

## 24.231 Ethics – Handout 12

### My notes, Norcross, “Comparing Harms: Headaches and Human Lives”

**Consequentialism:** (V1) An action is morally *permissible* if and only if (iff) produces at least as much good as all the available alternatives.  
Or: (V2) An action is morally *permissible* if and only if no alternative action produces more good.  
(Also: An action is morally *required* if it produces more good than any available alternative. An action is morally *impermissible* if it is not permissible.)

**Deontology:** An action is sometimes *not* permissible even when it produces more good than all the available alternatives. In other words, the right action might not always be the one that produces as much good as possible.

### Problem Case for Consequentialism:

***Life for Headaches (LfH):*** There is some finite number of headaches such that it is permissible to kill an innocent person to avoid them.

*Deontologists* can reject LfH by appealing to agent-centered restrictions or side-constraints: they allow that there are some things we may not do, even for the sake of bringing about more good.

*Consequentialists* cannot respond in this way. To avoid LfH, they must argue that a world that that contain a vast (but finite) number of headaches *cannot be worse* than a world that differs from it only in lacking the headaches and containing one more premature death.

### How might a consequentialist argue for this conclusion?

- I. *First suggestion:* Adopt ***Incomparable***: the badness of any number of headaches is incomparable to the badness of a premature death – that is, it is neither better than, nor worse than, nor the about the same as the death.

*Problems:* If we accept *Incomparable* and (V1) of Consequentialism, we’ll be able to claim that killing the person to prevent the headaches is impermissible, since it will be false that killing the person is as good as or better than not killing him. But it will also be impermissible allow him to live, because again, it will be false that doing so will be as good or better than killing him. So *Incomparable* plus (V1) generate a moral dilemma; but if we instead accept (V2), adopting *Incomparable* won’t allow us to avoid LfH: killing the person will turn out to be permissible, since there’s not alternative action that will produce *more* good.

II. Norcross concludes that a consequentialist should resist LfH by arguing for

**Less:** The state of affairs constituted by *any* number of fairly minor headaches is less bad than even one premature death of an innocent person.

*Problem:* But Norcross argues that *Less* is problematic. It may seem plausible that one premature death is worse than any number of mild headaches. But he asks us to consider a series of judgments, S: surely, although a death is worse than a severe mutilation, there will be some (possibly large) number of severe mutilations ( $n^1$ ) that together are worse than a single premature death. And there will be some much higher number ( $n^2$ ) number of slightly less severe mutilations that are worse than  $n^1$  severe mutilations. And there will be some again much higher number ( $n^3$ ) of again slightly lesser harms that will be worse than  $n^2$  less severe mutilations. And following this chain of reasoning, we'll eventually get to a harm, perhaps a sprained ankle, such that some huge number ( $n^{m-1}$ ) of sprained ankles is not as bad as some much vaster number ( $n^m$ ) of headaches. But then by the *transitivity of better-than*, it will follow that since  $n^{m-1}$  ankle sprains are worse than  $n^m$  headaches, and  $n^{m-2}$  broken ankles are better than  $n^{m-1}$  ankle sprains, and so on, that  $n^1$  deaths are after all better than  $n^m$  headaches. So we'll have to reject *Less* after all.

III. Perhaps consequentialists should instead reject *the transitivity of better-than*.

*Problem:* but rejecting the *transitivity of better-than* will introduce the possibility of moral dilemmas, which consequentialists might not be happy with (it was one of the appealing features of their view that it always gave us an answer about what to do). And in any case, the *transitivity of better-than* seems more plausible than *less* does: we sometimes do seem to reject *less* in our actual reasoning (speed-limit example).

IV. Arguments for the *intransitivity of better-than*:

**Quinn's Self-Torturer example:** perhaps a lot of individually imperceptible differences in pain levels, whose disutility at each step can be outweighed by a large monetary prize, together add up to a level of pain that itself cannot be outweighed by any monetary prize. At each step, the imperceptibly greater amount of pain plus the prize is better than the previous amount of pain. But the total pain at the end, plus all the prize money, is not better than the beginning amount of pain.

*Replies:* Should we accept the claim, on which the example depends, that each move to a new level makes no difference in the subjective level of pain experienced by the ST? Clearly, each step *must* make a difference to the objective level of pain. If each level involved no additional pain over the previous level, the ST could reach level 1000 and still feel no pain. And the ST can use the knowledge that each move to a higher setting must be increasing the objective damage to his well-being to take that damage into account in his decision-making.

