Two Mistaken Strategies of Deontology

Abstract

Two distinctions between parallel strategies commonly used by deontologists and consequentialists to derive or defend their respective views have been almost completely ignored in the literature on both. The first is a difference between the type of moral universalizability test used by each, the second is a difference in testing the rationality of a principle (whether universalized or not). Once explicitly identified, the strategies used by deontologists are easily seen to be defective compared to those their consequentialist counterparts.

Paper

Two distinctions between parallel strategies commonly used by deontologists and consequentialists to derive or defend their respective views have been almost completely ignored in the literature on both. Once explicitly identified, the strategies used by deontologists are easily seen to be defective compared to those their consequentialist counterparts.

The first distinction is between two kinds of universalizability test. Many philosophers mistakenly associate universalizability tests with deontology, and more specifically with Kantian ethics (cf. Jens Timmermann 2005: 254 note 27). Actually consequentialists frequently claim that their norms are uniquely universalizable—cf. Peter Singer (1981: 229-30; 1993: 11-14), Richard Hare (1963; 1981), David Cummiskey (1996), and Phillip Pettit (2000)—often using the language of the golden rule or moral supervenience. Their differing conclusions are partly explained their appeal to two very different types of universalization tests. The deontologists' universalizability test requires that moral principles be those which a rational agent can will (value, endorse, approve, etc.) *all* agents following simultaneously, cf. the question "What if

everybody did that?" Consequentialists' universalizability tests instead require that moral principles be those which a rational agent can will that *any* agent follow, in any situation (including cases where not all other agents are following the same principle); cf. the question "What if somebody did that to you?" We may distinguish these as universal practice (UP) and universal approval (UA) tests, respectively. (Note that a common version of rule consequentialism uses a UP test—and supports some deontological norms as a result.)

UA tests are stronger than UP tests, for their test conditions include the ideal worlds of universal conformity which the latter appeal to, as well as many more realistic ones. Hence they are intuitively more plausible and relevant, while UP tests have problems identifying sound principles for non-ideal situations, as when confronting evildoers, or solving coordination problems.

The vast logical gulf between these tests can be made invisible through at least two strategies, both of which are commonly used—perhaps quite unconsciously—by defenders of UP tests. The first is to use negative (failing) examples; for since the UA condition includes the UP condition as a limiting case, any principle which fails the latter will also fail the former. But the reverse is not the case, so passing a UP test is a necessary condition of morality, not a sufficient one; sadly the history of Kantian scholarship is strewn with instances of this elementary logical fallacy. The second strategy is to use examples of principles which produce certain harmful effects (in virtue of which we intuitively judge them to be immoral) in linear proportion to the number of agents following them. For instance, if one person litters, hits or kills another agent, lies (while promising or otherwise), or refuses to help another agent in need, some predictable harm ensues; and if a thousand or a billion agents do the same, roughly a thousand or a billion times as much total harm will ensue, making it increasingly likely that any agent taken at

random—including, pointedly, any single agent considering acting on such a principle—will suffer from it. If our agent imagined that he could put up with a few other agents acting as he does, given the low likelihood that he will be harmed thereby, a UP test puts that harm in a starker light by showing the magnitude of harm he would suffer if he was guaranteed that all agents were doing the same as he, making escape from the ill-effects of his principle impossible.

While stress on the deprivational effects of a universalized practice can highlight the harm caused by even a single instance of it, a huge problem with this approach is that many principles cause harm at non-linear rates. E.g., principles of absolute non-violence or always driving on the left side of the road cause no harm at all when everyone is following them, but lead to utter disaster when some people do but others reject them. Hence these easily pass UP tests, but only pass UA tests if further qualified, requiring different behavior in response to the predatory violence of others, or to a local practice of driving on the right.

