Perceiving Imperceptible Harms (With Other Thoughts on Transitivity, Cumulative Effects, and Consequentialism)

Donald H. Regan

University of Michigan Law School, donregan@umich.edu

Available at: https://repository.law.umich.edu/book_chapters/301

Follow this and additional works at: https://repository.law.umich.edu/book_chapters

Part of the Applied Ethics Commons, and the Law and Philosophy Commons

Publication Information & Recommended Citation
Many writers believe there can be cases which satisfy the following description: starting from an initial state of affairs, it is possible to make a series of changes, none of which alters the value of the state of affairs in any way, but such that the final state of affairs that results from the series of changes is worse than the initial state of affairs.¹ I shall call the claim that there can be such cases the "ex nihilo" claim, since in a sense it asserts that the bad effects of the complete series of changes arise ex nihilo. Proponents of the ex nihilo claim — ex nihilists, as I shall call them — usually advance the claim as part of an argument against act-utilitarianism.² If there were cases such as the ex nihilist imagines, then it would be possible to construct variants in which act-utilitarianism unequivocally required behavior which in the aggregate produced sub-optimal consequences. We could construct the sort of case I have called (while denying its possibility) an "act-utilitarian prisoners' dilemma".³ Act-utilitarianism would be, in Derek Parfit's phrase, "directly collectively self-defeating".⁴

Or so the ex nihilist thinks, and so I carelessly thought when I wrote *Utilitarianism and Co-operation*. But Wlodzimierz Rabinowicz has pointed out that the ex nihilist's case would not in fact be an act-utilitarian prisoners' dilemma.⁵ The reason is easiest to see if we have an example. According to the ex nihilist view of the lawn-crossing problem, there are no bad consequences from one person's crossing a lawn, regardless of how many others do so. So, if each walker gains some small benefit from crossing, it is better than one person cross than none, better that two people cross than one, and so on, right up to its being better that 1,000 cross than 999. But, it is supposed, the consequences of 1,000 crossing are worse than if none cross, which is to say, it is better that none cross than 1,000. But now we have a cycle. 1,000 crossings is better than 999, which is better than 998, . . . , which is better than 1, which is better than 0, which is better than 1,000. If this is all true, "better than" is intransitive, and as Rabinowicz and Michael Otsuka⁶ have pointed out, it becomes impossible to say that any pattern of behavior is either optimal or sub-optimal. So universal
satisfaction of act-utilitarianism in such a case cannot entail sub-optimal results. There is no act-utilitarian prisoners' dilemma.

I overlooked this point in Utilitarianism and Co-operation because I was intent on showing that there simply cannot be such a case as the ex nihilist imagines — which I still think is true and important to establish. If the ex nihilist's case did exist, the intransitivity of "better than" would destroy consequentialist reasoning as we know it. It would be small solace that act-utilitarian prisoners' dilemmas disappeared in the wreckage. I shall revisit my earlier argument against the ex nihilist claim, and elaborate on it slightly, in section II below. I shall point out that the main premise of the argument is effectively equivalent to the denial of intransitivity, which is not surprising in view of what we have just seen. This means that I must confront Larry Temkin's very impressive "continuum argument" for intransitivity.7

But I want to begin by focusing on a sort of case in which the ex nihilist claim has its greatest plausibility — cases in which the harm lies in the worsening of the subjective experience of some agent or agents. In such cases the ex nihilist can try to exploit the possibility of imperceptible differences.

1. INDISCRIMINABLE DIFFERENCES

Consider a standard example. Jones is wearing a wristwatch-like device, capable of delivering finely graded electric shocks. At the setting 0 there is no current. At each higher setting, the current is increased by a very small amount over the previous setting. We now perform the following experiment (with Jones's eager participation — he's a philosopher himself, and curious). First, we set the dial of the device at 1 and ask Jones what it feels like. He says, "Not much. Maybe I can feel a little something, I'm not even sure. But it's not uncomfortable, definitely not painful in any way. Just something, maybe." Next, we move the dial to 2 and ask Jones if he can feel any difference. To which he responds, "Nope. No change. I don't feel anything different at all." Next, we move the dial to 3 and ask Jones if that feels any different from the way it felt at 2. Again, he says, "No, no difference that I can feel. It feels the same." Now, we move the dial from 3 back to 1 and ask if that makes any difference. At that point, Jones says, "Yeah, as a matter of fact, it does. I wouldn't have expected it to, given my previous responses, but it does, and now I realize that 3 was actually a bit uncomfortable. Not bad, or I would have felt the difference sooner. But there was a tiny bit of discomfort at 3 — just an almost unnoticeably slight pain, but still, something that I would prefer not to have continued indefinitely." So far so good. This is an instance of the
intransitivity of "indiscriminably different". 2 is indiscriminably different from 1; and 3 is indiscriminably different from 2; but 3 is discriminably different from 1.

At this point it is very tempting to reason as follows: (a) the badness of pain is entirely a matter of how things feel; (b) Jones cannot feel any difference between the way things are at 1 and the way things are at 2; therefore, (c) things are no worse, in respect of Jones's pain, at 2 than at 1. By the same argument, things are no worse at 3 than at 2. But Jones can feel the difference between 1 and 3, so things are worse at 3 than at 1. In sum, two marginal changes that each make no difference at all to the value of the state of affairs (from 1 to 2 and from 2 to 3) combine to make a change (from 1 to 3) that does make a difference.

The ex nihilist, of course, embraces this conclusion. For myself, I think the reasoning, plausible as it sounds, is mistaken. But how, exactly? Some people who reject ex nihilism are driven by cases such as this to believe in "imperceptible harms". They claim that the change from 1 to 2, for example, harms Jones even though he cannot feel any difference. This is paradoxical. If the sort of harm we are talking about is entirely a matter of how things feel, how can a change be harmful that cannot be perceived? My sympathies are much more with the believers in imperceptible harms than with the ex nihilists. But the bare assertion that there must be imperceptible harms seems merely to replace one paradox with another. Also, if we simply assert the existence of imperceptible harms without any further explanation, it seems that we have no choice but to assume, as believers in imperceptible harms usually do, that each step in a sequence such as we have imagined must account for an equal share of the total harm. But this amounts to assuming, at least for this sort of case, what I have rejected as the "contributory consequences" approach and Parfit as the "Share-of-the-Total View". So, can we do better?

Consider a different experiment. We begin by setting the dial to 1 and asking Jones what it feels like. He responds exactly as he did above, "Not much. Maybe I can feel a little something, I'm not even sure. But it's not uncomfortable, definitely not painful in any way. Just something, maybe." Next, we move the dial to 3 and we ask him what it feels like there. Notice we do not ask for a comparison with 1. We just ask, as we asked about 1, what it feels like. His response is again the same in substance (allowing for the difference in the form of the judgment) as what he has told us before. He says, "That's a bit uncomfortable. Not bad. But there is a tiny bit of discomfort. It's an almost unnoticeably slight pain, but it is something that I would prefer not to have continue indefinitely." Now we move the dial to 2, and once again we ask
the non-comparative question: we ask what it feels like. At this point, I don't know or care what Jones says. The important point is simply this: it is not logically possible that what Jones says about 2 should be both the same as what he said about 1 and the same as what he said about 3, since what he said about 1 and what he said about 3 are different. Whatever he says about 2, it must either differ from what he said about 1, or differ from what he said about 3, or both.