What if we allow that each move to a higher setting involves a cost in well-being, only a very very small one. The prize of \$10,000 surely would offset such a small cost. (Compare the decision to eat one more cookie when I'm trying to lose weight. Each cookie does only imperceptibly small damage to my weight-loss goals. But the benefit in pleasure I get from each cookie is definitely perceptible.) Won't the paradox still arise?

Not necessarily. One thing the example leaves obscure is the marginal utility curves of the pain and the reward. We can imagine that the marginal utility curve of the pain is flat – perhaps that what it is to be pain. But what about the money? Money plausibly has a diminishing marginal utility. So each further installment does less to improve our well-being. Why is it ever better to take the prize in exchange for moving to the next setting, give that (as ST knows, though he can't tell) each higher setting leaves him in one respect slightly less well off? It's better because, presumably, the money allows ST to buy things that improve his well-being enough to offset the (very small) loss involved in moving to the next setting. (If we're simple utilitarians, we can imagine the money can be converted into pleasure, which offsets the pain.) But given the diminishing marginal utility of wealth, each new \$10,000 will outweigh the pain by less. Eventually, the additional pleasure to be gained by another \$10,000 will itself be imperceptible, and ST will no longer want to accept the deal – and we'll reach a point where the imperceptibly small benefit of the prize will in fact be outweighed by the imperceptibly small cost of moving to the next level of torture. Can't we adjust the example, to keep the marginal utility function of the prize-money flat as well, e.g., but steadily increasing the reward for moving to the next level? Imagine we do that. Now each utility curve (of the pain and of the prize money) is flat. In that case, if it's ever best for ST to take the prize, it will always be best for ST to take the prize. ST will never reach the point, even at level 1000, where it's a mistake to take the prize money – after all, all that pain will be offset by so much pleasure. Maybe this will seem implausible – maybe it seems to us that if ST is in a certain amount of pain, he can't really enjoy the pleasure. But this is just to say either that the marginal utility of the prize will decrease after all (it won't be able to generate as much pleasure as ST moves up the settings), or that the marginal disutility of the additional levels of torture is increasing – in addition to causing pain, it's increasingly hampering ST from enjoying pleasures, so each move to a new level of torture will damage him more than each previous move. In that case, again, there will be a point at which the functions cross – the utility to him of getting the prize will no longer outweigh the utility to him of not moving to the next level. And the circle is one again broken. (Of course all this is compatible with ST's being unable, because of the imperceptibility of the relevant differences, to exactly pinpoint where the functions cross, and it makes sense for him to turn down the offer.) It seems to me that the example is playing off of our instinctive sense of the diminishing marginal utility of wealth (that's how it gets us to react to the case of level 1000) without acknowledging that this breaks the circle and allows us to avoid intransitivity.

## **V. Temkin's argument for intransitivity:**

Temkin argues that intransitivity follows from three independently plausible claims:

**Claim 1:** for any unpleasant or "negative" experience, no matter what the intensity and duration of that experience, it would be better to have that experience than one that was only a little less intense but twice as long. **Claim 2:** there is a continuum of unpleasant or "negative" experiences ranging in intensity, for example, from extreme forms of torture to the mild discomfort of a hangnail. **Claim 3:** a mild discomfort for the duration of one's life would be preferable to two years of excruciating torture, no matter the length of one's life.

If we accept these claims, it will again be possible to imagine a chain of comparison-claims, beginning with the claim that excruciating torture for two years is better than slightly less excruciating but still terrible torture for 4 years, and ending with the claim that, e.g., a somewhat uncomfortable hangnail for many thousands of years is better than a very slightly less uncomfortable hangnail for twice that long. But the two years of excruciating torture are surely not better than having the hangnail for one's whole life, no matter how long the life. So better than must be intransitive.

*Replies:* Norcross says Claims 1 and 2 seems plausible. But 3 holds up less well on reflection. The torture looks really bad to us, in part because 2 years is a significant portion of our lives. But can we really form a reliable judgment about how bad it would be if it occupied a period of, say, a 32 million-year-long life that compares to 2.5 minutes of our life? And can we really judge how bad a hang-nail that we never get used to would be if it lasted 32 million years?! (Also, we may be judging the two years of torture to be particularly bad because we think we could not meaningfully survive it – we'd go crazy or lose our identity. But that would break the continuum assumed in Claim 2. And we may be judging the hangnail not so bad because we imagine if we had one for years we'd get so used to it that it would stop bothering us. But that, again, is not what we're supposed to be imagining.) Norcross says, "I conclude that we have no good reason to suppose that any possible pain experiences could span the pain spectrum from torture to hangnail in few enough small enough steps that A and Y were 'merely' a few hundred (or even thousand) years long." And beyond that, our intuitions leave us at a loss.

