Is act-consequentialism self-effacing?

Nikhil Venkatesh

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

1. Introduction

Act-consequentialism is a theory that can be expressed thus:

C Agents ought to do whatever produces the best outcome.

A theory T is self-effacing for an agent A iff it implies that

A ought not accept T.[1] C is self-effacing for A iff A's not accepting C would produce the best outcome. Some hold that if C is self-effacing that would be a mark against it: it would 'usher itself from the scene' (Williams 1973: 134), failing to be action-guiding. The claim that C is self-effacing is also used to defend C. Many objections to C stem from problems associated with agents accepting it. Williams claims that accepting consequentialism is incompatible with engaging in the projects that make life meaningful (Williams 1973). It is also said that someone who accepts C could not be a good friend (Kapur 1991) or uphold social practices such as promising (Hodgson 1967). Call these the Problems of Acceptance. If C implies that individuals ought not accept C, the Problems of Acceptance are not counterexamples to it. Parfit (1984: Ch. 1) and Railton (1984) make such arguments.

If C is self-effacing, then, it can escape the Problems of Acceptance. However, it would face the objection that it ushers itself from the scene. If C is not self-effacing, it escapes this objection, but not the Problems of Acceptance.

In this paper I will show that one argument, drawn from Parfit and Railton, fails to establish that C is self-effacing for individuals. However, a slightly different argument could succeed in showing that C is self-effacing for groups. This raises the possibility that it might be right for an individual, but not for a group of which they are a member, to accept a moral theory. This possibility, odd though it seems, might help consequentialists to mitigate the Problems of Acceptance whilst affirming C as a guide to action.

Two preliminary notes: first, all plausible moral theories are self-effacing for some agents in some circumstances. Parfit demonstrates this with examples involving Satan (Parfit 1984: 43–45 – see also de Lazari-Radek and Singer 2010 and Eggleston 2013). For any theory, we can imagine that Satan credibly threatens to unleash great amounts of suffering on the world if some agents accept some theory. If that theory is plausible, it will tell those agents to do whatever it takes to avoid this outcome (it is not only consequentialists who care about avoiding disastrous outcomes). Therefore, it will be self-effacing for

[1] To accept a theory, broadly, is to be disposed to use it to inform one's practical reasoning, and to take certain practical attitudes towards it (Van Fraassen 1980, Harman 1986, Ross 2006).

Analysis Vol. 00 | Number 0 | 2021 | pp. 1–9 doi: 10.1093/analys/anab042

© The Author(s) 2021. Published by Oxford University Press on behalf of *The Analysis Trust*.This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

them. Secondly, the claim that C is self-effacing is not equivalent to the claim that C recommends that agents, in some circumstances, act on motives other than C. This has been noted by consequentialists going back to Mill (2008: 149–53) and Sidgwick (1962: 413), and again, similar claims are likely true of any plausible moral theory.

2. *The argument from collective self-defeat*

The Problems of Acceptance purport to show that one's accepting C precludes one from enjoying certain valuable things (friendships, projects, trust). This claim is shared ground between those pushing the Problems of Acceptance as objections to consequentialism and those arguing for consequentialism's self-effacingness as a response. It implies that if everyone accepted C, these valuable things would completely disappear. I will assume that on any plausible version of C the complete disappearance of these things would be suboptimal: it would make things worse than they would otherwise be. Hence, C is (given these problems) 'collectively self-defeating' (Parfit 1984: ch. 1) – if every individual accepted C, the aims of C would be much less well achieved.

Parfit and Railton seem to suggest that it follows from the claim that C is collectively self-defeating that it is self-effacing for individual agents. Parfit writes:

> If we were all pure do-gooders, the outcome would be worse than it would be if we had other sets of motives. If we know this, C tells us that it would be wrong to cause ourselves to be, or to remain, pure do-gooders. (Parfit 1984: 28)

He makes a similar argument in later work, writing that, if it were true that if 'it would be best if everyone accepted some improved version of Common Sense Morality' (rather than C), then

> Act Consequentialism would imply that . . . everyone ought to accept, not [C], but this version of Common Sense Morality. (Parfit 2017: 415)

Meanwhile, Railton imagines someone justifying their non-consequentialist commitments – which are supposedly necessary for a loving relationship – on consequentialist grounds by saying,

