another. More precisely, of course, it would need to insist that there are no moral requirements concerning how to treat one another where the appropriate treatment is a function of their *status*. According to this more radical proposal, then, desert sensitive theories get into trouble precisely by virtue of the fact that they incorporate "status sensitive rules" regarding our treatments of one another, telling us to treat good people one way, and bad people another way. To

eliminate indeterminacy, it might then be suggested, we should eliminate all such status sensitive rules.

Now it must, I think, be admitted that eliminating status sensitive rules – all of them, not just the self-regarding ones – would eliminate the kind of indeterminacy that we have been examining. And so, in this sense, this latest proposal is indeed a *solution* to the problem. But it is not of course one that advocates of desert sensitive theories will find congenial. For it is, in effect, the proposal to *abandon* sensitivity to considerations of moral desert. Admittedly, not everyone finds the intuitions behind such sensitivity attractive. But *I* do (though I won't attempt to argue for them here). It does seem to me that people can be classified as being morally better or worse, and it does seem to me that people who differ in this regard differ in terms of the treatment that they deserve; and it seems to me, finally, that we have an obligation to give people the treatment that they deserve. But to accept all of this just is to accept the existence of status sensitive rules. And so the question for me remains whether there is a way to avoid indeterminacy while still retaining an acceptable form of desert sensitivity.

We should, however, note the following possibility. Even though eliminating self-regarding status sensitive rules cannot, by itself, provide a complete solution to the problem of indeterminacy, it might still be thought an important element in an acceptable complete solution. I don't actually believe this to be the case, but in order to avoid begging the question against this view I am going to assume in what follows that the only relevant status sensitive rules are *other*-regarding. (This also allows us to simplify the remaining examples, by disregarding each person's love or hate of himself.)

Of course, as we have seen, even if the only status sensitive rules are other-regarding, this doesn't eliminate the possibility of indeterminacy. So the question remains: can indeterminacy be avoided while remaining within the confines of a desert sensitive theory?

9. A fourth proposal suggests that we can indeed eliminate indeterminacy while retaining (other-regarding) status sensitive rules, provided that these rules are not the *only* moral requirements that an agent must obey. More precisely, indeterminacy can be avoided provided that the status sensitive rules are not the only rules relevant to determining the agent's moral status.

The thought here is a simple one: in the examples we have been considering indeterminacy arises from the fact that to be good one must conform to morality. But in our simple model the only requirements morality has put down are requirements to treat people in keeping with their moral status. Thus, each person's status necessarily depends on the status of others – and, in effect, on such facts about status alone. It is no wonder, then, that this generates indeterminacy. Status depends on status, and until status is fixed, status cannot be fixed. Thus, status cannot be fixed.

Suppose, however, that there were further, additional requirements laid down by a fuller moral theory. Then by seeing whether or not a given individual obeys these *other* requirements, we may be able to fix his status, which in turn will fix,

for example, how others are to treat him, which will allow us to go some distance toward settling the question of the extent to which those others are conforming to morality; and this – along with similar initial determinations of status for various other individuals – may allow us to arrive, eventually, at a uniquely stable assignment of statuses overall. Thus, according to this fourth proposal, introducing additional moral requirements will eliminate indeterminacy.

For our purposes, it won't much matter what the content of these other requirements comes to. It might be that there is a requirement to tell the truth, or to keep your promises, or to compensate those you have accidentally harmed. As I say, so far as I can see the details shouldn't matter, so we can simply say that in addition to the relevant status sensitive rules, there are one or more further duties, which for the moment we shall simply refer to as the duty to do X.[6]

Consider, then, the case where Jones and Smith hate one another. As we have already seen, if the only moral requirement is the obligation to treat (other) people as they deserve, then there are two stable ways to assign status (Jones good and Smith bad, or Jones bad and Smith good). But we are now assuming that there is a duty to do X as well. How do Jones and Smith fare with regard to X? Let's imagine that Smith does do X, but Jones does not. What assignments are possible now?

Since Jones does not do X, he does not perfectly conform to morality, and so he is bad. Since he is bad, he is appropriately hated by Smith, who thus meets the obligation to treat people as they deserve. What's more, Smith does X, and so meets this further moral obligation as well. Smith thus meets all of the requirements of morality, and is therefore good. Finally, given Smith's goodness, Jones's hatred of Smith is inappropriate, a further violation of morality, confirming Jones's badness. Thus one stable assignment has Jones being bad and Smith being good.

What's more, it is the only stable assignment of status. Since Jones does not do X, he does not meet all of the requirements of morality, so we must assign him the status of being bad. At the same time, we *must* assign Smith the status of good, since he perfectly conforms to morality (given that he does X and appropriately hates Jones, who is bad). In short, the *only* stable assignment of status holds that Jones is bad and Smith is good. Here, then, we have no indeterminacy.

Clearly, then, the introduction of further moral requirements can sometimes guarantee the existence of only a single stable assignment of status. Unfortunately, however, it does not suffice to eliminate indeterminacy in all cases.