Can 3 be argued for directly? Norcross: "If Temkin is to explain why A [e.g. two years of torture] is worse than Y [a long life of hangnail], rather than simply assert that it is, he must appeal to a property possessed by A but not by Y (e.g., having a badness score of more than 1, being miserable). Given the 'other things being equal' clause that applies to all the lives in the sequence, the possession by A but not by Y of the property is sufficient for the judgment that A is worse than Y. There will, however, be a first life in the sequence that does not possess the property. Therefore, there will be an adjacent pair of lives, say m and n, such that m does but n does not possess the property. Therefore, n is not worse than m. But this contradicts claim 1."

## **VI. Defending LfH:**

All this goes to show that we should not reject the transitivity of better-than. But the judgments in S are overwhelmingly plausible. So we should instead reject Less, after all. But rejecting Less and accepting Consequentialism leads to LfH. So we should either

reject Consequentialism or bite the bullet, and embrace LfH. Norcross suggests that the right approach is the latter:

Consider that we raise speed limits for the sake of the added convenience of many people, even though doing it will cause a few more deaths. According to *Lives for Convenience* (LfC), we have no moral obligation not to raise the speed limit. But LfC looks just like LfH. Is there a morally significant difference?

Norcross considers several possibilities:

- (i) In LfH, but not in LfC, we know who the victim will be.
- (ii) In LfH, but not in LfC, there is a determinate victim.

Norcross argues (convincingly) that neither difference has *moral* significance, though it may have *psychological* significance. But he seems to think the *psychological* difference can explain why LfH is true, even though it's counterintuitive. It might do this, but it might also explain why we failed to recognize the badness of raising the speed limits (just like it causes us to fail to recognize the badness of the natural disaster that effects anonymous children).

- (iii) Victims voluntarily consent to the risk of being killed by voting for higher speed limits.
- (iv) Victims benefit from having higher speed limits, even if they did not vote for them.
- (v) Victims at least chose to assume the risk of the higher limits.

(iii)-(v) aren't true of all such victims. And anyway, if we built similar choice or benefit assumptions into the headache example, would it change our intuitions about it?

- (vi) We're not sure raising speed limits will kill people. (A non-starter.)
- (vii) When we raise speed limits, our actions allow more people to die. But when we trade a human life for headaches, we *kill* someone. There's a morally significant difference between killing and letting die.
- (viii) When we raise speed limits, it's an *unintended* but foreseen consequence of our actions that people die. When we trade a human life for headaches, we *intend* to kill. An *intended* death is morally much worse than a *merely foreseen but not intended* death. (Doctrine of Double Effect)

But, Norcross argues, for this to make the difference, killing will have to be *lexically* worse than letting die, not just *proportionally* worse, and intended deaths will have to be *lexically*, not just *proportionally*, worse than merely foreseen ones. And it's implausible that killing or intended deaths are *lexically* worse. (*But*: We could decide that killing/intended death is proportionally worse, but *much, much* worse, than letting die/merely foreseen death. We often in fact reason like this, since it's in most of our

power to prevent many deaths all the time by donating money to charity, but we don't think we're guilty of anything *like* mass-murder. Then we might conclude that LfH is true, but that the number of headaches required is *very very* high – too high for us to be able to intuitively assess their badness. For all intuitively assessable cases, we ought never to trade a life for headache relief.)

Also: would our intuitions about LfH change if the life exchanged is one we merely allow to end, rather than actively take away? Consider the case of *Fred*. (*But*: also consider deciding how to distribute a drug that can either save one person's life or cure any number of mild headaches. We all the time decide to fund drug research into cure for, e.g., athlete's foot, but wouldn't fund research into drugs that might cure a disease so rare it claims only a single victim. Is this wrong? What makes this case different from that of Fred?)

Perhaps we can question the causal claim – higher speed limits don't cause more deaths, bad driving causes them.

Perhaps restricting the speed limits too much costs us more than mere convenience – it costs us our freedom to live and act as we choose (consider extending such paternalistic laws indefinitely).

***Norcross's Conclusion: LfH is not a decisive strike against consequentialism.***

MIT OpenCourseWare  
<http://ocw.mit.edu>

24.231 Ethics  
Fall 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.