Both deontologists and consequentialists have routinely confused these two tests, treating them as interchangeable or even identical, even while relying upon one or the other form in their actual derivations of moral norms. A common approach is to begin by describing the requirements of morality as a UA test, then claiming to derive a UP test from this, as if the sufficiency of the former shows the sufficiency of the latter, or as if they were not even distinct. The most infamous, though unrecognized, example of this is Kant's famous derivation of the formula of universal law (FUL). He begins by describing the sought-for categorical imperative as a principle *for an individual will*—i.e., governing our judgments or evaluations of behaviors as good or bad according to some universal law [my emphasis in all quotes below]:

> "[the categorical imperative must be a] principle of volition [or] *of the will*" (G 4:399-400) "which alone is to *serve* the will as *its* principle" (G 4:402) "[and] which contains *no conditions* to which it could be limited [beyond] the universality of a law as such." (G 4:421)

But then this immediately becomes a principle whose universalization ranges, not over our acts of will, valuation, or judgments, but over a group of persons in a world:

> Act only in accordance with that maxim through which you can at the same time will that it become a *universal law*. (G 4:421)

Kant's examples of how to apply this law (G 4:423-424) amply demonstrate that what he means is that it must be possible both for all persons to follow the maxim, and to thereby satisfy the ends thereof, so the universalization ranges, not over an agent's will that some or any agent follow the maxim, but over the behavior of all agents at once, whose universal compliance the agent must be able to will as a single unitary state of affairs. In other words, after insisting the principle should be limited by no conditions, he limits it severely by saying that it must consider only how an agent would fare if all other agents were following it simultaneously—a bizarre situation we are almost never going to be in.

Many Kantians have followed him in this error, either again inadvertently making the UP-UA slide in their own language, or sliding from the claim that passing the UP test is a necessary condition of morality to claiming that it is a sufficient condition, or simply by describing Kant's test as if it was a UA test, and then naively asserting that this demonstrates the adequacy of his (UP) FUL test.[1]

---

[1]   "[it follows from the requirement that if an agent acts on some principle] then it must be possible for him to will that [everyone do so], that the test of the moral acceptability of your action is whether…you could will that everyone follows such a law [of doing this action]." (Patricia Kitcher 2004: 571)

"The maxim of your action…must contain nothing which would prevent that maxim from serving as an unconditioned practical principle: as a universal law for all beings" (Barbara Herman 1990: 170)

"To regard one's reason for acting in a certain way as good is to assume its legitimacy for all rational beings in similar circumstances…[hence] rational agents cannot reject the universalizability test [of the FUL]" (Henry Allison 1990: 205)

"[Kant's formula] 'Act only on that maxim through which you can at the same time will that it should become a universal law' [means that] I have, that is, to will it not only for the present situation, in which I occupy the role that I do, but also for all situations resembling this in their universal properties, including those in which I occupy all the other possible roles. (Richard Hare 1997: 153)

"[A maxim is authoritative if you can will that any other agent] "no matter how situated" (Andrews Reath 2006: 73) regard it as giving a sufficient reason for action, or "more precisely, [if] you can regard the maxim as

G.E. Moore (1903: 161) and C.D. Broad (1916) are among the few authors who have explicitly distinguished between these two kinds of universalizability as a criticism of Kantian ethics. Neither criticism has been either adequately answered by Kantians, or (perhaps even more strangely) taken up at large by other consequentialists. However it has long been the basis of criticism against ideal rule consequentialism (IRC), the theory that we should follow those principles whose universal practice would create at least as much good as the alternative practice of any alternative principles would—this is a UP test which unsurprisingly supports many deontological norms. IRC has been accused of "rule worship" in cases where an individual's following the principle does harm precisely because not everyone else follows it (J.J.C. Smart 1973: 10). In response, Richard Brandt (1992) and Brad Hooker (2000) proposed a modified IRC which produces at least as much good as any alternative rule in the test condition of 90% conformity to the principle within some population, forcing such principles to give us utility-maximizing directions for responding to non-conformists. But this particular test condition is just

---

stating a sufficient reason for action while willing that everyone regard it as stating a sufficient reason for action, without inconsistency" (2006: 107)

"Every rational agent must will in accordance with a universal law…[which] ranges over all rational beings, that is, it commands you to act in a way that any rational being could act, because you could find yourself in anybody's shoes, anybody's at all, and the law has to be one that would enable you to maintain your integrity, in any situation, come what may." (Christine Korsgaard 2009: 214)

"[An] end or action [passes the FUL just when it] is implicitly understood to be such as would be appropriate for any subject capable of practical knowledge, provided that it is in the conditions on which the judgment is based." (Stephen Engstrom 2009: 124-125)

The error is not unique to Kantians and their critics:

"When we evaluate rules according to second-order standards, we judge the effects of their being followed. We consider what it would be like if one rule rather than another were generally adopted as a norm for right conduct" (Paul Taylor 1961: 45-46)

"[From the] generalization principle [that] what is right (or wrong) for one person must be right (or wrong) for any (relevantly) similar person in (relevantly) similar circumstances" (Marcus Singer 1961: 19) "[we can derive the] generalization argument [according to which] if the consequences of everyone's doing x were undesirable, then it would be wrong for anyone to do x." (1961: 61)

as arbitrary as universal compliance is, and generates counter-examples of its own, e.g., it would make it forbidden to drive on either side of the road, since doing so if only 90% of the population was driving on the left (or right) would be excessively dangerous. In response, Michael Ridge (2006) suggested a "variable-rate consequentialism" which evaluated principles by averaging their success at producing good across all acceptance rates within the relevant population. Holly Smith (2010) pointed out that principles which passed this test would be precisely those which require different behavior depending upon our knowledge of what principles other agents were following so as to do the most good we could through our choices regardless of what acceptance-rate our principles had throughout society, which essentially transforms IRC's UP test back into a UA test.

A few Kantians have essentially admitted that the FUL's UP test is only adequate precisely for situations in which an agent is, by oneself or in cooperation with others, choosing a general institutional or social practice, and is less adequate for choices outside of such contexts, which might require unspecified "special principles" (Korsgaard 1993: 154; 1996b: 100, 2005: 98). Tamar Schapiro (2003: 333) proposed a fix to this aspect of the FUL by treating certain behaviors as different actions in different contexts—including contexts in which others are not following the same principles as you, e.g., turning an act of "truth telling" into an act of "facilitating a murder." However she seems to think that we can get away with this re-assessment selectively, in certain problematic cases, and does not recognize that if such judgments are made non-arbitrarily, across the board, this again turns the FUL back into a UA test, individuating and hence assessing our actions on the basis of the consequences they will likely produce in interaction with the behaviors of other agents—as consequentialists have been urging us to do all along.

6

Yet another advantage of UA tests over UP tests is visible if we formalize their conditions in symbolic logic. Let Px mean "x follows principle P" and $\mathbf{W}\varphi$ be a modal operator meaning "I consistently will φ to be the case." (Set aside for now what counts as "consistency" or its lack, which is discussed further in the second part of this paper.) The conditions under which P satisfies the two tests are then:

UP: $\mathbf{W}(x)Px$ Where (x) ranges across all agents in a given world.

UA: $(x)\mathbf{W}Px$ Where (x) ranges across all possible agents in all possible worlds.

It is immediately clear from this symbolic formulation that if the UA test is satisfied this implies $\mathbf{W}Pi$ as a simple instantiation of the universal generalization, where "i" names the present agent. But this by no means follows from satisfaction of the UP test. Hence there is no particular reason to use it as a test for whether I can rationally will myself following the principle in question.

One might of course wonder, as some critics of universalizability tests often have, why we must commit to anything beyond $\mathbf{W}Pi$ to make it rational to follow the principle ourselves. While this point goes slightly beyond the scope of this paper, I would suggest that while purely instrumental reasoning need only evaluate our principles in terms of their efficacy towards our given ends, as currently tokened in our behavior, moral reasoning evaluates our principles as types which could also be instantiated in other agents. Approval of my following some principle, that is, approving of myself insofar as I am some agent following it, constitutively involves approving of any agent following the same principle as such. Of course, it is not metaphysically or empirically necessary that we evaluate ourselves in this way. But if we don't, then we are taking a kind of autistic, or even psychopathic view of other agents, and are simply not making

moral evaluations; we can then at best see the behavior of others (and our own) as collections of space-time events which happen to promote or frustrate our ends as the case may be, but cannot favorably (or unfavorably) evaluate our behavior (including our choice of ends) as a whole, nor compare this larger behavioral pattern to that of other agents, to be judged favorably or unfavorably, i.e., morally.