Consider once again the case where Jones and Smith hate one another. But this time suppose that both Jones and Smith do X. What assignments are possible here? It is possible, of course, that Jones is good: since he does X, he at least meets that requirement, and it could be that he meets the other requirements of morality as well. Suppose that he does. Then it is inappropriate for Smith to hate him, and thus Smith is bad: despite the fact that he meets the requirement to do X, Smith does not meet the relevant status sensitive rule, so does not conform perfectly to morality. Of course, if Smith is bad it is appropriate to hate him, and Jones, appropriately, does. Thus one stable assignment has Jones being good and Smith being bad.

Obviously, however, the opposite assignment is stable as well. It could just as easily be the case that Smith is good, and Jones bad. Thus we have two stable assignments, and indeterminacy has not been eliminated, despite the introduction of further moral requirements.

I hope it is obvious that nothing at all turns on the details of these further requirements, or for that matter on the *number* of the further requirements. Even if we had various additional requirements – an obligation to do X, an obligation to do Y, and an obligation to do Z – there will still be cases where someone has fulfilled these obligations, and thus the question of their moral status cannot be settled until we know whether they have fulfilled their various status sensitive obligations as well. And this means, I take it, that the problem of indeterminacy cannot be eliminated simply through the addition of further obligations.

This last point has an interesting implication which it might be good to make explicit. As we have just seen, indeterminacy cannot be eliminated through the mere introduction of further moral requirements – X, Y, and Z – in addition to the various status sensitive rules. But this means that no matter how pluralistic our theory is – no matter how much the obligation to give people what they deserve recedes into the background, merely one obligation among many – we still face an indeterminacy problem. So long as we are dealing with a desert sensitive theory at all, we run the risk of indeterminacy.

10. A fifth proposal suggests that indeterminacy arises from the stipulation in our original model that an agent is good only if he conforms to morality *perfectly.* This stipulation is obviously an implausible oversimplification. Suppose we replace it, instead, with the recognition that one can be good even if one has *not* obeyed morality perfectly. Will this help?

It is easy to see why one might think that it would: if one needn't obey morality perfectly to be good, then perhaps even if one has violated one or more of the status sensitive rules, one could be good anyway, in which case we will not need to settle everyone's status first, since it won't necessarily matter whether one has fulfilled the status sensitive rules or not. In this way, one might hope, the problem of indeterminacy could be eliminated.

In fact, however, I think the new proposal won't suffice to eliminate indeterminacy either. To see this, suppose that we retain two possible classifications here: all agents are to be classified as either good or bad. But we jettison the claim that one can be good only if one has obeyed morality perfectly, accepting in its place a "lower" standard than the one we have just rejected. Instead of requiring an agent to obey all of morality if he is to be good, it will suffice if he meets "enough" of the moral requirements. Different versions of this proposal will presumably offer different accounts of what is to count as "enough." And no doubt much will turn on the details of the further moral requirements – X, Y, and Z – that we have added to our theory. But let us suppose that all these proposals are similar in being *precise.* That is, we are able to specify precisely how much is enough.

One can still be good while violating "this many" requirements of morality (or this many of these particular requirements), but if the agent were to violate *more* requirements than this, then that would be too much, and now we would have to classify the agent as being bad.

It seems to me that any account of this sort will still be subject to indeterminacy, given the assumption that the relevant standard for assigning status is precise. For imagine that we have a case involving two agents, Jones and Smith, both of whom have just enough violations of the various other requirements (that is, X, Y, and Z) so that the following is true: if these are the *only* violations of morality, then each is appropriately classified as good. But these violations leave each agent only just above the line, so that if he has one *more* violation, that will be "too much," and he will be classified as being bad rather than good. And now imagine that each hates the other.

If Smith is bad, then Jones does right to hate him; he meets his status sensitive obligations, and thus has no further violations, and can be properly classified as good. Smith, meanwhile, fails to meet his status sensitive obligations, since he inappropriately hates Jones, who is good; thus Smith has one violation too many, and is properly classified as bad. In short, the assignment of Jones as good and Smith as bad is a stable one.

Obviously, however, the opposite assignment – Jones as bad, and Smith as good – is stable as well. Thus we have not actually avoided the problem of indeterminacy after all. Even though we have moved to a "lower" requirement for being good, given that the rule for assigning an agent the status of good or bad is precise it seems that we will still run the possibility of indeterminacy.

Nor would it be plausible, it seems to me, to propose instead that the relevant standard for assigning status should be *imprecise.* After all, to say that there is no precise standard for assigning status is to say that in at least some cases there is no fact of the matter whether the person is good or not, and thus no fact of the matter how he is to be treated. This is simply to embrace indeterminacy anew, and to insist that it is not objectionable after all. And this, as I have already suggested, seems to me unacceptable.

11. Perhaps, however, the indeterminacy problem arises from the fact that we have been confining ourselves to only two possible classifications? Instead of insisting, as our original model does, that everyone is to be assigned one of exactly two moral statuses – good or bad – we should allow for *several* distinct statuses. For example, it might be that people are to be classified in one of seven different ways: as being extremely good, moderately good, minimally good, neutral, minimally bad, moderately bad, or extremely bad. (Obviously, even more fine-grained classificatory systems would be possible as well.) This then is a sixth proposal: introducing such fine-grained classificatory systems will allow us to eliminate indeterminacy.

Unfortunately, this too seems to me a misguided suggestion. Although it certainly seems plausible to suggest that we should have fine-grained distinctions between varying levels of moral worth, I don't think doing this helps with the problem of indeterminacy.

