Draft 5 Feb 99

The Cost of Non-Consequentialism

Philip Pettit

Non-consequentialism comes at a cost; in particular, at a cost that is not often recognised amongst those who embrace the doctrine. If non-consequentialists are to embrace the requirement of universalisability and still hope to maintain their non-consequentialism, then they have to strike a surprisingly relativistic stance. They have to say that each speaker predicates a distinct property in speaking of what it is right for her to do but that when the person universalises and then speaks of what it is right for another to do in those circumstances, she has to be predicating, not rightness as she conceives of it, but rightness as that other person thinks of it. Or so at any rate I argue.

Consequentialism and non-consequentialism[1]

Non-consequentialism comes in many forms. Kantians say that agents ought to act on the categorical imperative, or ought to treat other people always as ends. Virtue ethicists say that they ought to manifest certain virtues in their behaviour. Deontologists hold that they ought to discharge certain duties: they ought to tell the truth, keep their promises, be non-violent, etc.. Rights theorists maintain that they ought to respect certain rights that others have against them. Contractualists assert that they ought to conform to principles that no one could reasonably object to as the bases of social life. Theorists of special obligation say that they ought to deal in a certain way with those who are bound to them. And so on.[2]

What is the common thread in these positions? All non-consequentialists speak, at whatever level of abstraction, about what any or every agent ought to do; in that sense they are universalists. All non-consequentialists prescribe certain patterns of behaviour or psychology or relationships for agents: acting on the categorical imperative, manifesting certain virtues, respecting the rights of those who deal with them, nurturing their own friendships, and so on. And all non-consequentialists say that the right thing for an agent to do is to instantiate those patterns — so far as they are co-instantiable — in their own behaviour or psychology or relationships. In particular, they say that that is the thing to do even if instantiating a pattern in their own life means, because of the perversity of the agent's circumstances, that the pattern will be less fully realised than otherwise in the world as a whole.[3]

Consequentialism, as I understand the doctrine, makes a dual assertion. First, it asserts that there are certain neutral values by means of which possible states of affairs can be ranked, though perhaps not completely; these may be patterns in human behaviour or psychology or relationships, of the kind that the non-consequentialist prescribes, or they may be outcome-patterns that are more detached from agency: patterns such as the maximisation of human happiness or the saving of

uninhabited wildernesses.[4] And second, consequentialism maintains that the right choice for an agent to take in any decision is one of those choices, assuming there is at least one — assuming incompleteness is not a problem — that promote the overall realisation of such values or patterns: we need not rule on what it is to promote values, though the obvious gloss would equate it with maximising expected value.

Consequentialism in this sense is a very open-ended doctrine. It says nothing on what the relevant values are. It need not amount, for example, to utilitarianism; it may even espouse as values any of the patterns recommended by non-consequentialists. And consequentialism in this sense leaves open the question as to what day-to-day habits of decision-making ought to be cultivated by agents. It does not entail, for example, that the agent ought to calculate every choice by reference to the relevant values. Such calculation might do much less well for the promotion of relevant values than a policy of following certain decision-making routines: than a policy, say, of in general taking one's cue from the categorical imperative, or from what the manifestation of virtue requires, or from the apparent rights that others can assert.

But however various are the forms of non-consequentialism, and however open-ended is this characterisation of consequentialism, the difference between the two approaches is salient. Where non-consequentialism holds that certain patterns are to be instantiated in the agent's behaviour and relationships, even if this means that their overall realisation is not thereby promoted, consequentialism reverses that order. It says that certain patterns are to be promoted, even if this means that the agent fails to instantiate corresponding patterns in their own life.

The non-consequentialist pacificist says that non-violence is to be instantiated by agents, for example, even if this means that there is less non-violence overall. The consequentialist counterpart says that non-violence is to be promoted overall, even if this means instantiating violence: say, waging the war to end all wars. In practice the two sorts of pacifist will often converge in their recommendations, as non-consequentialist Quakers converged with Bertrand Russell in their opposition to World War I. But there is always a possibility of their coming apart, as of course the Quakers and Russell came apart in their attitudes to the second World War.