What do we learn from this experiment? Recall that after our first experiment, we argued from the premises (a) that the badness of pain is entirely a matter of how things feel, and (b) that Jones could not feel any difference between the way things are at 1 and the way things are at 2, to the conclusion (c) that things are no worse at 2 than at 1 (and similarly, no worse at 3 than at 2). But the second experiment casts grave doubt on this inference. Once again, we may assume that the badness of pain is entirely a matter of how things feel. But we now have evidence either that 2 feels different from 1 or that 2 feels different from 3. And we have this evidence from exactly the same privileged source as the evidence in our first experiment, Jones himself. With regard to at least one of these pairs, Jones's reports of how the conditions feel are not the same. For convenience, let us assume that there is a difference between Jones's description of how 2 feels and his description of how 3 feels. (Obviously, the argument would proceed the same if the difference were between 1 and 2.) We now seem to have evidence both that 2 and 3 feel the same (since Jones felt no difference in the first experiment) and that 3 feels worse than 2 (from his descriptions in the second experiment).

What should we think, given this conflicting evidence? The first question, of course, is just how serious the "conflict" is anyway? In most contexts, we would see no genuinely problematic conflict between (reliable) evidence that Jones could not distinguish perceptually between two situations and other (reliable) evidence that there was a significant difference nonetheless. We would conclude without hesitation that there was a difference, which Jones was unable to perceive. This would account easily and naturally for all the evidence. But if we attempt that move in this context, we are brought up against the question how there can be any difference between two situations in respect of a predicate grounded in Jones's feeling, if Jones cannot feel any difference between them. Of course, this "if Jones cannot feel any difference between them" overstates the case a bit. Asked for a comparative judgment (between 2 and 3), Jones can feel no difference. But asked for separate non-comparative judgments, he reveals that he does feel 2 and 3 differently nonetheless, at least when they are presented and inquired about in that way. The real puzzle in this area is just this incongruity in Jones's reports. How can it be that he feels 2 and
3 differently if they are presented separately, but that he feels no difference between them if they are presented side-by-side?

Puzzle it is, but not a logical paradox (like the ex nihilo claim). The answer to the puzzle will lie in the facts of the neurobiology of pain. I do not know the true answer; it is possible that no one currently could give a really complete answer. But just to emphasize that the question here is an empirical one (and for no other purpose), I shall offer a couple of speculative answers. One possibility is that when Jones is presented with two very similar stimuli in succession and asked to compare them, the effect of trying to compare the present stimulus with the recent remembered one is to muddle them together so that each appears to be an average between them. 1 and 2 may be close enough together for this effect to operate, and likewise 2 and 3, but not 1 and 3. Notice that if this is the explanation of the puzzle, then when Jones says 2 feels the same as 1, he is reporting on a different internal presentation of 2 than when he says that 2 feels the same as 3. His internal presentation of 2 when asked for a non-comparative judgment must be different from at least one of these other internal presentations of 2, so it is no longer surprising that his non-comparative judgments should reveal differences that are obscured when he makes comparisons. And it is natural to assert that the non-comparative judgments are a better guide to the "way things are" — at least when we are dealing with a chain of comparisons that causes inconsistency in the internal presentations of the intermediate stimuli.11

Here is another possibility. It may be that when 1 and 2 are close and 1 is presented first, then the internal presentation of 2 is just assimilated to the internal presentation of 1; and further that if 3 is close to 2, it is then assimilated to 2, and thus to 1, in the same way. So it may in fact be that after we have started with 1, and moved to 2 and evoked a comparative judgment, and then moved to 3 and evoked another comparative judgment with 2, 3 really does "feel the same" as 1. But the situation is unstable, like a supercooled liquid. As soon as we ask about the comparison between 3 and 1, Jones "realizes" that 3 is enough different from 1 to feel different, and so now it does — as we may assume it would have from the start if we had solicited only non-comparative judgments.

(It might seem that I have just allowed the existence of a case where two imperceptible changes add up to a perceptible worsening, even though neither imperceptible change in fact made things worse. We can imagine that after we moved from 1 to 2 and from 2 to 3, we asked Jones for non-comparative reports as well as comparative, and that he not only said he could feel no difference but gave identical non-comparative reports as well. And yet,
there was an overall change. The answer to the puzzle should be obvious. Our latest version of the experiment actually involves three significant interventions in Jones's situation: (1) the move from 1 to 2, (2) the move from 2 to 3, and (3) asking Jones for a comparative judgment on 1 and 3. Neither of the first two "imperceptible" interventions caused any real harm, even as judged by the non-comparative reports; but then, neither did the first two interventions together. After the first two interventions, Jones's non-comparative reaction to 3 was the same as his original non-comparative reaction to 1. So there was as yet no harm. It is only the complete chain of three interventions that has produced any overall harm; and we can see that the third intervention in particular made Jones worse off, even though it did not involve changing the setting on his shock-device. The third intervention, by some cognitive-perceptual mechanism we do not and need not understand in detail, changed how 3 felt to Jones. So in fact, our general claim is vindicated. At exactly the point where some harm appears from the sequence of interventions, we are able to identify a particular intervention in the sequence which causes harm at the margin. This example emphasizes that how a particular condition feels may depend on history and context – but however it feels, there will be no instance of the ex nihilo claim.

I have observed that in general, if we are trying to decide whether two situations are different, we would give more weight to reliable evidence of difference between them than to equally reliable evidence of Jones's inability to perceive a difference. And my neurobiological fantasies are meant to suggest that that is still the sensible strategy here, even though the ultimate question is about how things "feel" to Jones. We should credit the evidence of difference. It is important to remember, as I have stressed before, that the evidence for the proposition that the two stimuli feel different comes from exactly the same source as the evidence that they feel the same, namely, Jones himself. So, there is no danger of "undermin[ing] the distinction between the physically objective and the experientially subjective", which Warren Quinn warns us against in his defense of the ex nihilo claim.