Of course, showing this through an example requires that we introduce a new rule for assigning a particular status, one sensitive to the fact that we now have several distinct levels of moral worth, rather than just two. But we can keep the general approach of the original model, where one's moral status is a function of the extent to which one has conformed to the various requirements of morality, including – presumably – the various status sensitive requirements to treat people in keeping with their particular (now, fine-grained) status. We will need some new status sensitive rules as well, to reflect the fact that there are now more than two levels of moral worth, since people with different levels of moral worth deserve to be treated differently. It might be, for example, that the extremely good are to be loved very much, the moderately good are to be loved somewhat less, the minimally good only slightly, and so forth.

Now consider a case in which Smith and Jones each love each other very much. But suppose, as well, that each has just enough violations of the various other moral requirements (X, Y, and Z) so that if one were to assign status on the basis of these various violations alone each would remain assigned the status of being extremely good, but each would be just above the line. That is, for each it is true that if there are no other violations, then each is indeed extremely good, but if there are any additional violations each will have "too many" violations to be extremely good, and will instead be only moderately good. So to determine their exact status we must consider whether or not they meet their status sensitive obligations.

Suppose, then, that both are indeed extremely good. Then the fact that each loves the other very much is appropriate, and so there are no further violations beyond those already noted, and so both are, as just stipulated, extremely good. This is, then, a stable assignment of status. But it is not the only stable assignment. For if we suppose, instead, that each is merely moderately good, then it is inappropriate for each to love the other very much – rather, each should only be loved a moderate amount. So each has a further violation not yet taken into account, and in light of it they cannot actually be taken to be extremely good; instead they must be assigned the status of being only moderately good. Thus there are two stable assignments: both can be extremely good, or both can be moderately good.[7] And this means, of course, that we have not escaped indeterminacy, after all, despite the introduction of multiple levels of moral worth. Adding extra levels of moral worth doesn't eliminate the problem, it just complicates the examples.

12. Nor would it help to suggest – a seventh proposal – that we relax the assumption that conformity to morality is the *only* factor relevant to determining someone's moral status. While it is certainly plausible to suggest that *several* factors may be relevant to determining status (and not only the extent to which one conforms to morality), provided that we maintain that status is indeed fixed at least in part by considerations of conformity, then we won't yet have escaped indeterminacy.

For suppose we have a case where these various other factors – those not involving conformity to moral requirements – leave the status of the concerned individuals just "above the line" for some particular level of moral worth. Thus, which precise status they have will turn on the matter of their conformity to morality. And we will then be able to imagine that the verdict of that latter factor – conformity – will hinge in turn on whether or not the various status sensitive requirements have been met. And this means that indeterminacy can still arise.

Put another way: if conformity to morality matters at all in determining one's status, and if status sensitive rules are among the rules conformity to which makes a difference, then we haven't yet found a way to avoid indeterminacy.

13. Perhaps, then, we should maintain that while it is *generally* true that conformity to morality matters (among other factors) in determining one's moral status, conformity to the particular obligation to give people what they deserve is *not* relevant to determining status. On this proposal we need not deny the existence of status sensitive obligations. We can still recognize that one who fails to meet these obligations fails to conform perfectly to morality. We simply abandon the thought that one's conformity to such rules plays a role in determining one's moral status. Perhaps it is only the additional rules – the requirement to do X, or Y, or Z – that are relevant to fixing one's moral status. One's goodness is (in part) a function of the extent to which one meets these *other* requirements, but not at all a function of whether or not one fulfills one's status sensitive obligations.

I find this eighth proposal implausible. Assuming, as we have so far, that one's moral status is indeed at least in part a function of whether or not one obeys the requirements of morality, it is very hard to see why only some of these rules should be relevant to determining one's moral status. At any rate, it is quite difficult to see why the status sensitive rules should be singled out as being irrelevant. Surely, if we believe in the existence of status sensitive rules at all, such rules – instructing us to treat people as they deserve to be treated – will be important ones. There is no obvious justification for holding that violation of such rules has no impact on an agent's moral status, while violation of other rules does have such an impact.

14. But this, in turn, suggests a more radical proposal: that we simply abandon the thought that one's moral status can depend, at least in part, on the extent to which one conforms to morality. Perhaps moral conformity is altogether irrelevant to one's moral worth. Whatever it is that does determine one's exact level of moral worth, it is something other than one's degree of conformity to morality.

It seems to me that an approach along these lines – our ninth suggested solution – does hold out a genuine means of avoiding indeterminacy, while still retaining a desert sensitive theory.

To see why, let's consider how indeterminacy arose previously. By considering alternative assignments of status, we were able to change whether or not status sensitive rules were to be taken as obeyed, which meant that we were able to change the extent to which individuals had conformed to morality. And since conformity was relevant to fixing one's status, this meant that alternative assignments of status might then be justified. Thus alternative assignments of status sometimes turned out to be stable.

But if conformity is irrelevant to status, then this path to indeterminacy is blocked. Instead, individuals will be assigned particular moral statuses on the basis of something else – something having nothing to do with conformity. So even if we do go on to consider alternative assignments of status, although this will change whether or not status sensitive rules have been obeyed, and thus change the extent to which individuals have conformed to morality, this will do nothing to justify reassigning the status of the various individuals, since conformity per se will be irrelevant to fixing one's status. Thus the alternative assignments of status will not be stable. No indeterminacy arises.