> Look, it's a better world when people can have a relationship like ours and nobody could if everyone were always asking themselves who's got the most need. You'd make things worse in a hurry if you broke up those close relationships for the sake of some higher goal. (Railton 1984: 150)

Parfit and Railton seem to infer from the fact that C is collectively self-defeating that it is self-effacing for individual agents – that is, from the fact that outcomes would be suboptimal if everyone accepted C to the claim that, according to C, individuals ought not to accept C. However, it is not in general

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

true that if the outcome of our all doing something would be suboptimal, consequentialism implies that each of us ought not do it. Consider this case from Feldman:

> Suppose a group of adults has taken a group of children out to do some ice skating. . . . [W]hile they are out skating, it just so happens that all the adults are spread out around the edge of the pond. A lone child is skating in the middle, equidistant from the adults. Suddenly, the ice breaks, and the child falls through. There is no time for consultation or deliberation. Someone must quickly save the child. However, since the ice is very thin, it would be disastrous for more than one of the adults to venture near the place where the child broke through. For if two or more were to go out, they would all fall in and all would be in profound trouble. (Feldman 1980: 171)

Now, imagine you are one of the adults and you know that no other adult will go to save the child. C directs you to save the child: this would result in the best available outcome. This is true even though if every adult did the same, the outcome would be worse. C determines what an agent ought to do according to the expected effects of their actions, taking facts such as what others will do as given.

Accepting C could be like going to save the child. The Problems of Acceptance make it plausible that although everyone accepting C would be suboptimal, the expected effects of *my* accepting C, given the likely behaviour of others, would be optimal. If so, C would not be self-effacing for me, despite being collectively self-defeating.

First, consider social institutions such as promising. Perhaps trust would be undermined if everyone accepted C: people would not accept promises that they knew would be disregarded whenever the greater good required it. But if everyone else carries on as they are, *my* acceptance of C would not undermine trust very much. Most people who rely on promises never rely on *my* promises, and don't know what I think about morality. Of course, some people do, and their trust in me (and perhaps subsequently, their trust in others) might diminish. But this would be a small negative effect on trust, which – unlike the complete disappearance of it – could very plausibly be outweighed by the good my accepting C might do. (Similarly, in Feldman's case my going to save the child has a small effect on the safety of the ice, but not one consequential enough to threaten the rescue attempt.) For instance, widespread poverty makes outcomes significantly worse, and the most effective way to relieve such poverty might be through radical changes in public policy. I might stand a better chance of making such changes if I accept C, and am therefore always ready to betray promises in order to advance my political career (which I will then use to implement value-maximizing policies). In this case my accepting C would produce better outcomes than my not doing so, even though (and partly because) it would preclude me from making sincere promises.

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

It is a similar story with commitments to projects and intimate relationships. If it is impossible for people who accept C to have these things, it follows that the universal acceptance of C would lead to a world absent of them, which would be, according to any plausible version of C, suboptimal. But if I accepted C and everyone else carried on as they are, although I would be unable to pursue such projects and relationships, others would continue to do so. Moreover, if I accepted C, I might help them to pursue their projects and relationships more successfully (for example, by expending my energies on policies that save their lives or increase their capacities) than I would if I simply pursued my own projects and relationships. Given that these things are good, C holds that the more of them exist, the better. C would then tell me to accept C; it would not be self-effacing.

These arguments depend on the claim that most others will not accept C, whatever we ourselves do. The Problems of Acceptance give us good evidence for this claim. They show that accepting C precludes certain things that make one's life go better, such as projects and intimate relationships. This may be offset, morally speaking, by the good things one can bring into the lives of others. But individuals seldom sacrifice their projects and intimate relationships for the good of others. Therefore, if accepting C amounts to making such a sacrifice, there is little risk of everyone accepting it, and each of us has the opportunity to increase value by sacrificing ourselves to help those who do not accept C escape poverty, pursue their projects, enjoy their relationships and so on.