Of course, the strategy of preferring the evidence of difference which comes from Jones's non-comparative reports over the evidence of sameness from his comparative reports would not be sensible if it could be argued that the evidence of the non-comparative reports is suspect. I can imagine someone arguing, for example, that Jones's report on condition 2 in the second experimental protocol does not count, because we have somehow "forced" Jones to respond in a way that will differentiate 2 either from 1 or from 3. But that is not true. We haven't forced any response. Jones has been perfectly free at every point to say exactly what it seems to him appropriate to say. He could have
given exactly the same description of how things felt in all three conditions, if
that was how they struck him. It is true that once Jones has given differing
reports on 1 and 3, he cannot report on 2 in a way that will fail to differentiate it
either from 1 or from 3, or both. In that sense, he is forced to differentiate it, if
he continues to respond. But it is the limits of the logically possible that "force"
him. There is no sense in which we are forcing him to invent or to report a
distinction where he feels none. (It is worth remembering that in a sense Jones
is not "reporting a distinction" at all. If he is able to focus his mind exclusively
on each of the non-comparative questions as we ask it, he might in principle be
as surprised as we to find that he has reported differently on stimuli that he
cannot distinguish when they are presented side-by-side and he is asked to
compare. But it seems to me that the right reaction, even for Jones himself,
would be to believe that the stimuli do in fact feel different when presented and
attended to separately, even if the difference is imperceptible as such, that is, in
the making of a single comparative judgment.)

It may yet seem that there is something misleading and improperly
compulsive about the questions we put in the second experimental protocol. If it
seems that way, the reason must be that we are troubled by a worry that we are
mistakenly reifying "the way it feels [always to Jones, of course] at 1 or 2, or
wherever". And yet, we regularly ask and answer questions like "How does it
feel?" or "Does that hurt?" And it is clear that we are asking for and giving
more than a report on how "that" compares to the state just before it, since the
answer may be, "Yes, it hurts. It hurts less than a moment ago, but it still hurts."
It might be suggested that what pass for non-comparative judgments about
whether something hurts are all comparative at a deeper level, because they all
involve comparison of the state in question to some standard "neutral" feeling
state. Even if this is true, comparison to a standard reference point gives us in
effect a one-place predicate; whereas what we have been referring to as
"comparative" judgments involve comparison between two states neither of
which is a standard reference point, and such judgments ineliminably involve a
two-place predicate. So, on any analysis, what we have called "non-
comparative" judgments are different from what we have called "comparative"
judgments, and they are very much a part of our ordinary conceptualization of
the phenomenon of pain.

Not only does our ordinary intuitive understanding presuppose a "way
it feels", but so does all consequentialist reasoning. (And so, as we shall see,
does the ex nihilist's reasoning, even if, like Quinn, he is attacking
consequentialism.) The basic bearers of value for a consequentialist theory are
states of affairs, which include elements like "the way it feels at 2". There is no
way (either for the consequentialist or for the ex nihiloist) to translate the fact that Jones cannot distinguish between stimulus 1 and stimulus 2 into any conclusion about how Jones (or anyone) should act without positing, explicitly or implicitly, the intermediate proposition that Jones is in fact no worse off with 2 than with 1. That is to say, without positing that "how things feel at 2" is no worse than "how things feel at 1". Ultimately, Jones's inability to distinguish is relevant only as evidence for the proposition that "how things feel at 2" and "how things feel at 1" are the same. And this should be Jones's view of the matter too. What he is really concerned about is not whether he can tell the difference between two settings, but how he fares at each. Whether he can tell the difference, when they are presented side-by-side and he focuses directly on the comparative question, is evidence on the issue of whether he fares the same or differently under each. But it is no more than evidence. It is not a final answer to the ultimate question. Some writers may have overlooked this because of a subtle and unconscious confusion about what act-consequentialism requires us to evaluate. Loosely speaking, whether the act-consequentialist should move the dial on Jones's device from 2 to 3, say, depends on the value of the consequences of that change. And the consequences of the change are the difference between the state of affairs at 3 and the state of affairs at 2. So it appears that the act-consequentialist propriety of the move depends on the value of the difference between 3 and 2; and it appears further that we can evaluate that difference directly by asking Jones to give us a comparative judgment as he moves from 2 to 3 about whether it feels worse. In most ordinary cases, this procedure would get us the right answer. But if the situation is one where Jones reports no difference between 2 and 3 when asked for a comparative judgment but describes them differently when asked for non-comparative judgments, then the procedure misleads us. The basic bearers of value for a consequentialist, as I have said, are states of affairs, and changes in states of affairs are not themselves states of affairs. So strictly speaking, a change is not something that has value at all. When we speak of the value of a change, this is an elliptical way of referring to the difference between the values of the final and initial states. This difference in values is what really matters. So what we are primarily interested in is not the value of the change from 2 to 3, but rather the difference between the value of 3 and the value of 2. Which is to say, the difference between the value of "how it feels at 3" and the value of "how it feels at 2". And the most direct evidence we have about what each of those states is like (from our second experiment) suggests that "how it feels at 3" is different from "how it feels at 2", even though Jones cannot distinguish between them in a single comparative judging.
It may seem that I put too much credence in Jones's non-comparative reports. If we run our second experimental protocol a second time, it is unlikely that Jones's descriptions of "how it feels" at each setting of the dial will be just the same as they were the first time. Of course, we can assume that the descriptions of how it feels at 1 and how it feels at 3 will still be different, since the existence of a difference between these is one of the basic features of the puzzle situation. And that means that in any particular trial, the description of how it feels at 2 will differ from at least one of the other descriptions, just as we have said. But the ex nihilist might claim that the variability of Jones's descriptions casts doubt on whether there really is anything he is describing, and thus supports the idea that the second protocol has somehow forced Jones into an illusory discrimination.

This is not persuasive. Jones's descriptions may vary for a number of reasons that are fully consistent with there being a determinate "way things feel" at each setting in each experimental trial. The descriptions may vary because how things feel is actually slightly different in each trial, which could be because of different detailed history and context, or indeed because of random neurological events in Jones. The descriptions may also vary because language is a very imperfect medium for reporting inner events and states. This second point does reveal that Jones's reports do not give us perfect access to how things feel to him at each setting in each experiment. But the point of introducing the second experimental protocol does not depend on our having perfect access through Jones's reports to the way things feel. Our access is often good enough so that Jones's reports will allow us to identify genuine differences in the way things feel, even at "indistinguishable" settings. But the real point of the second protocol, with its seeming reliance on our having access through Jones's reports to how things feel, is just to focus our attention on that "how things feel". It is to remind us that Jones himself can make, and makes, non-comparative judgments on individual settings as well as judgments about discriminability. As do we all. We should not be led by the fact that individual experiments are neither perfectly informative nor perfectly repeatable to doubt the existence of the underlying phenomena.

Let us now change our focus slightly and consider Warren Quinn's well-known ex nihilist argument. (As we shall see, Quinn seems to anticipate something like my argument from the second experimental protocol, and he responds partly by doubting the determinateness of "the way it feels".) Quinn imagines Jones wearing a shock-device with a thousand-and-one settings, numbered 0 to 1,000. Neighboring settings give currents which are indiscriminable. The lowest setting is completely painless. The highest is
torturous. Once a week, Jones is given an opportunity to be paid $10,000 if he increases the setting on the dial by one. At each choice point Jones is allowed a brief period of free experimentation with various settings, but at the end of the period for experimentation, the dial is reset to where it was at the beginning of the period, and Jones's only choice is to leave it where it is or to advance it. Except temporarily during the experimentation period, the dial can never be set back.