Of course, this doesn't show that indeterminacy cannot arise in some *other* way – even for views that do take conformity to be irrelevant to fixing status. But at least it should allow for the possibility that indeterminacy can be avoided.

Put another way: indeterminacy arises when status is dependent upon status. And if status depends on conformity, including conformity to status sensitive rules, then the invidious dependence of status upon status remains in place, generating indeterminacy. If we insist, however, that status cannot depend on conformity, then we can at least break this particular connection between status and status, and open up the possibility of avoiding indeterminacy altogether.

15. There is a somewhat less radical suggestion that is worth considering as well. To break the dependence of status upon status that can generate indeterminacy, we may not need to go quite as far as to claim that conformity is altogether irrelevant to determining status. Consider more carefully the precise (indeterminacy generating) way in which status has depended upon status in our theories up to this point. One's status depends (at least in part) on the extent to which one conforms to the requirement to give people what they deserve, where what they deserve depends in turn on their status. But not all aspects of this dependence of status upon status are equally responsible for generating indeterminacy. What actually creates the problem is the fact that one's status at a given time depends (at least in part) on what status people have at that very same time. (Thus, alternative assignments of status at a given time affect the extent to which people are giving people what they deserve at that time, affecting their level of conformity, and thus affecting their status at that same time. This is what opens up the possibility of multiple stable assignments of status.)

This suggests that we might be able to avoid indeterminacy by denying the dependence of status at a given time upon status at that *same* time.

One way we might try to do this would be to claim that one's status at a given time – whether one is good or bad – could not depend at all upon the extent to which one is conforming to morality *at that time.* But this doesn't seem an especially plausible route. If we continue to maintain that one's conformity to morality at *other* times is relevant to whether one is currently good or bad, it is difficult to see how it could not also be relevant whether one is conforming to morality *now.* If my conformity to morality is relevant to my moral worth at all, it

seems hard to deny that whether I am good or bad now depends, at least in part, on whether or not I am currently meeting morality's requirements.

But there is a less implausible alternative. We could admit that one's current level of conformity is relevant to one's current *status,* but insist nonetheless that it is *not* relevant to how one *deserves to be treated* now. That is, the extent to which I am currently conforming to morality may well determine (in part) whether or not I am currently a good person; but how I deserve to be treated now depends not at all on whether or not I am a good person now. Rather, it depends on whether or not I have *been* a good person. Put another way, to the extent that conformity makes a difference to what I deserve, it is only past conformity that affects what I currently deserve. It is what I have done in the past, not what I am doing now, that affects what I deserve now.[8]

If we accept a view like this then the kind of indeterminacy we have been exploring is blocked. When we imagine alternative assignments of status at a given time, these can only affect how the relevant parties deserve to be treated at a *later* time. Thus alternative assignments of status can affect the extent to which people are later meeting the obligation to give people what they deserve, and can thus affect the extent to which people are later conforming to morality, and thus affect the moral worth of individuals at that later time. But none of this can affect how people deserve to be treated now, hence cannot affect current levels of conformity, hence cannot affect current status. In short, alternative assignments of status cannot "circle back" to justify themselves. Alternative assignments of status won't be stable, and indeterminacy will be avoided.

This approach manages to avoid indeterminacy, while at the same time allowing us to maintain that what one deserves depends in part on the extent to which one has obeyed morality. It simply insists that at least as far as this aspect of desert is concerned, it is past deeds, not current ones, that are relevant. (Current conformity will, of course, still be relevant to how one should be treated in the *future.*)

While some will certainly find this an attractive solution to our problem, I myself think that on reflection we should not accept it. For I don't see why how someone deserves to be treated now should depend only upon his *past* deeds.

Suppose, for example, that Smith will fail to conform to morality at some time in the future, in a way that will sufficiently lower his moral worth, so that punishing him will then be appropriate. If it will be appropriate to punish him in the future, why shouldn't it be appropriate to do it now as well? Of course, now that we have introduced temporal considerations into our discussion there will be legitimate questions about the amount or extent of punishment that is deserved (for example, how long someone should be hated), and we might worry that punishing now as well as later may well exceed this amount. But it might not. Indeed, it might be that it will be impossible to punish Smith later: we must do so now, or not at all (perhaps you won't be able to do it later). In a case like that, I can't see why we shouldn't think it legitimate to punish now, despite the fact that the infraction won't come until later. If we think that by virtue of a later infraction Smith will deserve punishment, and we know that he won't actually be given what he deserves later, then why should we think it illegitimate to give it to him now?

Obviously enough, in typical cases familiar epistemic limitations will leave us uncertain as to whether the future will go as we predict, and this might well make it illegitimate normally to impose the sanction prior to the infraction. But such epistemic considerations don't support the general moral claim that how one deserves to be treated now can only depend upon past deeds. If one *will* fail to conform, then it seems to me that this might well be relevant to what one *deserves now* – whether or not we are in a position to appreciate this fact.