A second distinction lies between two ways of determining when following a principle is irrational. For many deontologists, this occurs when an agent thereby wills (perhaps via universalization) the *impossibility* of satisfying some end (or equivalently: wills the absence of *necessary* conditions thereto). This end is one either presumed to be morally good by prior argument or intuition, or which the present agent purports to be good via their present agential behavior—cf. Kant (G 4:417) and Alan Gewirth. However consequentialists identify irrational universalizability with willing a greater *probability* of frustrating rather than satisfying the agent's own ends, should another agent be following the same principle (Hare 1981: 110-111).

Again, the latter condition is stronger than the former: any means *necessary* to some end is one whose absence makes its satisfaction less *probable*. But the reverse is not true; if any (or even all) other agents follow a principle of rolling ten dice each morning and killing an innocent person that day just if they are all 6's, it is less likely, but not impossible, for me to satisfy any other ends—including of killing someone if my own dice demand it. The probability condition is also intuitively more plausible, and used in most ordinary practical reasoning; it is irrational to use a less effective tool for a job when a more effective one is equally available, even if it remains possible to succeed with the former.

However deontologists typically make their strategy *look* stronger by fabricating and equivocating on "necessary" conditions. By incorrectly claiming that certain arbitrarily privileged means are necessary for satisfying some end—when in fact they are just some instrumentally useful conditions amongst many others—they privilege the provision of those means as obligatory, and their denial as forbidden. E.g., Kant (G 4:422-423) claims that the universal practice of false promising or benevolence will make it impossible to satisfy certain ends, when actually this is just made less likely; after all, cosmic rays or a fit of madness could hit the potential lender to make her strangely compliant to our desires, or our ends might be satisfied without ever needing the help of others. Ironically, this strategy can actually be made to look even stronger when conjoined with the weak UP test, by focusing on the role of the intentional activities of other agents as means to our ends. If we failed to get a desired loan or some other help from one person (through bad luck, or because our reputation preceded us), we might hold out hope that some other person might come to our aid; imagining a world in which everyone falsely promises, or is rigorously selfish, seems to close off all such hope. But this ignores the fact that other (possibly quite unusual) circumstances could conspire to satisfy our ends. However, Kant's examples are rarely challenged with these sorts of counter-examples, and so his claims that the universal practice of certain principles makes it truly impossible to satisfy certain ends by depriving us of their logically necessary conditions are too often glibly accepted.

Kantians have often tried to privilege the value of rational agency and its associated capacities or prerequisites over all other values because they are supposedly necessary for successful action, while all other means are merely helpful or conducive to various ends. We can agree that rationality is pervasively helpful for successful satisfaction most other ends—including the end of choosing morally good ends. But this does not mean that the former

9

infinitely surpasses all other means in value, trumping them in every contest regardless of the degree of infringement of or risk to each.

Quasi-Kantians like Alan Gewirth and Michael Smith have each made such arguments to privilege certain means over others. Gewirth (1978: 63-64) insists that "freedom and well-being" are absolutely privileged ends which each agent must treat as "rights" he deserves from others and must accord to them in turn (here, using a straightforward UA argument). But in practice he equivocates wildly as to when our moral rights—such as the right not to have cancer inflicted upon us—are absolute (1982: 186-189) and when they are just values to be traded off against each other (1982: 195). Incidentally, Gewirth (1998: 85n24) is one of the few authors who hints (inadvertently) that his use of the necessity condition may be due to a confusion between the rational necessity to act morally, and the idea that moral demands must be based on "premises that themselves also embody such necessity." The two necessities are actually quite different; it could be morally necessary to help improve the probabilities of satisfying good ends however we can, whether or not this done by providing "necessary" conditions for such ends.

And after arguing that non-interference with the rational capacities of other agents is an absolute moral obligation because the similar non-interference from others is "required," something each agent "must have" from others, Smith (2011: 361) then denies that promotion of other agents' welfare, and even reduction of their pain, is not so strictly obligatory, because "the overlap [between welfare and exercise of rational agency] is not perfect." He forgets that the overlap is almost never perfect; I could perhaps exercise my agency adequately despite others' attempts to interfere with it, while pain or the lack of direct assistance with my ends could be more serious problems than someone directly trying to mess with my rational capacities.