Quinn's claim is that if Jones is guided by standard tenets of rationality, he will have the dial set forward each week. After all, Jones cannot feel any difference between the present setting and the next higher, so he is no worse off in regard to the pain, and he gets $10,000. This goes on for a thousand weeks, at the end of which Jones (aside from being twenty years older) has a fortune of $10,000,000 and is committed to torturous pain for the rest of his life. At this point, Jones wishes he could give back the $10,000,000 and take off the device. But it is too late. And in Quinn's view, it is the standard principles of rational decision-making that have brought disaster on Jones's head. (Actually, if Quinn's story was the whole story on Jones's preferences, we could not say the outcome was a disaster. Quinn equips Jones with intransitive preferences, which makes it impossible to say that any result is globally worse for Jones than any other; and his attempt to avoid this difficulty seems to me completely inadequate. But since I think Quinn's story is not the whole story, and since I agree that if Jones keeps moving the dial right up to 1,000 that is a disaster, there is no need to pause over this.)

As I have mentioned, Quinn anticipates an argument something like mine. He imagines someone suggesting that we should assign to each setting a "discomfort-index", quantifying how uncomfortable Jones is at that setting. The index will start at 0 at the setting 0, and will be some very high positive number at 1,000, so there must be some first setting s at which the discomfort-index is positive. And therefore, whether Jones can feel it or not, his comfort level must decline in moving from s-1 to s. Quinn responds: "But the measure of [Jones's] discomfort is indeterminate. There is no fact of the matter about exactly how bad he feels at any setting. And if so, we cannot argue that the measure of his discomfort must increase in some single step."

Now, there are a number of possible responses to this, including the assertion that Quinn mistakenly thinks he sees indeterminacy where there is really only variability according to context and the limitations of language as a medium for reporting inner states. But the point I want to make now is that we can concede Quinn his indeterminacy without conceding his conclusion. If indeterminacy prevents us from saying that one of two neighboring comfort
levels is better or worse than the other, which is Quinn's claim, then it must also prevent us from saying that the two comfort levels are the same. But, as I have noted previously, what Jones is ultimately concerned about is not his powers of discrimination, but his comfort level at various settings. And at any point where it is impossible to say the neighboring comfort levels are the same (which seems on Quinn's implied view to be every point), Quinn's argument that rationality requires Jones to advance the setting breaks down. His argument depends on the premise that if Jones cannot tell the difference (within the protocol) between neighboring settings, then they have the same comfort level, so the $10,000 for advancing the setting can be secured at no (local) cost. But this premise is now unavailable.

We can see essentially the same point in the context of another of Quinn's objections-and-refutations. The objection is given thus: "[Jones] feels no worse at 1 than at 0, and the comfort-comparison between 1 and 0 is the same for him as the comparison between 2 and 0. [This Quinn claims to have established previously.] From this it seems to follow that he feels no worse at 2 than at 0. Reiterated enough, the argument implies that he feels no worse at 1,000 than at 0. Here we find a kind of sorites puzzle. Some of his clear and immediate judgments about his comparative comfort are true only if others are false." Quinn attempts to refute this charge of inconsistency in his method by making two observations. First, he notes truly that his argument does not require Jones to actually rely on the empirically false conclusion of a sorites argument. (Jones never bases any reasoning on the premise that he feels no worse at 1,000 than at 0.) And he continues, "What naturally matters to [Jones] is that the comfort status of $s$ and $s+1$ are, introspectively and behaviorally, no different . . . . It is enough for him that the empirical data give him no reason to suppose that his comfort declines, either directly or relative to some fixed point, in any single step." This second point is either wrong or too weak to carry Quinn's conclusion.

Consider. There is going to be some point long before the dial is set to 1,000 at which Jones realizes that the current going into his wrist hurts. (That there will be some such point, Quinn agrees. He insists we cannot identify a first such point, but, right or wrong, that does not matter.) Now, at any point $s^*$ where he realizes that the current hurts, a rational Jones will also realize that something is wrong with the reasoning Quinn recommends. As we have seen, Quinn's reasoning tells Jones that he feels no worse at any $s$ than at 0. But Jones can now observe empirically that that is false. He has arrived at an $s^*$ where he feels worse than at 0. So something is wrong with the reasoning. And it is clear
in a general way what is wrong. Somehow the judgments of indiscernibility, however confidently made and even correct, do not tell Jones all he needs to know. There is more to the facts about how things feel than those judgments encompass.

When Quinn says "the empirical data give [Jones] no reason to suppose that his comfort declines ... in any single step," he apparently means to insist that the judgments of indiscernibility are still enough to require Jones to move along one step at a time. But that is surely mistaken. Given the complete ensemble of empirical data, which includes the fact that s* hurts, Jones has reason to think that either (a) his comfort level has declined, unnoticed, at some step, in which case it could happen again, or (b) he doesn't have any idea what is going on, in which case he has no business regarding any particular choice as "rationally required". Quinn would in effect have Jones believe that he understands perfectly what is going on at each individual step, even though he cannot figure out how the steps combine to produce a result that is inconsistent with what logic, plus his understanding of the individual steps, demands. This is not reasonable. Quinn might object that I still have not said what precisely Jones ought to do. That is true. On my view, there are facts about how things feel at different settings that Jones may not have adequate access to in the situation Quinn describes. That is a problem for Jones; but how he should proceed is not my present concern. My present concern is just to point out that when Jones realizes the comparative judgments of indiscernibility are not telling the whole story, he will see that he is not rationally required to keep advancing the dial. Whatever Jones should do at this point, Quinn's story breaks down.

A final remark on Quinn, just for its heuristic value. As Quinn imagines the case, Jones gets $10,000 each time he advances the setting on his shock-device. But notice that if Quinn's argument were sound, it would work exactly the same if Jones got only $10 for each advance. That alteration, however, would make the argument seem vastly less compelling. So the argument does not work as advertised. Which leaves us with no reason to think it works at all.

Let us pause to take stock. My main object in this first section has been to suggest that cases involving indiscriminable differences look very different, and cease to offer any special threat to act-consequentialism (in moral reasoning or in prudential), if we redirect our attention from the subject's judgments of indiscernibility to his judgments of individual states of affairs. In a sense our second experimental protocol vindicated "imperceptible harms", but only by showing that they are not really imperceptible after all. And nothing about the
argument suggested either that every physical change in a sequence such as we imagined must occasion imperceptible harm, or that the harm must be evenly spread over the discrete changes.