Of course, one might hold one or another metaphysical view according to which currently there simply is no fact of the matter concerning how Smith *will* act. (This might follow, for example, from certain views about future contingents, or from certain views about free will.) Obviously, if there is currently no fact of the matter concerning how Smith will act, then such "facts" cannot play any role in determining what Smith deserves now.

But whatever force such metaphysical qualms might have with regard to Smith's future acts, they seem to me considerably less compelling when it comes to the *present.* Presumably there can be a fact of the matter concerning what it is that Smith is doing right now. It might be, for example, that right now he is failing to give someone the treatment that they deserve (say, hating someone who has been, and will remain, good). And if there *is* a fact of the matter concerning whether Smith is currently conforming to morality (and often, at least, it seems uncontroversial that there are such facts), then it is difficult to see why such facts shouldn't be relevant to what one currently deserves. More precisely, it is difficult to see why facts about present conformity shouldn't be relevant to current desert – *given* the assumption that facts about *past* conformity *are* relevant.

I conclude then – albeit with some hesitancy – that we shouldn't accept this more modest solution to the problem. Admittedly, denying the relevance of current conformity to what one currently deserves (while accepting the relevance of past conformity) does indeed avoid indeterminacy. But it is not, I believe, a position that can be maintained. We must either insist that conformity is altogether irrelevant, or admit that current conformity is relevant along with past conformity. Yet once we admit the relevance of current conformity, the possibility of indeterminacy returns. I believe, therefore, that if we are to avoid indeterminacy while retaining the framework of a desert sensitive theory we must accept the more radical conclusion: we should insist that one's moral status is not at all a function of whether or not one has conformed to morality.

Put another way, so long as conformity to morality matters at all in determining one's status, then current conformity should matter, and so there will be cases where current conformity or violation will be decisive to settling one's status, one way or the other. And this means that indeterminacy will be unavoidable. If we are to avoid indeterminacy, it seems, we must insist that conformity to the rules of morality simply plays no role at all in determining one's moral status.

This is, I believe, a surprisingly strong result. At least it surprises me. (It appears to mean, for example, that one cannot hold the intuitively plausible view that sometimes one ought to punish a given individual because he deserves it, where the reason he deserves it is because he has violated one or another of the rules of morality.)

16. Very well, then, if we are to avoid indeterminacy we must hold that one's moral status is not at all a function of one's conformity to morality. But what, then, is it a function of?

No doubt, there are various suggestions that might be made at this point, though the challenge, obviously, is to think of ideas that are reasonably attractive. It would, for example, avoid indeterminacy if we assigned status solely on the basis of one's birthday! But this is hardly a plausible view.

Luckily, there is a familiar, and natural, proposal to make at this point. Perhaps one's status is a function of one's *motives*. If one's status turns on one's motives – rather than on whether or not one has actually met one's various moral obligations – then we should, in principle, be able to avoid indeterminacy.

Now the view that motives are relevant to one's moral status is of course a popular and attractive one. It is less obvious to me how popular would be the claim that motives *alone* are relevant to status. (More precisely, the requisite claim is that conformity is irrelevant. It might be that some third factor – beyond motives and conformity – is relevant as well.) But in any event, I don't mean to suggest that this proposal is especially troubled or implausible. My main aim – already accomplished – is to have argued for the conclusion that if indeterminacy is to be avoided, motives, or something similar, must be the sole basis for assigning moral worth: conformity to morality must be irrelevant.

17. Still, there are dangers to be avoided, even by those who accept the suggestion that motives are the only basis for assigning status. For if this view is developed in the wrong way, indeterminacy can return.

Suppose we accept the view that the moral worth of a given individual is indeed simply a matter of that person's motives. Even if he has failed to conform to morality, this is in and of itself irrelevant to determining his moral worth (since, for example, the violations might be inadvertent). All that matters is what the person's motives are. If the person has good motives, he is himself good; if he has bad motives, then he is himself bad. (Obviously, more fine-grained approaches are possible, and common, here as well.)

But what marks one motive as good, or another as bad? One common suggestion is that we should rank motives "instrumentally," by comparing the consequences of having them. On this approach, good motives are those motives that will actually lead the agent to conform to morality (to "do the right thing").

Unfortunately, within the context of a desert sensitive moral theory this approach can still generate indeterminacy. For it will sometimes be indeterminate which motives are the best motives. This is a possibility by virtue of the fact that it will sometimes be indeterminate in a given situation which acts conform to

morality. After all, within a desert sensitive theory there is a moral requirement to give people what they deserve. Thus sometimes we cannot determine what someone is required to do until we have fixed the status of the relevant individuals. But we cannot do this until we have determined which motives are the best motives, and we cannot tell that until we have determined what one is required to do. In effect, determining moral status requires determining the best motives, which requires determining what one is required to do in particular situations, which requires determining moral status. Here then we have a return of the kind of regress which is the hallmark of indeterminacy.

Consider, once again, a world in which Jones and Smith hate one another. Let us suppose, initially, that Smith is bad. This means that he deserves to be hated. Jones does hate Smith, so Jones has the best kind of motive, a motive which actually leads him to do the right thing (hating Smith). So Jones is good. Since he is good, he doesn't deserve to be hated. But Smith does hate Jones. So the motives he has actually lead him to do the wrong thing (hating Jones). This means that Smith has bad motives, and so Smith is indeed bad. Thus one stable assignment holds that Smith is bad and Jones is good.

