

## 24.231 Ethics Handout 15 Singer, “Is Act-Utilitarianism Self-Defeating?”

**Act-Utilitarianism (AU):** An act is right if and only if it would have the best consequences, that is, consequences at least as good as those of any alternative act open to the agent.

(This is really a statement of act-*consequentialism*, not act-*utilitarianism*, but we can assume, for the sake of evaluating the argument, that what makes consequences best is that they involve as much total welfare as the consequences of any other available act.)

**Hodgson’s charge:** “to act upon the act-utilitarian principle would probably have worse consequences than would to act upon more specific moral rules, quite independently of the misapplication of that principle”

(c) D. H. Hodgson. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/fairuse>.

Hodgson, D. H. *Consequences of Utilitarianism*. Oxford, United Kingdom: Oxford University Press, 1967, p. 3. ISBN: 9780198243120.

*First Question:* If Hodgson’s charge holds up, how damaging is it to AU? It establishes that people’s accepting AU would not have the best consequences, if that led them to apply it, even if they applied it correctly. But this does not, at least not obviously, show AU to be false. It merely shows that if AU is true, then we shouldn’t believe it, or at least, shouldn’t (always) live by it.

### **Hodgson’s argument:**

- (1) If AU is true, then truth-telling and promise-keeping (as they exist in our world) are valuable practices. They allow us to communicate information and to make fixed plans.
- (2) The value of truth-telling and promise-keeping depends on our expectations that people (except in unusual circumstances) will tell the truth, and keep their promises.
- (3) In a society of perfect act-utilitarians, people will tell the truth and keep promises only when doing so will have the best consequences overall.
- (4) But truth-telling and promise keeping will have the best consequences overall only if the addressee has reason to expect the speaker to be speaking truly or promising faithfully.
- (5) So a perfect act-utilitarian agent will only (reliably) tell the truth or make a promise if he believes his listener to have such expectations.
- (6) But a listener will have such expectations only if he believes the speaker to will (reliably) tell the truth or promise faithfully – that is, only if he believes the speaker believes the listener expects him to do so.
- (7) But the listener has no reason to believe this is what the speaker believes (since before the listener’s expectations are set, the speaker has no reason to believe it).
- (8) So the listener will not have any expectation that the speaker will tell the truth/promise faithfully.
- (9) So the perfect utilitarian agent will have no reason to tell the truth or promise faithfully.
- (10) But if AU is true, this would be a very bad consequence.

- (11) So very bad consequences, from the standpoint of AU, would follow from everyone's perfectly adopting AU as their rule of action.

In short, if AU is true, perfect act-utilitarian speakers will tell the truth or keep a promise only if they have reason to believe their perfect act-utilitarian listeners expect them to do so. But those listeners will expect them to do so only if they have reason to believe the speakers will tell the truth or keep the promise. This leaves both parties' expectations and intentions underdetermined. We're at an impasse. The speaker has no grounds for concluding the listener will believe him, and the listener has no grounds for concluding that the speaker will tell the truth. So the speaker will have no reason to do so. That's bad.

### **Singer's Replies:**

- (1) It's hard to imagine a case where a perfect AU agent would have any reason to lie or break a promise.

*Concerns:* Is it that hard to imagine? What about white lies, like "you look lovely in that hat"? Singer might respond that a perfect AU agent wouldn't care how he looks in the hat, but this seems wrong. In order for AU to be able to function as a guide to our lives at all, we must have some preferences not determined by AU. (Otherwise the principle would give us nothing to do.) These preferences introduce the possibility of beneficial white lies. In any case, if the speaker also has no reason to tell the truth, rather than lie, this might not break the impasse.

- (2) There's only one way for a statement to be true, but many ways for it to be false. So I will have a better chance of influencing someone else to do the best thing by telling the truth than by telling a lie, even if there's only a 50-50 chance that I will be believed.

*Concerns:* This seems to be a removable feature of the bus example. What if the question were, instead, "Is there a late train out of the city on Monday evenings?" This question has only two possible answers. Singer suggests that any question could be rephrased to include more than two options, providing the speaker with an incentive to tell the truth. But I'm not sure that's right. I could ask "Is it the case that there's a late train and the moon is made of cheese, or that there's a late train and the moon is not made of cheese, or that there's not a late train and the moon is not made of cheese?" Now there is more than one way for the answer to be false, but one of those ways does not run the risk of influencing my action in the wrong way, and so can be disregarded by both speaker and listener, leaving us again with just two options.