Throughout this section I have been arguing against the ex nihilo claim, but it may seem that the precise form of my opposition changed between the discussion of our original experiments and the discussion of Quinn's scenario. In fact, it did. The ex nihilo claim says that there can be cases such that, starting from an initial state of affairs, it is possible to make a series of changes, none of which alters the value of the state of affairs in any way, but such that the final state of affairs that results from the series of changes is worse than the initial state of affairs. The denial of this claim has a somewhat different purport depending on whether each state of affairs has a determinate value. If each state of affairs does have a determinate value, then the negation of the ex nihilo claim entails that whenever a series of changes results in a final state which is worse than the initial state, there must be at least one particular change which has made things worse. If it is not the case that each state of affairs has a determinate value, then the negation of the ex nihilo claim says only that if a series of changes results in a final state which is worse than the initial state, then it cannot be that every individual change has determinately left the value of the state of affairs unchanged. (Even this, as we saw in the discussion of Quinn, is enough to block the ex nihilist's argument against act-consequentialism.) For myself, I am inclined to think that each state of affairs has a determinate value (more on that in the next section), but it is certainly worth having both versions of the negation of the ex nihilo claim in play.

2. THE PROBLEMS FOR CONSEQUENTIALIST THEORIES

Let us now consider from a more general perspective what all this means for consequentialist theories. First, let me state the most general version of the claim that we have been tending towards in section I:

(S) If we move from one state of affairs to another by a series of intermediate changes, then the difference between the value of the final state of affairs and the value of the initial state of affairs is equal to the sum of the marginal differences made by each intermediate change along the way.

I shall call this the "summation" principle. It presupposes that we can associate with each state of affairs a real number which represents its value (a presupposition I shall consider further in a moment). But given that presupposition, it is a simple truth of
Let us call the various states of affairs 1 through N. The value of state of affairs i we call $V_i$. The difference in value brought about by any particular change from state i to state $i+1$ is $V_{i+1} - V_i$. So, what (S) says is that

$$V_N - V_1 = (V_2 - V_1) + (V_3 - V_2) + \ldots + (V_{i+1} - V_i) + \ldots + (V_{N-1} - V_{N-2}) + (V_N - V_{N-1}).$$

This, as I say, is a truth of arithmetic.

If instead of focusing on marginal and total changes to overall value (benefit net of harm) over the series of changes, we focused either on marginal and total benefit over the series of changes or on marginal and total harm (assuming benefit and harm are separately identifiable), it is obvious that analogues to (S) would also be true. In addition, the summation principle (S) entails the following "intermediate change" principle (I), which is a more limited generalization of the conclusion from section I:

(I) If we move from one state of affairs to another by a series of changes; and if the value of the final state of affairs is different from/greater than/less than the value of the initial state of affairs; then it must be the case that at least one of the intermediate changes alters/increases/decreases the value of the state of affairs.

This 'intermediate change' principle is what most directly prevents the existence of act-utilitarian prisoners' dilemmas such as the ex nihilist imagines. It tells us that whenever a series of changes makes things worse, then some particular change along the way must make things worse, and thus cannot be required by act-utilitarianism.

Now, what about the presupposition that we can associate with each state a real number which represents its value? Some possible difficulties are easily dealt with. We have noted that the value of the state of affairs produced by a particular stimulus or resulting from a specified change may vary with context or history. We can either think of this as one incompletely specified state of affairs having different values depending on context or history, or we can think of it as the same intervention bringing about different completely specified states of affairs, each with a different value, depending on context or history. How we think of it does not matter, provided we remember what we are doing. In either event, the number representing the value of the state of affairs brought about by a particular intervention will need to vary with context or history. But provided the valuations of various states are made under consistent assumptions about context and history, (S) will be satisfied.

Similarly, to claim that (completely specified) states of affairs have determinate values is not to claim that we always know precisely what they are. We may be ignorant of precisely what state of affairs an intervention produces in particular circumstances, or we may be ignorant of just what value the state of affairs
possesses. In such cases it may be that all we really have is a probability distribution (over the possible states of affairs, or over the possible values, or, if there is uncertainty about both, over the combination). But even this will allow us to assign an expected value to the result of each intervention, and once again, satisfaction of (S) by these expected values is guaranteed by arithmetic.

In conversations about the intransitivity of indiscriminability I have encountered the suggestion (not in print so far as I know) that intransitivity is best accounted for by thinking of the values of states of affairs not as real numbers, but as some sort of fuzzy numbers — blurs, as it were, on the real line. Two states that correspond to overlapping blurs are indiscriminable, but if one blur is entirely to the right (left) of the other, then the state corresponding to the first blur is better (worse) than the state corresponding to the second. It is difficult to know how to respond to such a suggestion in the absence of more detail about how the blurs are supposed to work. And yet, the idea is not transparently silly. It even has a sort of appeal, since it does sometimes seem a heroic assumption to suppose that every bearer of ultimate value has a value that is completely definite. This is no doubt why Quinn felt free to assert without argument that Jones's comfort levels were indeterminate. So — it seems worth mentioning the "blur" possibility, even if I need someone else to develop it before I can deal with it properly.

This much I can say. The blurs have to be individuated somehow if it is to be possible for any two states to have distinguishable values. The blurs must therefore have properties beyond their blurriness. The two most obvious ways to think of the blurs are (a) as bounded intervals on the real line or (b) as some sort of density functions over the real line. If the blurs are bounded intervals, they will have left endpoints (greatest lower bounds) and right endpoints (least upper bounds) and midpoints between their endpoints. Similarly, if the blurs are density functions, they will have something like means and medians, and so on. With regard to any of these numbers that help to define the blur, the analogue of (S) will be satisfied by those numbers. So we are not as far from the world of determinate values as it may seem.

Additionally, the blur idea may not be as intuitively satisfactory as it seems at first. The idea is appealing because it relieves us of the claim that every bearer of ultimate value has a perfectly definite value. And it also seems at first to give a nice representation of the intransitivity of indiscriminability, since "overlaps" is an intransitive relation on blurs. But it is not clear that this model of intransitivity actually captures intuitively the real-world phenomena. We have observed that blurs must have some properties other than their blurriness; and Jones might be able to discriminate on the basis of those other properties between some blurs that overlap. If, in our original experiment, the pain at 1 and the pain at 2 are represented by blurs which overlap just a little bit at the right end of the 1-blur and the left end of the 2-
blur, would not Jones regard these as different? More particularly, would it not be irrational of Jones to regard them as interchangeable in every choice situation? To avoid this objection, the blur-theorist might say that Jones does not himself perceive the blurs; rather they represent somehow the range of possibilities for what he may perceive. But then what does he perceive, and what reason have we at all for thinking that overlappingness of blurs corresponds to Jones's inability to perceive any difference? (Note that we cannot say the pains are indiscriminable only if the blurs are identical, or have the same endpoints, because then indiscriminability would be transitive.) Enough about blurs.