The universalisability challenge

Every prescription as to what an agent ought to do should be capable of being universalised, so that it applies not just to that particular agent, and not just to that particular place or time or context or whatever. So at any rate we generally assume in our moral reasoning. If we think that it is right for one agent in one circumstance to act in a certain way, but wrong for another, then we commit ourselves to there being some further descriptive difference between the two cases, in particular a difference of a non-particular or universal kind.

Thus, if we say that an agent A ought to choose option O in circumstances C — these may bear on the character of the agent, the behaviour of others, the sorts of consequences on offer, and the like — then we assume that something similar would hold for any similarly placed agent. We do not think that the particular identity of agent A is relevant to what A ought to do, any more than we think that the particular location or date is relevant to that issue. In making an assumption about what holds for any agent in C-type circumstances, of course, we may not be committing ourselves to anything of very general import. It may be, for all the universalisability constraint requires, that C-type circumstances are highly specific: so specific, indeed, that no other agent is ever likely to confront them.

There is no difficulty in seeing how the universalisability challenge is supposed to be met under consequentialist doctrine. Suppose that I accept consequentialist doctrine and believe of an agent that in their particular circumstances, C, he or she ought to choose an option O. For simplicity, suppose that I am myself that agent and that as a believer in consequentialism I think of myself that I ought to do O in C. If that option really is right by my consequentiaist lights, then that will be because of the neutral values that it promotes. But if those neutral values make O the right option for me in those circumstances, so they will make it the right option for any other agent in such circumstances. Thus I can readily square the prescription to which my belief in consequentialism leads with my belief in universalisability. I can happily universalise my self-prescription to a prescription for any arbitrary agent in similar circumstances.

In passing, a comment on the form of the prescription that the universalisability challenge will force me to endorse. I need not think that it is right that in the relevant circumstances every agent do O; that suggests a commitment to a collective pattern of behaviour. I will only be forced to think, in a person-by-person or distributive way, that for every agent it is right that in those circumstances he or she do O. Let doing O in C amount to swimming to the help of a child in trouble. Universalisability would not force me to think that everyone ought to swim to the help of a child in such a situation; undoubtedly they would frustrate one another's efforts. It only requires me to think, as we colloquially put it, that anyone ought to swim to the help of the child; it only requires a person-by-person prescription, not a collective one.[5]

So much for the straightforward way in which consequentialism can make room for universalisability. But how is the universalisability challenge supposed to be met under non-consequentialist theories? According to non-consequentialist theory, the right choice for any agent is to instantiate a certain pattern, P; this may be the pattern of conforming to the categorical imperative, manifesting virtue, respecting rights, honouring their special obligations, or whatever. Suppose that I accept such a theory and that it leads me to say of an agent — again, let us suppose, myself — that I ought to choose O in these circumstances, C, or that O is the right choice for me in

these circumstances. Can I straightforwardly say, as I could under consequentialist doctrine, that just for the reasons that O is the right choice for me — in this case, that it involves instantiating pattern, P — so it will be the right choice for any agent in C-type circumstances? I shall argue that there are difficulties in the path of such a straightforward response and that these raise a problem for non-consequentialism.

The problem

Suppose that I do say, in the straightforward way, that pattern P requires, not just that I do O in C, but also, for any agent whatsoever, that that agent should do O in C as well. Suppose I say, in effect, that it is right for me to do O in C only if it would be right for any agent X to do O in C. Whatever makes it right that I do O in C makes it right, so the response goes, that any agent do O in C.

This response is going to lead me, as a non-consequentialist thinker, into trouble. Judging that something is right gives one a normative reason to prefer it; the judgment of rightness must provide such a reason if it is to have an action-guiding role. When I think that it is right that I do O in C, therefore, then I commit myself to there being a normative reason for me to prefer that I do O. And when I assert that it is right that anyone should do O in C-type circumstances, then I commit myself — again, because of the reason-giving force of the notion of rightness — to there being a normative reason for holding a broader preference. I commit myself to there being a normative reason for me to prefer, with any agent whatsoever, that in C-type circumstances that agent do O.