But of course the opposite assignment – with Jones bad and Smith good – is stable as well. Thus we have indeterminacy. We cannot say who is good and who is bad without first determining which motives actually lead people to do the right thing; but we cannot determine that without first saying who is good and who is bad. Alternative assignments are equally stable, and despite the appeal to motives as the basis of assigning status, we are still plagued by indeterminacy.

Nor will it help to suggest, plausibly, that in many cases we can determine the best motives by examining which motives *generally* produce right action, and that in many cases the right act can be determined independently of having first fixed the relevant statuses. This may well be the case. But it remains true that in *some* cases the best motives cannot be fixed except with an eye to the rightness of particular acts, acts whose rightness cannot be determined until questions of status are settled. And we have long since agreed that the mere possibility of indeterminacy suffices to render a theory inadequate.

I conclude that even if we fix status on the basis of motives, if we stipulate that the best motives are those that actually lead to conformity to morality then we cannot avoid indeterminacy. This common "instrumentalist" approach is not available to desert sensitive theories.

18. There are, happily, still other approaches to specifying the relevant motives. Most famously, of course, we have the proposal that to be good one must have the motive of acting for the sake of duty, doing the right act because of the very fact that it is morally right.[9] I will leave subtle questions concerning the best articulation of this position aside. It seems to me, at least in many moments, a plausible enough view. More importantly, for our present purposes, it also seems to avoid indeterminacy.

For consider a case in which status is assigned to the relevant individuals on the basis of whether they have, or lack, the motive of duty. Even if we now

tentatively alter some part of that assignment of status, the result will simply be a change in the extent to which one or another individual meets his status sensitive obligations. But this change – a change in the extent to which various agents actually fulfill the requirements of morality – does not force us to alter our view concerning whether or not they are *motivated* to act for the sake of duty, and so cannot by itself justify assigning an alternative status. Thus no further stable assignments of status get introduced, and indeterminacy is avoided. This latest proposal, then, blocks the kind of indeterminacy we have been examining.

Suppose then that we agree that one's moral status turns on whether or not one acts for the sake of duty, and we incorporate this view into our desert sensitive theory. We can then recognize that there are status sensitive requirements to love the good and hate the bad. And we can do this without running the risk of indeterminacy.

19. Of course, if we do accept a desert sensitive theory along these lines, there will be still more to say about what it is to act for the sake of duty. At least in those cases where the agent recognizes the truth of the desert sensitive theory, he will recognize, for example, that he has a moral obligation to love someone if and only if he is good. Accordingly, if he is motivated to do the right act because it is the right act, it seems to me that he will also have various derivative motives as well. He will, for example, have the motive to love someone if and only if he is good. Plausibly, he will also have the motive to view someone as good if and only if that person is indeed good. And he will have the motive to love someone if and only if he takes that person to be good. And so on.

This has some interesting implications. Assuming that both Jones and Smith recognize the truth of the desert sensitive theory, then, if each is good, each will attempt to determine whether the other is good or bad, and as we have just seen, this means, at least in part, that each is trying to determine whether the other attempts to correctly identify whether he himself is good or bad. Thus, for example, Jones is motivated to try to determine whether or not Smith is motivated to try to determine whether or not Jones himself is motivated to try to determine whether or not Smith is motivated to try to determine, and so forth and so on.

This is, of course, yet another regress. But in this case, at last – at least, so far as I can see – there is nothing vicious about the regress. Morally good individuals do indeed attempt to identify whether others are morally good or not, and this involves, at least in part, the attempt to identify whether others are attempting to identify whether one is oneself attempting to identify, and so forth and so on. This is all as it should be, so long as no indeterminacy is introduced.[10]

To reassure ourselves that there is indeed no new indeterminacy here, let us look, one last time, at the case in which Smith and Jones hate one another. What assignments are possible under our current proposal?

Interestingly enough, there are four distinct assignments compatible with the case as it has been described so far. It could be, for example, that Jones is good, and recognizes that Smith is bad, and thus Jones appropriately hates him, in keeping with Jones's motivation to love someone if and only if they are good.

Smith meanwhile recognizes that Jones is good, but hates him anyway, since Smith isn't motivated to give people what they deserve, thus confirming the claim that Smith is indeed bad. This first assignment is thus stable. And, of course, the opposite assignment – with Jones bad and Smith good – is stable as well. Furthermore, it could be that both are good. Of course, if both are good, then neither should be hated, thus each fails to meet a status sensitive requirement. But this is still compatible with both being good, for each might still be motivated to love someone if and only if that person is good, and might simply have failed to correctly identify the other person as good. (Such misidentification is obviously quite compatible with being motivated to try to identify people correctly.) Finally, it could be that both are bad. If both are bad, then each deserves to be hated, which means that each actually meets his status sensitive obligations. But this is still compatible with both being bad, for each might care not one whit about giving people what they deserve (or it might be that one, or both, tries to hate people who are good, and has misidentified the other as good).