To my mind, the most powerful attack on the idea that every state of affairs has a determinate value does not involve indiscriminability, or indeterminacy on the model of blurriness, at all. I refer to Larry Temkin's "continuum argument for intransitivity",19 which I shall summarize in a moment. The conclusion Temkin purports to establish is that "all things considered better than", the fundamental concept for moral evaluation of states of affairs, is intransitive. Now, if every state of affairs has a determinate value which can be represented by a real number (and if, just to be explicit, the "better than" relation is modeled by "greater than" on the real numbers – as we have naturally assumed all along), then "better than" is transitive, just because "greater than" (on the real numbers) is transitive. So, conversely, if "better than" is not transitive, it cannot be the case that every state of affairs has a determinate value representable by a real number. Of course, for those of us who have a strong intuition that states of affairs do have values representable by real numbers, that amounts to a strong intuition against Temkin's conclusion. But Temkin recognizes that for many people his denial of transitivity is "wildly counterintuitive".20 And the intuition that states have real-number values is not an additional intuition against Temkin, since that intuition and the intuition that "better than" is transitive are on reflection pretty much the same.

So, what is Temkin's argument? It rests on three claims. First, Temkin claims that for any experience of pain, of whatever intensity and duration, another experience which involves a slightly less intense pain lasting twice as long is worse. Second, he claims that the possible intensities of painful experiences form a continuum from the pain of excruciating torture at one end to something like the slight discomfort of a hangnail at the other. Third, he claims that however long one's life (and this may include lives of millions of years), suffering the slight discomfort of a hangnail for one's entire life would be better than suffering two years of excruciating torture somewhere in that life.

These three claims are enough to generate a sequence of possible lives on which "better than" is intransitive. Begin with a (very long) life, call it A, that includes two years of excruciating torture. By Temkin's first assumption, A is better
than B, a life of the same length that includes four years of a pain, perhaps still
torturous, that is slightly less intense. B, in turn, is better than C, which includes
eight years of a pain slightly less intense than the pain in B. We proceed downwards
along the continuum of painfulness guaranteed by Temkin's second assumption, at
each step reducing the pain slightly but doubling its duration, until we have got to a
life Y in which the pain of a hangnail is continued for however many years our
process of repeated doubling has now got us to. Since A is better than B, and B is
better than C, and so on right down to X is better than Y, we can see that if "better
than" were transitive, A would be better than Y. But, by Temkin's third assumption,
Y is better than A. So "better than" is not transitive.

I am not persuaded. It's not that I have a crushing counterargument. My
not being persuaded results mainly from a very strong reluctance to be persuaded. I
share the view Temkin attributes to Derek Parfit, that if "better than" is not transitive,
practical reasoning is seriously endangered, perhaps destroyed. But if I haven't got a
counterargument, I can at least point out a specific ground for unease with Temkin's
argument. My first reaction to the argument was to be uncomfortable with Temkin's
first assumption — that given any painful experience, of whatever intensity and
duration, another painful experience involving a slightly lesser pain lasting twice as
long is worse. This assumption of Temkin's is not an intuition about a particular
comparison between two well-specified experiences. Rather it is a schema for
intuitions about an infinity of possible comparisons. I do not deny that it has a
certain plausibility as a universal schema, but I have less confidence in this sort of
intuition-schema than in particular intuitions.

This thought led me to wonder if we could do without the schema. Could I
identify some single intermediate experience somewhere between the two years of
torture and the immensely long life with a hangnail which would be both worse than
the two years of torture (because of being more prolonged) and better than the
lifelong hangnail? If such an experience, call it Q, could be identified, then A would
be better than Q, Q would be better than Y, and Y would be better than A, so we
would have a simpler example of intransitivity, based on particular well-specified
intuitions. But I have not been able to identify a Q. For Q to exist, we need a level
of pain which (a) is sufficiently bad so that a long stretch of it can be worse than two
years of excruciating torture, and yet (b) is sufficiently not-bad so that that same long
stretch of it is still better than some very much longer period with a hangnail. It
seems that the floor put under the intensity-of-pain by (a) is higher than the ceiling
put on the intensity-of-pain by (b).

Now, the fact that I cannot come up with a simpler example than Temkin's
to prove his point may not seem like much of an argument against him. But these
thoughts led me to the further observation that if Temkin's conclusion is correct, then
there must be a simpler example. Temkin's example involves a long chain of states, but whenever we have a long chain that violates transitivity (which I shall call an n-cycle), there must be some subset of three elements in the chain, not necessarily contiguous, which violate transitivity also. (In fact, this is the only reason we are justified in speaking of the n-cycle as violating transitivity, since transitivity is defined in terms of how the relation behaves on three-element sets.) Consider a chain 1 to N, where i+1 is always worse than i (1 i i N-1), but N is better than 1. (This is Temkin's example, renamed for convenience.) Now consider the relation between 2 and N. If 2 is better than N, then the set consisting of 1, 2, and N violates transitivity. (1 is better than 2; 2 is better than N; but 1 is not better than N.) If, on the other hand, 2 is not better then N, then we turn our attention to the reduced chain from 2 to N (which incidentally is now known to violate transitivity) and repeat the argument, in the reduced chain, by asking about the relation between 3 and N. If 3 is better than N, then the set \{2, 3, N\} violates transitivity. (2 is better than 3, and 3 is better than N, but 2 is not better than N.) If 3 is not better than N, we transfer our attention to the chain from 3 to N, and we ask about the relation between 4 and N. And so on. It is clear that we must eventually find a set of three elements that violates transitivity. If the process we have been describing continues up to the point where we find that N-2 is not better than N, then the set \{N-2, N-1, N\} violates transitivity. QED. I should note that I have made one assumption beyond Temkin's, which is that all pairs of the form i,N are comparable in terms of "better than". (Either i is better than N, or it is not.) But that is not an assumption I would expect Temkin to deny.

The upshot is that if Temkin's conclusion is correct, there must be a very much simpler example to demonstrate it, an example involving only three states. If we were confronted with any putative three-state example, then we could think about the plausibility and robustness of just three well-specified intuitions, rather than the intuition-schema that is Temkin's first assumption (and various other complexities such as he discusses in his article). It seems reasonable to regard a conclusion so counterintuitive as Temkin's as unproven until he has either offered the three-state example that we know must exist, or at least explained why, even though it must exist, we cannot specify it.

3. OTHER EX NIHILO ARGUMENTS

The one sort of ex nihilo position that I cannot understand at all is that of someone like Jonathan Harrison, at least if we take what he says at face value. Harrison's general view is that one should do [not do] an act if the consequences of
its being done by everyone who is making a serious choice about whether to do it would be good [bad].²¹ He calls this view "cumulative-effect utilitarianism", but it seems the same as what is usually referred to as "utilitarian generalization", so for convenience I shall sometimes refer to Harrison as a "generalizer". At one point, Harrison considers whether an act-utilitarian could justify the same behavior that cumulative-effect utilitarianism requires, by an argument based on probability. On its face, act-utilitarianism would sometimes require individual defection from a generally useful practice. But Harrison imagines the act-utilitarian arguing that each agent should do what it would be best that everyone do in order to avoid the risk that his act might occur just at a margin where, given other people's behavior, his defection would undermine the good to be achieved by general participation or bring about the harm to be prevented by general avoidance.