The problem with these reasons and these commitments, however, is that they may come apart. For it is often going to be possible that, perversely, the best way for me to ensure or increase the chance that for any arbitrary agent, X, that agent does O in C-type circumstances, is to choose non-O myself in those circumstances.[6] The best way to satisfy the preference as to what the arbitrarily chosen agent should do may be to go against the preference as to what one should do oneself. The best way to get people to renounce violence may be to take it up oneself; the best way to get people to help their own children may be not to press for the advantage of one's own; the best way to minimise murder may to commit a murder; and so on. More generally, the best way to promote the instantiation of pattern, P, where this is the basic pattern to which one swears non-consequentialist allegiance, may be to flout that pattern oneself. The best way to increase the chance that for any arbitrary agent, X, that agent instantiates P may be not to instantiate P oneself.

How can I avoid the conclusion that in such a perverse situation I ought to promote the overall instantiation of my cherished pattern, even at the cost of not instantiating it myself? How, in other words, am I to keep faith with the non-consequentialist commitment to the rightness of

instantiating P, even where this means that the overall realisation of the pattern falls short of what it might have been?

It is hardly going to be plausible for me to say that normative reasons bearing on preferences over my own choices trump normative reasons bearing on preferences over how other people behave. Both sorts of reasons are supported in the common language of what is the right choice or of what ought to be done.[7] And it would surely run against the spirit of universalisability — the spirit in which I deny that my own particular identity is important to the prescription defended — to say that a reasoned preference as to what I do myself should not be responsive to a similarly reasoned preference as to what people in general do — what arbitrary agent, X, does — in the sorts of circumstances in question.

The upshot is that if as a non-consequentialist theorist I straightforwardly universalise the prescription that in a certain situation I should instantiate a favoured pattern, P, then the prescription to which I thereby commit myself — that in that situation any X ought to instantiate pattern, P — may force me to revise my original self-prescription. I have equal reason to prefer both that I instantiate P and that any agent instantiate P — this reason is expressed by the use of the word 'right' or 'ought' in each case — and the spirit of universalisability blocks me from treating myself as in any way special. Thus, if the preferences are inconsistent in a certain situation — if the choice is between my instantiating P alone, for example, or my acting so that many others instantiate P instead — then I will have reason not to instantiate P myself.

As a would-be non-consequentialist thinker, my initial claim must have been that the point is to instantiate P in my own life, not promote it generally. But I countenance the general claims of the P-pattern when I universalise in the straightforward way: I prescribe general conformity to that pattern, not just conformity in my own case. Thus it now seems that what I must think is that this general conformity is to be promoted, even if that means not myself instantiating the pattern in my own behaviour or psychology or relationships. It seems that what I must embrace, in effect, is a consequentialism in which conformity to pattern P is the ultimate value to be promoted.

The sort of consequentialism envisaged here is not an outlandish doctrine to will on some self-described non-consequentialists. It is a consequentialism in which the goal is maximal conformity with a reflexive pattern: with a pattern whereby, for any X, X instantiates a certain pattern in their behaviour, or psychology, or relationships. X does not treat other people only as means; X displays a certain pattern of virtue; X keeps his or her promises; X respects the rights of those with whom they come into contact; or whatever. It may well be that many self-described non-consequentialists are really reflexive consequentialists in this mould. They may be reflexive consequentialists who think that it is extremely unlikely that promoting the reflexive pattern they

celebrate will ever require flouting it and who have little incentive, therefore, to emphasise the consequentialist aspect of their commitment.

One example of a reflexive consequentialist is the person described by Robert Nozick as a utilitarian of rights: someone who thinks that an agent like the state ought to act so that there is more rather than less rights-satisfaction, even if that means flouting some rights itself (Nozick 1974, 28). And another is the just-deserts theorist who thinks that the courts ought to pass occasional exemplary sentences in order to ensure that a regime of just deserts is promoted: in order to minimise the number of cases where offenders go unpunished (Braithwaite and Pettit 1990). Such thinkers often describe themselves as non-consequentialists but clearly what they really espouse is a consequentialism in which the overall aim is to promote a certain regime of rights or deserts.[8]

The only recourse

But many non-consequentialists will shrink from embracing reflexive consequentialism. So is there any other recourse available? Is there any way in which a non-consequentialist can consistently endorse the universalisability constraint?