So which is it? Is Jones good, or bad? Is Smith good, or bad? We can't tell. Given all that we have been told so far, either of them could be either. But doesn't this mean that we have an unacceptable indeterminacy after all? No, I don't think it does mean that, for we haven't been given all the relevant information. In particular, obviously enough, we haven't yet been told whether or not Jones and Smith are motivated to act for the sake of duty. Since under our current proposal it is the presence or absence of this motive that determines whether one is good or bad, it is hardly surprising that we cannot yet determine who is good and who is bad.

Still, the fact remains: until we are told what their motives are, we cannot tell which determinate assignment of status is the correct one. Isn't this still an unacceptable indeterminacy?

Again, I don't think so. It seems to me not at all problematic if we cannot determine the correct assignment of status when the reason for this is that we haven't yet been told some of what is admitted to be relevant information. The situation was quite different in the cases of indeterminacy that we faced previously. For in those cases, by way of contrast, all of the relevant facts were *in*. By hypothesis, there was no further information – relevant to settling the status of the various actors – and yet for all that there was no way to fix on a particular stable assignment. This does seem unacceptable. Here, however, we simply lack some of the admittedly relevant information. Tell us more about the case – tell us, in particular, what motives Jones and Smith have – and we will be able to fix upon a determinate assignment of status. And so here, it seems to me, there is no troubling indeterminacy.

20. Let me suggest, finally, that in solving the problem of indeterminacy – or at least noting one possible solution – we have also solved the problem of *instability* that we noted at the beginning of this paper, only to put it aside. Recall that in this type of situation the problem is not that there are too many stable assignments, but rather that there is not even a single assignment of status that is stable.

We illustrated this with the simple case in which Jones hates himself. If he is good, then he is wrong to hate himself, and so – under the terms of our original model – he is bad; but if he is bad he does right to hate himself, and so is good. Thus neither status – good or bad – can be assigned stably.

But we are now in a position to see easily that the instability arises from the assumption in the original theory that Jones's conformity to morality – whether or not he actually treats himself as he deserves – is relevant to his moral status. Each particular assignment of status carries with it a corresponding judgment concerning whether or not morality has been obeyed, which forces in its turn a different, incompatible, assignment of status. Hence the instability.

Suppose however that we adopt the suggestion that one's moral status is a function solely of one's motives, and that it depends, in particular, on whether or not one acts from the motive of duty. Then we can say the following: if Jones is good, he is motivated to treat himself as he deserves, although in point of fact – since good people deserve to be loved, not hated – he fails to do so. (This is, of course, a coherent possibility, for he may simply have misidentified his own status.) Similarly, if Jones is bad, he is not motivated to try to treat himself as he deserves. As it happens, he does in fact give himself the treatment he deserves, but this may be a sheer accident. (He might, for example, love and hate randomly, without regard to the person's goodness or badness.) Either assignment of status is perfectly compatible with the facts of the case as it has been described so far. Of course, we can't yet say which particular assignment of status is the correct one, for we haven't actually been told anything at all about Jones's motives. But as we have already seen, this kind of merely epistemological indeterminacy is not at all the same thing as the troubling metaphysical kind. And more directly to the point at hand, the instability that originally plagued the example has now been removed as well.

I believe the same result would emerge were we to consider other examples that would have displayed instability under our original model. In each such case, I believe, instability is generated by the claim that one's status may depend on whether or not one conforms to morality. Once we accept the proposal that status depends instead on whether one has or lacks the motive of duty, the mechanism through which instability is generated is blocked.

Note, finally, that here too it would not have sufficed to hold that status is determined by *motives,* if the best motives are taken, instead, to be those that will actually produce the right act. For there will not always be a stable answer as to whether the relevant motives are among those that actually produce right acts. (If, for example, we assume that Jones is good, then we must assume that his motives lead him to do the right acts, which is to say we must assume that it is right for Jones to hate himself. But this we cannot do, under the assumption that Jones is good. On the other hand, if we assume that Jones is bad, then his motives must lead him to act improperly. But hating himself is not improper, under the assumption that Jones is bad.) Apparently, to avoid instability, we cannot merely appeal to motives; we must appeal to motives in the right way. Happily, however,

our favored proposal – that the relevant motive is the motive of duty – does seem to offer us a way to avoid instability, as well as indeterminacy.

21. I conclude – although, again, a bit hesitantly – that desert sensitive theories can indeed escape the problem of indeterminacy, as well as the problem of instability, provided that they accept the view that one's status turns only on whether or not one has the motive of duty. No doubt there are other possible solutions to these problems as well – though the challenge, as I have already noted, would be to think of others that are comparably attractive.

Notes

[1] Sometimes it is claimed that consequentialists can only appeal to "nonmoral" concepts in articulating their theory of the good. This would obviously rule out the possibility of desert sensitive consequentialism; but I see no reason to accept this restriction. I should note that I am myself particularly interested in desert sensitive consequentialist theories. But to avoid needless distractions in the main body of the text, I'm going to restrict all further discussions of consequentialism to the notes. (An important recent example of a desert sensitive consequentialist theory can be found in Fred Feldman, 1997, *Utilitarianism, Hedonism, and Desert*, Cambridge University Press, Cambridge.)