Harrison attempts to answer this argument by saying that there is no risk. He focuses on the proverbial case of piling straws on a camel's back. Assuming there are enough people standing around waiting to add a straw so that everyone's doing so would break the camel's back, cumulative-effect utilitarianism calls for general avoidance. But according to Harrison, any act-utilitarian in the group must add his straw (if he has any positive reason to do so), because it is "quite untrue that anyone's putting one single straw on the camel's back ever does the camel any harm".²² It does not matter how many other people are putting their straws on the camel's back: "The camel's back is never sensitive to individual straws . . . . Those who suppose that there must be [straws which individually do some harm] simply do not grasp the fact that the effects of a number of repetitions of an action, which individually do no harm, can do harm."²³ And more to the same effect.

But Harrison's "fact" is no such thing. Harrison is in the position of someone who asserts all the premises of a sorites argument, and then denies the conclusion, with no further comment. If no one puts a straw on the camel's back, the beast is unharmed. According to Harrison, no individual straw produces any effect. ("There is a threshold such that stimuli of certain degrees of magnitude which fall below it produce no effect.")²⁴ So, the first straw produces no effect. Therefore the condition of the camel is exactly the same after the first straw is added as when there were no straws. Similarly, the second straw produces no effect. Therefore the condition of the camel is exactly the same after the second straw is added as after the first, which means, by the transitivity of "exactly the same", that the condition of the camel is exactly the same after the second straw is added as when there were no straws. And so on, until the camel's back is broken, but the condition of the camel is still exactly the same as when there were no straws.

The sorites is a remarkably durable paradox (perhaps because, as Crispin Wright has suggested, it is really a number of paradoxes similar in form),²⁵ and there
is no general agreement about how to dissolve it. But surely the one thing we cannot
do is just assert all the premises and deny the conclusion and say no more about it.
Even Michael Dummett, one of the few writers on sorites who does not attempt to
explain it away, at least offers to explain it by the assertion that the observational
predicates involved in the versions he discusses are incoherent.26

Almost everyone who discusses sorites proposes to dissolve the paradox by
finding some fault or other in the induction premise – the premise that moves us
step-by-step along the chain – in Harrison's version, the premise that no single straw
has any effect.27 And it seems obvious that the induction premise is where the fault
lies in Harrison's version of sorites. It is not true that no straw has any effect. It may
be that there is no straw that breaks the camel's back – that depends on how we
understand "broken" in this context. But it must be (logically) that some straw has
some effect, and indeed it is certain empirically, even if not logically, that many
(individual) straws do. At a microscopic level, any number of straws (individually)
increase the strain on the camel's bones and sinews; and any number of straws
(individually) change slightly the probabilities that the camel will be unable to stand
up under his burden or will stagger and collapse in mid-stride as a rock moves under
his foot.

The reader may still wonder how such tiny effects could matter to what an
agent ought to do. But we should remember, once we have blocked Harrison's
argument, that if the total harm from the whole chain of straw-additions is greater
than the total benefit from them, then there must be some individual straw-addition
where the marginal harm is greater than the marginal benefit. If this seems
implausible given the small size of the harms, we can increase its plausibility by
inquiring into the size of the benefits. If the benefit from each addition (or almost
every addition) is extremely small, then it is not so surprising that at some point even
a small marginal harm should outweigh the benefit. On the other hand, if the
benefits are not extremely small almost everywhere (that is to say, if a substantial
percentage of the straw-additions produce non-trivial benefits), and if it takes a lot of
straws to break the camel's back (as the "tiny effect" objection presupposes), then the
total benefit may well be such that it is best on balance, from a consequentialist
perspective, that the camel's back be broken. (This may sound odd, since we might
assume that the benefit of adding one's straw depends on ending up with an
ambulatory camel. But in fact that assumption would be inconsistent with another of
the generalizer's assumptions, usually left implicit, that the benefit is independent of
how many others add their straws.)

There is one point at which Harrison seems to recognize his overreaching.
In the middle of a passage where he repeatedly insists that no individual action has
any effect at all, he says just once that no individual action has any "appreciable"
effect\textsuperscript{28} -- which is, of course, a completely different claim. This new claim suggests two further arguments against act-utilitarianism. One is the "tiny effect" objection we have just answered. The other involves arguing, not for an objective act-utilitarian prisoners' dilemma, but rather for what we might call a subjective act-utilitarian prisoners' dilemma, where each agent's unequivocal subjective obligation is to behave in such a way that the set of all agents produces sub-optimal consequences. The argument in the camel case is clear: Even if some individual actions must have some bad effects, and even some net bad effects, no individual action ever has enough of a net bad effect to justify worrying about it. Hence every agent is subjectively required to defect from the desirable general pattern, and the pattern is not, cannot be, achieved.

There is an answer to this argument, which both Parfit and I have suggested (although I hid it away in a footnote\textsuperscript{29} and Parfit discusses only one special sort of case, voting).\textsuperscript{30} Let's do a simple but general example quickly and crudely. Suppose there are a thousand people, and the participation of some large number of them is needed to produce a benefit with value B. (We don't need to be precise about the number, nor is it necessary that there should be a perfectly sharp threshold.) Each individual can produce a benefit with value b* by defecting. This is the sort of case in which the generalizer is likely to think that everyone should participate because that is what it is best for everyone to do. This requires, of course, that B > 1,000 b*. On the other hand, the generalizer argues that from the point of view of each individual, the likelihood that his not participating will cause any significant loss in the overall benefit is very small. So, subjective utilitarianism requires him to defect. But if everyone follows this command of subjective act-utilitarianism, the result is disastrous.