The reason why the straightforward response to the universalisability challenge raises a problem for non-consequentialism is this. It would force me to think that just as it is right that I instantiate P in a certain situation, so in the very same sense of 'right', it is right than anyone instantiate P in that sort of situation. It would force me to employ the same currency of evaluation — the same language of 'right' or 'ought' or whatever — in expressing the value I attach to my instantiating P and to anyone's instantiating P. The only escape from the problem, then, will be for the non-consequentialist to argue that this is mistaken and that in universalising the particular judgment as to what it is right for me to do — in moving to a judgment as to what it is right for anyone to do — I introduce an evaluative vocabulary that does not commit me in the same way.

As it happens, there is a ploy by which non-consequentialists can effect this escape. In the straightforward mode of universalisation, 'It is right for me to do O in C' goes over to 'It is right for arbitrary X to do O in C', where 'right' in each case is used to express the same sort of prescription. But non-consequentialists can escape the problem raised by arguing that 'right' — or 'ought' or whatever — does not express the same sort of prescription in the two judgments.

There are two steps whereby non-consequentialists can achieve this result. First, they can argue that when a person, A, says that something is right, or at least right for them, then 'right' is indexical and is best represented as 'right_A'; this indexicalised predicate serves to identify the property of meeting A's standards, as the corresponding predicate 'right_B' — 'right' as used in B's mouth — would serve to identify the property of meeting B's standards (Dreier 1990).[9] And then,

second, they can say that the way to universalise A's judgment 'It is $right_A$ that I, A, do O in C' is not

(i). (X)(It is $right_A$ that X do O in C); but rather

(ii). (X)(It is $right_X$ that X do O in C).

If non-consequentialists take this line, then they can escape the pressure towards reflexive consequentialism. For by taking 'right' — and 'ought' and other such words — to be indexical, they ensure that my commitment to what I do myself has a hold on me that my commitment to what any other does in similar circumstances need not. Universalising no longer has to mean countenancing a universal prescription implicit in my prescription for what I should do myself. It need only mean recognising that as I prescribe for myself in my own position, so others must be allowed to prescribe for themselves in theirs.

The idea is this. The question raised when someone asks if it is right for them to take a certain choice, like the question raised when they ask if it is prudent, varies from person to person. A asks if it is $right_A$, B asks if it is $right_B$, and so on. The non-consequentialist claim is that it is $right_A$ for A to instantiate P in A's life and similarly that it is $right_B$ for B to instantiate P in B's life. And so universalisation does not commit a non-consequentialist, A, to thinking that any general pattern is $right_A$; that commitment would raise the same problem as before. It commits A to thinking only that as A's instantiating P is $right_A$, so B's instantiating P is $right_B$, and so on.[10]

Conclusion

It does not follow from this argument that non-consequentialism reduces to absurdity. However implausible it may be to take the two steps described — to indexicalise 'right' and to hold that universalisation shifts the index of the rightness predicate — it is not an absurd move to make. When we use the word 'prudent' we do often change the indexical reference of the term — here I mean prudent for me, there I mean prudent for you — and so there is no incoherence in saying that when I speak of rightness, now in relation to my actions, and now in relation to yours, the index of of the predicate shifts. [11]

But what the argument achieves, if it is sound, is to force non-consequentialists to face squarely the sort of relativisation implicit in their position. Where consequentialists hold that the successful justification of a moral judgment must always take people back to common values — it must show that the judgment is required by those shared values — non-consequentialists lower the standards that justification is required to satisfy. On their view all that justification need do is to show each person addressed that the judgment made, and the action taken, was reasonable from the agent's point of view; in particular, it was a counterpart of the judgment and the action that they might themselves have endorsed in their own case.

On this approach, speaking of what is right in one's own case never amounts to speaking about a common property. It amounts only to employing a common schema whereby each participant can refer to the property that matters to them: the property of being right, as that is indexed to their position. My trying to argue that something I did was right, therefore, is not an exercise pursued with reference to any common lights. It is an attempt to show that the lights whereby the action looked acceptable to me, while they are not necessarily shared with any others, are at least lights that others should understand. It is an attempt to display, not an identity in evaluative commitments, but rather an isomorphism of evaluative standpoints.