[2] Recognition of this apparent problem – or at least, recognition of problems in the same general neighborhood – crops up in unexpected places. Let me mention just one. In the *Groundwork of the Metaphysics of Morals*, Kant discusses various propositions about the nature of duty. At one point he says that "it is clear from the preceding that the aims we may have in actions, and their effects, as ends and incentives of the will, can impart to the actions no unconditioned and moral worth" (Ak 4:400). What Kant seems to mean is that one cannot have moral worth by virtue of aiming at good results. This is supposed to follow "from the preceding." With characteristic obscurity, however, Kant doesn't ever indicate what previous point is supposed to actually support this conclusion (let alone make it "clear"). But I think it may be this: Kant has already argued that the only thing that is good without qualification is the good will. If true, this apparently has the implication that ordinary goods, like happiness, are not always good. But if that's so, then it cannot be a mark of a good will that one aims at trying to produce happiness (or other ordinary goods), since that will be aiming at something that is not necessarily good.

Fair enough, we might reply, but what about the possibility of aiming at the happiness of those with good wills? How has Kant ruled out the possibility that one has a good will by virtue of being the kind of person who aims at giving happiness to those that deserve it? Kant doesn't say, but it might be that he sees the looming threat of regress with such a view: a good will would be a person who aims at giving happiness to those with good wills, that is, aims at giving happiness to those that aim at giving happiness to those that aim at giving happiness to those that aim at giving happiness, and so forth, and so on. If a view like this is philosophically unacceptable – and Kant may think it obvious that it is – then this may explain why Kant thinks that our aims cannot be the source of our moral worth.

[3] Even if a theory has additional obligations, or different obligations, it will be good enough for our purposes if these obligations are effectively *equivalent* to the single obligation to give people what they deserve. Thus, for example, a consequentialist theory will presumably have an obligation to promote the overall good. But provided that it accepts

a theory of the good according to which the *only* thing that can affect the goodness of an outcome is the extent to which people receive what they deserve, this will be equivalent to simply positing the obligation to give people what they deserve, and positing only that obligation. And this will suffice for our purposes.

[4] This case has an obvious similarity to the "liar" – the sentence that says of itself that it is false. (In effect, by hating himself, Jones is saying that he violates morality. Yet if he is right, then he is wrong, and if he is wrong, then he is right.) Might the literature on the liar hold out potential solutions to our own problem? This is an intriguing suggestion, and I don't know the literature on the liar well enough to assess it properly. But note, in any event, that many proposals with regard to the liar attempt to rule it out as being somehow ill-formed, and it is far from obvious that any comparable claim in the moral case will be plausible.

[5] Of course, you might worry that an advocate of desert sensitive *consequentialism* could not embrace this proposal. For consequentialists hold that one must promote the overall good, and if it is a good thing for a person to get the treatment that they deserve, then an agent can have self-regarding duties to bring it about that they themselves get the treatment that they deserve. However, it might be suggested that a proper understanding of the value of desert reveals that while it is a good thing whether one person gets what he deserves from *others*, it is a matter of complete indifference how the person treats himself. With regard to oneself, it might be suggested, desert is indifferent. Thus it could be held that it simply doesn't matter whether Jones loves himself or hates himself – none of this affects the goodness of the outcome.

[6] Of course, anyone who accepts desert sensitive *consequentialism* will have to derive any such further requirements from the basic requirement to promote the overall good. But we can easily do that by adopting a pluralistic theory of the good, one according to which other factors can have an impact on the intrinsic value of an outcome, besides the extent to which people are getting the treatment that they deserve. It might be, for example, that it is a bad thing if animals suffer pain, and thus it might be that each person, in addition to having an obligation to give people the treatment that they deserve, has an obligation to eliminate avoidable suffering of nonhuman animals.

[7] Might there be still other stable assignments possible as well? Suppose, for example, that we assume that each is in fact extremely bad. Then in loving one another very much, each has an *extremely* serious violation of the relevant status sensitive rule. Might this violation suffice, together with those already noted, to warrant assignment as being extremely bad? If so, this assignment is stable as well. But without more details of the status assigning rule, we cannot say.

[8] I owe this important suggestion to Peter Vallentyne, who notes as well that it may be necessary to supplement this proposal with the ("arbitrary") stipulation that *initially* one deserves to be loved. (That is, if people can have *first* moments – where they have no past deeds, or anything else, to fix what they currently deserve – then at their first moments people deserve to be loved.)

[9] The classic proponent of this view is, of course, Kant in the *Groundwork.* While Kant put this view forward as part of his overall deontological position, I see no reason why it cannot be accepted by a consequentialist as well.

[10] Let me quickly mention one other, related, regress problem that Kant may have had in mind in the passage from the *Groundwork* cited in note 2. Perhaps Kant was worried about the possibility of *defining* the good will. He may have thought that desert sensitive consequentialism would have to define a good person as someone who (among other things) aims at loving good people (that is, people who themselves – among other things – aim at loving good people, and so on). Perhaps he thought that this kind of regress is unacceptable

in an adequate definition. I am not sure we should accept this condition, but even if we should, it is easily met: desert sensitive consequentialism could *define* the good person as one who has the motive of duty, and then simply note that if the good person recognizes the existence of status sensitive duties, he will indeed aim at loving good people (who, if they recognize the status sensitive duties, will themselves aim at loving good people, and so forth). Here the regress acceptably *follows* from the definition, without actually being *part* of that definition.