In response, the first and crucial point is this. Subjective act-utilitarianism should not be understood as requiring the agent to maximize the likelihood that his act will be objectively right, which might indeed mean that he should defect. Rather, it should be understood as requiring him to maximize his act's expected utility.\textsuperscript{31} Now, we cannot give a precise calculation for the expected value of participating without definite assumptions about what other people are likely to do and about what levels of participation secure how much of the total possible benefit B. But let us assume crudely that it is equally likely that no one else participates, that one other person participates, that two others participate, and so on right up to nine-hundred ninety-nine others participating. (This assumption is in some respects quite implausible, but the reader who follows through the argument with this assumption will be in a position to see that more realistic assumptions will often strengthen the subjective act-utilitarian case for participation; and that when they do not, it will often be because other people's likely behavior and the value-achieved schedule are
such as to make the generalizer's recommendation of participation unrealistically optimistic about others' behavior.) On our crude assumption, the expected value $E$ of participating is $\sum_{i=1}^{1,000} \frac{1}{1000} b_i$, where $b_i$ is the marginal benefit secured by being the $i$th person participating. So, $E = \frac{1}{1,000} \sum_{i=1}^{1,000} b_i$. But $\sum_{i=1}^{1,000} b_i = B$ (the total benefit from universal participation), by an argument like the argument for the summation principle (S) at the beginning of the preceding section. So $E = \frac{1}{1,000} B$. But we saw above that if the generalizer's prescription is to be plausible at all, it must be that $B > 1,000 b^*$, where $b^*$ is the benefit from each individual act of defection. It follows that $E = \frac{1}{1,000} B > b^*$, so the expected value of participating ($E$) is greater than the benefit from defection ($b^*$), and subjective act-utilitarianism requires participation.

There is a puzzle about the stability of this conclusion in a society of act-utilitarians. If everyone is an act-utilitarian, and if it is understood that for reasons such as I have just indicated subjective act-utilitarianism requires participation, then the probable level of participation is very high. And if that is the case, then the expected value from one individual's participation may decline below $b^*$. So subjective act-utilitarianism now requires defection. And unfortunately, if everyone is an act-utilitarian who understands that subjective act-utilitarianism requires defection, then the defection rate will be high, the expected value of participation may again be very low, and subjective act-utilitarianism may still require defection. So this is a stable situation. It may seem that subjective act-utilitarianism fails after all. But that conclusion also is too quick. In *Utilitarianism and Co-operation* I pointed out that in cases such as the ones we are discussing there will standardly be at least two general patterns of behavior, in each of which act-utilitarianism is universally satisfied. One or more will be sup-optimal, but one will be the optimal general pattern of behavior. It would appear that this optimal pattern of behavior, all of which is fully act-utilitarian, will also be a pattern to which an appropriately formulated subjective act-utilitarian prescription can stably gravitate. It is not my primary goal to defend subjective act-utilitarianism (or objective act-utilitarianism), but we see that even the question of whether there can be a subjective act-utilitarian prisoners' dilemma is too complicated to afford an easy rehabilitation of Harrison's position. Possibly there are some subjective act-utilitarian prisoners' dilemmas. But it does not seem probable to me that they are so common as to make cumulative-effect utilitarianism generally preferable.
PRECEIVING IMPERCEPTIBLE HARMS

4. CONCLUDING REMARKS

The reader may wonder why, if my object is not to defend act-utilitarianism, I am so insistent that the ex nihilo claim is mistaken. One reason, as we have seen, is to preserve the transitivity of "better than". The other reason is that it is important to me that there cannot be cases where the achievement of optimal results requires non-act-utilitarian behavior (in other words, objective act-utilitarian prisoners' dilemmas). This is not because if act utilitarian prisoners' dilemmas are impossible, then universal satisfaction of act-utilitarianism always guarantees optimal results – which sadly is not true. But at least if act-utilitarian prisoners' dilemmas are impossible, then there is room for some other theory (like co-operative utilitarianism) to guarantee optimal results by the group of agents who follow it, without creating the tension in practical reasoning which would be generated if those optimal results required some agent to behave in a way which produced less than the best possible results individually.

Donald H. Regan
Department of Philosophy
University of Michigan
NOTES


2 Quinn argues even more ambitiously against consequentialism in prudential decision-making.

3 Utilitarianism and Co-operation (Oxford: Clarendon Press, 1980), 54-65. Włodzimierz Rabinowicz, "Act-Utilitarian Prisoner's Dilemmas," Theoria 55 (1989), 1-44, who is not an ex nihilist, makes a number of other ingenious suggestions about how act-utilitarian prisoner's dilemmas might arise; but since in the end he admits that all of his suggestions are either avoidable or seriously controvertible, I shall not pause here over the possibilities.


8 E.g., Parfit, Reasons and Persons, 79 (one might regard my remarks that follow as a spelling out of why we might believe, and how it could be, that "someone can mind his pain slightly less ... even though he cannot notice any difference," as Parfit suggests); Jonathan Glover, "It Makes No Difference Whether or Not I Do It," Proceedings of the Aristotelian Society Supp. Vol. XLIX, 171-190; Torbjörn Tännsjö, "Classical Hedonistic Utilitarianism," Philosophical Studies 81 (1996), 97-115.

9 Utilitarianism and Co-operation, 13-17.


11 Frank Jackson makes a similar point about chains of indiscernible color patches in Perception (Cambridge: Cambridge University Press, 1977), 113 ff.

12 Quinn, "Self-Torturer," 84.

13 It may occur to the reader that the ex nihilist's unavoidable appeal to valuing "the way it feels" at each stimulus would not really commit him to the coherence of this thought if the appeal could be viewed as a step in a reductio of consequentialist reasoning. But there is no reductio. As we shall see in the discussion of Quinn below, the ex nihilist fails to prove anything at all.

14 Quinn, "Self-Torturer".

15 Id. 81.

16 Id. 82.

17 Id. 82-83.

18 Cf. Regan, Utilitarianism and Co-operation, 63-65; Otsuka, "Group Beneficence," 146.


20 Id. 175.


22 Id. 32.

23 Id. 32-33.

24 Id. 33.

26 Michael Dummett, "Wang's Paradox", in Keefe & Smith, *Vagueness*, 99-118

27 Epistemological theorists, for example, argue that there is some particular step that makes the crucial difference to "redness", or "heapness", or whatever, even though we cannot know which step it is. E.g., Timothy Williamson, "Vagueness and Ignorance," in Keefe & Smith, *Vagueness*, 264-280. Supervaluationists argue that there is some step that makes the difference, even though it is not true of any particular step that it is the one. E.g., Kit Fine, "Vagueness, Truth and Logic," in Keefe & Smith, *Vagueness*, 119-150. Degree-theorists argue that the induction premise is not completely true at any step, but only almost true everywhere. E.g., Dorothy Edgington, "Vagueness by Degrees," in Keefe & Smith, *Vagueness*, 294-316. (It is not actually true that I have read nothing about vagueness and sorites elsewhere than the Keefe & Smith volume. But it is a collection of extraordinary quality, probably the best philosophical "reader" I have encountered.)

23 Harrison, "Cumulative-Effect Utilitarianism", 33

29 *Utilitarianism and Co-operation*, 231-232, note 6

30 Parfit, *Reasons and Persons*, 73-74

31 If it is not clear how these requirements differ, see Regan, *Utilitarianism and Co-operation*, 264-265, note 1, and 231-232, note 6.

32 Id. 12-53.

33 Id. If Michael Otsuka is denying this when he says "it is impossible for the maximization of the positive difference each individual makes to result in a state of affairs in which the benefits produced by all of humanity are not maximized", he is mistaken. Otsuka, "Group Beneficence", 148-149.