10

Sometimes the necessity argument is used to privilege certain more specific ends or institutions, like capitalism or traditional sexual norms, etc. Libertarians (cf. Rand 1970: 25-26; Otteson 2006: 22-23; Narveson 2008: 97) who wish to privilege capitalism and private property often claim that these are necessary conditions for all kinds of obviously good things, and hence that any principle demoting their value in favor of others are not universalizable. To make this work, of course, they must ignore the fact that capitalism and private property are not logically necessary for life and the goods therein, just highly conducive to the same for many people in many contexts—as are, for many people in many other contexts, social and physical public goods and many kinds of welfare or alternative economic institutions.

Religiously-inspired natural law theorists like Finnis also use this argument pervasively when they consider certain acts of contraception (1991: 86) or sexual preferences (1997: 127, 132) to be essentially "acting directly against" by denying the necessary means for the "basic good" of procreation, while denying that natural family planning, sex between infertile married heterosexuals, or (presumably) priests taking vows of celibacy do the same, even though the chances of thereby procreating are just as low, barring miracles. Euthanasia and abortion are in turn typically condemned by this school of thought for denying the necessary conditions for life, while refusing an abortion to a woman whose life is in danger is often considered virtuous, on the grounds that the abortion is not logically necessary to save her life—that being only a matter of probabilities. Anscombe (1957: 17 note) made the same move to avoid the hypothetical choice between judicially executing an innocent man and allowing a nuclear war to break out by suggesting that in any actual case there might be other possible options—i.e., that the lesser bad was not truly necessary to avoid the far worse alternative, while the intentional killing of a person always necessarily denies the victim the good of his life.

Once again, all such arguments are necessarily selective in their imputation of "necessities"—for perhaps in some actual case killing an innocent would be, by all our evidence, necessary to prevent a war, or an abortion necessary to prevent further harm and death, and euthanasia to prevent more needless and excruciating suffering unto inevitable death. Of course, it is logically possible that we could be wrong about any of these things—a miracle might even occur! But appeal to the vagaries of the divinity who works in mysterious ways only exacerbates the problem, for as Jonathan Bennett (1966; 1995: 209-25) pointed out, miracles could just as easily go the other way—perhaps your attempt to kill the innocent man will only result in the appearance of his death but still save the world, perhaps euthanasia and abortion will be compatible with the greatest goods for all (including the targets of the act), and perhaps Adam and Steve's lovemaking will somehow produce a child, through the inscrutable will and providence of the deity through whom all things (*all* things, mind you!) are supposedly possible.

Note further that some of the specific examples of such conservative arguments, like that based on the necessity of some heterosexual couple to engage in vaginal intercourse in order to procreate, show that even if a certain means was truly necessary for some admittedly good end E, such necessary conditions might be cheap and easy to come by compared to the more arduous task of raising a healthy and happy child, and so we have little need to value instances of the former above the latter.

Like advertisements urging us to buy lottery tickets on the grounds that we just might win the grand prize, while ignoring the ticket's low expected value, these appeals draw furtively upon our strong valuation of the best possible outcome, and hence to just part of the probability-based framework of practical reason which we use in most of our lives. They ask us to not look behind the curtain at the real probabilities, and urge us to just hold out hopes for the best possibility,

demanding that we not reject it by denying its necessary preconditions (buy that ticket!) At least, they say, do that when I tell you to—but you must certainly avoid doing so all the time, lest you realize that almost no action ever truly forecloses the logical possibility of any result whatsoever, paralyzing your every choice.

It is worth briefly noting that a conjoined UA/probability test can easily distinguish harmless coordination principles from immoral free-riding ones, a problem which frequently stumps Kantians (Herman 1993: 139; Wood 1999: 106). Both sorts of principles can pass a UP test, for if properly qualified to only permit self-benefitting action when not all other agents are behaving likewise, universal conformity to them will not lead to the actions' universal practice and hence to the denial of necessary conditions for satisfying any ends. But any level of conformity to them—universal or partial—will change the probability of agents' ends in general being satisfied—with maxims of fruitful coordination (buying clockwork trains when others are selling, playing tennis or becoming a dentist when others are less likely to do so, etc.) increasing, and free-riding maxims decreasing this chance, so that they pass and fail respectively a UA/probability test.

These two distinctions in strategy are not the only differences between deontologists and consequentialists. But given the frequent appeals to them, it is curious that they have not been previously analyzed in general, and especially by critics of deontology. For it seems that the real differences between them is not that the latter evaluate actions via their consequences while the former do not, but that the latter appeal to actions' real consequences in actual circumstances, while the former appeal, far more dubiously, to imaginary consequences in unreal worlds.