Interestingly, Parfit and Railton seem aware of such considerations, although they undermine their inferences from collective self-defeat to self-effacingness. Parfit writes:

> I know that most of us will continue to have the motives much like we have now. Most of us will love certain other people, and will have other strong desires on which most happiness depends. Since I know this, C may tell *me* to try to be a pure do-gooder. This may make the outcome better even though, if we were *all* pure do-gooders, this would make the outcome worse. If most people are *not* pure do-gooders, it may make the outcome better if a few people are. (Parfit 1984: 30)

Railton writes:

> [J]ust how demanding or disruptive [complying with C] would be for an individual is a function – as it arguably should be – of how bad the state of the world is, how others typically act, what institutions exist, and how much that individual is capable of doing. (Railton 1984: 161)

Parfit and Railton seem to recognize that which theory (or theories) C requires one to accept depends on what others accept. This being the case, it is plausible that if others don't accept C, I should, even if it would 'make things worse in a hurry' for everyone to do so. The fact that Parfit and Railton seem to suggest an argument that is not only invalid, but invalid for reasons they themselves

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

note, is puzzling. Next, I will suggest a different interpretation of their arguments.

3. Why C might be self-effacing for groups

Consider the following case, adapted from Postow (1977). Fred and Mary have a garden that needs attention. Each can weed, or water, or do both, or neither. It would be best if one waters and the other weeds simultaneously. It would be worse if neither task is done. Both of them watering would be disastrous, as it would waterlog their plants. However, if one of them does neither, it would be best if the other waters without weeding, since there is no time to do both in sequence, and watering is the more urgent task. In this case, it is true of each of Fred and Mary that, if the other does neither task, they ought to water without weeding. Both Fred and Mary are tired from the working week and reluctant to do either task, and each knows this is true of the other.

In the previous section I argued that although it might be worse if we all accepted C, it might be best for each individual to accept it, given what others will do. In this respect, accepting C is like watering without weeding. Whether you are Fred or Mary, given that your partner is unlikely to do either task, you should water without weeding. This is true even though it would be disastrous if you both watered, and best if both watering and weeding were done. But there is also, intuitively, a sense in which 'Fred and Mary ought not water without weeding' is true. This suggests that there might be a sense in which 'C implies that people ought not to accept C' is true – that is, that C is self-effacing – even if individuals ought to accept C.

According to Postow, statements such as 'Fred and Mary ought not water without weeding' are true when we conceive of Fred and Mary as a group (see also Jackson 1987 (pp. 91–110); although compare Smith 2009). Likewise, 'C implies that people ought not to accept C' could have a true sense if according to C, those people form a group that ought not accept C. I will argue that it is plausible that some such claims are true – and this (unlike the claim that C is self-effacing for individuals) is supported by C's collective self-defeatingness.

This offers an alternative reading of Parfit's and Railton's inferences. When Parfit writes that '[i]f we know [that C is collectively self-defeating], C tells us that it would be wrong to cause ourselves to be, or to remain, pure do-gooders' he might mean, not that for each of us, but rather for us as a group, it would be wrong to cause ourselves to accept C. Similarly, when Railton justifies non-consequentialist attitudes in intimate relationships by saying: 'You'd make things worse in a hurry if you broke up those close relationships for the sake of some higher goal', the 'you' could refer to the group, not any individual. This wouldn't justify any individual's non-consequentialist attitudes, but it could justify their acceptance by the group.

On this reading, Parfit and Railton are not claiming that C is self-effacing for individuals, but for groups. So, is C self-effacing for groups – that is, does C

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

hold that groups ought not to accept C? If C is collectively self-defeating then the answer is probably 'yes', for those groups that are coextensive with the collective within which C is collectively self-defeating.

I will assume that C makes claims about what groups ought to do. Many of the actions that have the greatest impact on outcomes are performed by group agents and cases like Postow's support the view that without giving recommendations to groups, distinct from those it gives to individuals, C risks being silent or incorrect about such actions (Dietz 2016). I will also assume that accepting moral theories is something that groups can do. We might think of historians and anthropologists describing which moral theories some community accepts – see talk of 'Victorian morality' or 'Indian morality'. Perhaps no community is homogeneous enough in this respect for it to be straightforwardly true to say that it accepts some theory, but we can make sense of the enterprise.

I will not take a stand on what it is for a group to accept a moral theory. Uncontroversially, a group's accepting C either implies that all or most of its members accept C, or it does not. Call these 'summative' and 'non-summative' accounts of group acceptance.[2] If C is collectively self-defeating within the group, then it would be suboptimal if every member accepted it, and if it would be suboptimal if every member of the group accepted it, then it is likely that it would also be suboptimal if most members of the group accepted it. Therefore, according to summative accounts of group acceptance, if C is collectively self-defeating, the group's accepting consequentialism would likely be suboptimal, and consequentialism would likely be self-effacing for the group.