This upshot will not drive non-consequentialism out of the domain of possible theories. But it will come as a surprise, I think, and it does count as a cost. It may give some non-consequentialists reason to rethink their point of view.[12]

References

Baron, Marcia, Philip Pettit, and Michael Slote. 1997. *Three Methods of Ethics: A Debate.* Oxford: Blackwell.

Braithwaite, John, and Philip Pettit. 1990. *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford: Oxford University Press.

Cummiskey, David. 1996. *Kantian Consequentialism.* Oxford: Oxford University press

Dreier, James. 1990. Internalism and Speaker relativism. *Ethics* 101:6-26.

McNaughton, David, and Piers Rawling. 1992. Honoring and Promoting Values. *Ethics* 102:835-43.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.

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

Humberstone, I.L. 1991. 'Two Kinds of Agent-relativity'. *Philosophical Quarterly* 41:144-66.

---

[1] The account given here of consequentialism and non-consequentialism — strictly, consequentialism and non-consequentialism about acts — draws on earlier work, most recently on my contribution to Baron, Pettit, and Slote (1997). The problem I raise for non-consequentialism is also raised in that contribution but in a somewhat different form.

[2] I focus here only on the prescriptions supported by non-consequentialists. I say nothing on the fact that because they do not cover all aspects of action and life, the prescriptions supported often entail corresponding permissions.

[3] Where I speak here of instantiating a pattern in one's behaviour or relationships, I have spoken in earlier work of honouring the pattern. I consider only non-consequentialism that requires an agent-over-time to instantiate certain patterns. A stricter version would require an agent-at-a-time always to instantiate those patterns, even if that meant that the pattern was less well realised over the agent's life as a whole.

[4] Where I assume that all patterns of the kind that non-consequentialists think should be instantiated in an agent's life can be taken as neutral values to be promoted, I am not now inclined to think that all patterns of the kind that consequentialists think should be promoted can be taken as patterns to be instantiated in an agent's life. See McNaughton and Rawling (1992).

[5] The difference is one of scope. I need not say for every X, that X ought to do O in C-type circumstances. Rather, for every X, I say that X ought to do O in C-type circumstances. This helps explains why Harean universalisation is different from Kantian generalisation. See Hare (1981).

[6] Why do I say that his holds often, not always? Because some options may be characterised in a way that does not leave open the perverse possibility. I am thinking of an option such as that of swimming to save a drowning child when no one else is doing so.

[7] A non-consequentialist might think of arguing at this point that the notion of rightness has reason-giving force for someone who makes a judgment of rightness only where the judgment bears on their own action. But it is hard to make good sense of this move, short of the indexicalisation introduced below. If self-prescription gives me reason to prefer that which I prescribe for myself, so other-prescription would seem to give me reason to prefer that which I prescibe for others. Lloyd Humberstone has urged on me that this is the move made by R.M.Hare. See Hare (1981).

[8] Another example of a reflexive consequentialism would be the consequentialism that David Cummiskey (1996) derives from Kantian principles: a consequentialism under which the maximand is the treatment of rational beings as ends-in-themselves and not as means only.

[9] On the recourse to indexicalisation, and other reasons for invoking it, see Humberstone (1991). See in particular pp.148-49, where he envisages a formulation of ethical egoism that parallels the formulation that I envisage here for non-consequentialism.

[10] Such a non-consequentialist may hold that for any X and Y, if it is right_X that X do something, then it is right_Y that Y allow X to do it; the important point is that they do not hold that in such a case it has to be right_Y that X do it.

[11] But notice that the sort of move envisaged is not consistent with an expressivist, as distinct from a cognitivist, viewpoint. For according to expressivism 'right' in the speaker's mouth will always express the speaker's attitudes, never the attitudes of an arbitrary agent, X; it will always serve in

parallel to the way ‘$right_A$’ serves, on the story told here, never in parallel to the way ‘$right_X$’ serves. The upshot is that expressivism is inconsistent with the combination of non-consequentialism and universalisability.

[12] My warm thanks to Don Regan who challenged me at every step in the argument. My thanks too to the audiences at a number of seminars where I presented the paper: at the Colloque de Cerisy; at La Trobe University, Melbourne; at the University of Canterbury, Christchurch; and at L’Universite de Montreal. I was particularly helped by comments from Diane Proudfoot and Jack Copeland and by discussion with Lloyd Humberstone.