- (3) Singer sometimes seems to suggest that even a perfect utilitarian agent would be at least a fraction more likely to tell the truth, thereby breaking Hodgson's described stale-mate.

*Concerns:* But why think so?

- (4) Singer suggests that the promise-keeping case be treated as a special case of a the truth-telling case: to make a faithful promise is just to truly express a strong expectation, based on your intention, that you will perform some action (unless something very unexpected occurs). Singer suggests that so long as the listener can believe the speaker, she can get everything that's valuable out of the practice of promising, since she can interpret the speaker's "I promise to x" as "I will do x so long as doing x doesn't have worse consequences than not doing x, taking into account the effects on you of my failing to do x, and I know believe doing x will have the best consequences, so I predict that I will do x, and I realize that you will therefore expect me to x."

*Concerns:* Is this as useful as our ordinary practice of promising? Part of what's at issue is what the effect on the listener's expectations will be of making such a "promise." Will it be any different than merely saying "I'll do x if I think it best, and right now I think it best, though that's subject to change"? I don't see how. But the listener's expectations in this case will be quite different (and less fixed), it seems to me, from the case of an ordinary promise. So the practice of promise-keeping in the perfect AU society, even assuming truth-telling to be possible, is quite different from our actual practice. It's really more like a practice of making predictions. And that might be less valuable than our own practice.

- (5) Singer argues that the value of the practices of promise-keeping and truth-telling will provide just the sort of reason perfect AU agents need to decide to tell the truth or keep promises, and so break the stale-mate described by Hodgson. Their reason for telling a truth or keeping a promise will be that doing so will contribute to a useful practice.

*Concerns:* As Hodgson says, while AU may give us reasons to develop such practices, "such steps ... , although perhaps justified by AU, ... would amount to a partial rejection of AU and so would be inconsistent with our assumptions. These steps would amount to a partial rejection of AU, because the persons would be forming habits to do acts know not to be justified according to AU; and they could form these habits only if they resolved to refrain from applying AU in relation to these acts."

Two worries: (i) even if the contribution to a beneficial practice could give us a small reason to keep a promise, this reason (since one broken promise would do *very little* to undermine the practice on its own) may be a very weak reason, easily outweighed by any positive utilitarian considerations not to keep a promise; (ii) any practice of promise-keeping that allowed for breaking promises whenever doing so is best on the whole will (taking into account the effects of doing so on the survival of the practice, would look little like our own practice; and in any case, such a practice would *not* be undermined by breaking a promise whenever doing so is best on the whole. See my concerns for (4) above. (Note that our actual practice of promising is not *absolute* – it allows for the breaking of promises when the results of not doing so are *much* worse from an AU perspective.)

Finally, Hodgson's point here, as I understand it, is that we might have good AU reasons for developing *habits and dispositions* which will lead us to act counter to AU reasons. This will be true of the consequences of our having such dispositions are on the whole better than the consequences of our not having them, even if our having them will lead us, sometimes, to act in ways that are not ideal. A disposition that is a part of the set of ideal dispositions may sometimes lead us to act non-ideally, but that doesn't mean it would be better to abandon the disposition. Hodgson has in mind the disposition to tell the truth or keep promises, even if doing so is a bit worse, all-things-considered, in a particular case. But the point might be more clearly made by appealing to a different AU-desirable disposition: the disposition to love our children. Having this disposition seems clearly better, from the standpoint of AU, than not having it. But having it may lead us, in some cases, to give some benefit to our children instead of giving a significantly greater benefit to a stranger. In this case, it would be better if we gave the benefit to the stranger. But we could be motivated to do this only if we lacked the disposition of loving our children. And it would be very bad, according to AU, if we lacked this disposition (much worse than it is for us to sometimes give undue priority to our children's interests). This suggests that there are some dispositions which it is good for an AU to adopt, but which may sometimes lead him to act wrongly. Nonetheless, an AU should adopt them. So it is better, if AU is true, if we weren't always motivated only by the AU principle, even if we were applying it correctly.

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>.