Bibliography

Allison, Henry. (1990) *Kant's Theory of Freedom*. New York: Cambridge University Press.

Anscombe, Elizabeth. (1957) *Intention*. Oxford: Blackwell.

Bennett, Jonathan. (1966) "Whatever the Consequences." *Analysis* 23: 83-102.

Brandt, Richard. (1992) *Morality, Utilitarianism, and Rights*. New York: Cambridge University Press.

Broad, C.D. (1916) "On the Function of False Hypotheses in Ethics." *International Journal of Ethics* 26: 377-97.

Engstrom, Stephen. (2009) *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge, MA: Harvard University Press.

Finnis, John. (1980) *Natural Law and Natural Rights*. Oxford: Clarendon Press.

--. (1991) *Moral Absolutes: Tradition, Revision, and Truth*. Washington, D.C.: Catholic University of America Press.

--. (1997) "The Good of Marriage and the Morality of Sexual Relations: Some Philosophical and Historical Observations." *American Journal of Jurisprudence* 42: 97-134.

Gewirth, Alan (1978) *Reason and Morality*. Chicago: University of Chicago Press.

--. (1982) *Human Rights: Essays on Justification and Applications*. Chicago: University of Chicago Press.

--. (1998) *Self-Fulfillment*. Princeton, NJ: Princeton University Press.

Hare, R.M. (1981) *Moral Thinking: Its Levels, Method, and Point*. Oxford: Clarendon Press.

--. (1997) *Sorting Out Ethics*. Oxford: Oxford University Press.

Herman, Barbara. (1993) *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.

Kant, Immanuel. (1996) "Groundwork of the Metaphysics of Morals," In *Practical Philosophy*.

      Mary Gregor, translator. New York: Cambridge University Press. Cited as (G).

Kitcher, Patricia. (2004) "Kant's Argument for the Categorical Imperative." *Nous* 38: 555-584.

Korsgaard, Christine. (1996a) *The Sources of Normativity*. New York: Cambridge University

      Press.

--. (1996b) *Creating the Kingdom of Ends*. New York: Cambridge University Press.

Narveson, Jan. (1998) *The Libertarian Idea*. Philadelphia: Temple University Press.

Moore, G.E. (1903) *Principia Ethica*. New York: Cambridge University Press.

Otteson, James. (2006) *Actual Ethics*. New York: Cambridge University Press.

Pettit, Philip. (2000) "Non-Consequentialism and Universalizability." *The Philosophical*

      Quarterly 50: 175-190.

Reath, Andrews. (2006) *Agency and Autonomy in Kant's Moral Theory*. Oxford: Clarendon

      Press.

Ridge, Michael. (2006) "Introducing Variable-Rate Rule-Utilitarianism*." The Philosophical*

      Quarterly 56: 242-253.

Schapiro, Tamar. (2003) "Compliance, Complicity, and the Nature of Nonideal Conditions." *The*

      Journal of Philosophy 100: 329-355.

Singer, George Marcus. (1961) *Generalization in Ethics: An Essay in the Logic of Ethics, with*

      the Rudiments of a System of Moral Philosophy. New York: Alfred A. Knopf.

Singer, Peter. (1981) *The Expanding Circle: Ethics and Sociobiology*. New York: Farrar, Straus

      & Giroux.

--. (1993) *Practical Ethics*. New York: Cambridge University Press.

Smart, J.J.C. (1973) "An Outline of a System of Utilitarian Ethics." In J.J.C. Smart and Bernard
Williams. *Utilitarianism: For and Against*. New York: Cambridge University Press.

Smith, Holly. (2010) "Measuring the Consequences of Rules." *Utilitas* 22: 413-433.

Smith, Michael. (2011) "Deontological Moral Obligations and Non-Welfarist Agent-Relative
Values." *Ratio* 24: 351-363.

Timmermann, Jens. (2005) "Why Kant Could not Have Been a Utilitarian." *Utilitas* 17: 243-264.

Wood, Allen W. (1999) *Kant's Ethical Thought*. New York: Cambridge University Press.