What, if not the fact that a majority of members accept some view, makes it the case that a group accepts it? Non-summative accounts tend to emphasize widespread, conventional and norm-governed behaviour within the group. For instance, the willingness of group members to let that view stand as the view of the group, to publicly defend it, to rebuke group members who dissent from it or be quiet about their own dissent from it, and to allow group actions to proceed from it (Gilbert 1989, 1994, Tuomela 1992, Schmitt 1994). Thus, if a community accepted C, C would be proclaimed, enforced, taught to children and so on. Failing to do what C required, or openly dissenting from C, would make one fit for rebuke or punishment by the community. Whether individuals in this group accept consequentialism or not, such a regime incentivizes behaviour that is similar to the behaviour of individuals who do. If consequentialism is collectively self-defeating within the group and so every member's accepting it would be suboptimal, then it is likely that much of the membership acting as if they accept it would be suboptimal too. Furthermore, such a system is likely to cause more individuals to personally accept C. It would be the moral theory to which they are most exposed, which best

[2] It is highly plausible, of course, that summative accounts are apt for some groups, and non-summative ones for others.

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

explains the actions of their community, and with which it is least costly to comply. Thus, the group's accepting C, on non-summative accounts of acceptance, could cause (though not entail) most of its members to accept C. As argued above, if C is collectively self-defeating within some group, it would likely be suboptimal if most of its members accepted C. If the group's accepting C would cause the latter, C will likely say that the group ought not accept C.

So, if it is true that C is collectively self-defeating, then it is likely self-effacing for groups, although not necessarily for individuals. A variant of the argument from collective self-defeat therefore succeeds.

4. *Conclusion*

What are we to make of the possibility that C might be self-effacing for groups, although not for individuals?

The possibility might soften the Problems of Acceptance. If C is self-effacing for groups, it does not ask for society to press people to abandon their relationships and projects for the greater good, even if it asks individuals to make such sacrifices. It need not ask for people to be rebuked or punished for spending money on their hobbies when there are mosquito nets to be bought, even whilst it asks individuals themselves to buy mosquito nets. Moreover, if C is self-effacing for groups, it is consistent with groups performing the irreducibly group action of establishing institutions such as promising, and of individuals rebuking those who break promises (and other conventions of such institutions) – even for the greater good. All this could be true whilst individuals ought to accept C – so C cannot be accused of failing to guide action.

However, such a result seems odd. It suggests that (if C is correct) groups ought to discourage their members from accepting a morality that is not only true, but also morally right for each member to believe.[3]

Funding

This work is part of a thesis project supported by a studentship from the London Arts and Humanities Partnership and a grant facilitated by Longview Philanthropy.

University College London, UK
nikhil.venkatesh.16@ucl.ac.uk

[3] Thanks to: Véronique Munoz-Dardé, Mark Kalderon, Ulrike Heuer and Joe Horton for helpful discussions of this topic; an audience at the London Graduate Philosophy Conference 2020 for their helpful feedback; Alexander Dietz, Christopher Woodard, two anonymous reviewers and the editor at *Analysis* for comments on drafts of this paper.

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

References

Dietz, A. 2016. What we together ought to do. *Ethics* 126: 955–82.

Eggleston, B. 2013. Rejecting the publicity condition: the inevitability of esoteric morality. *Philosophical Quarterly* 63: 29–57.

Feldman, F. 1980. The principle of moral harmony. *Journal of Philosophy* 77: 166–79.

Gilbert, M. 1989. *On Social Facts*. London: Routledge.

Gilbert, M. 1994. Remarks on collective belief. In *Socializing Epistemology: The Social Dimensions of Knowledge*, ed. Frederick F. Schmitt, 235–56. Lanham, MD: Rowman & Littlefield.

Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.

Hodgson, D.H. 1967. *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory*. Oxford: Clarendon Press.

Jackson, F. 1987. *'Group Morality'. In Metaphysics and Morality: Essays in Honour of J.J.C.*

Kapur, N.B. 1991. Why it is wrong to be always guided by the best: consequentialism and friendship. *Ethics* 101: 483–504.

de Lazari-Radek, K. and P. Singer. 2010. Secrecy in consequentialism: a defence of esoteric morality. *Ratio* 23: 34–58.

Mill, J.S. 2008. Utilitarianism. In *On Liberty and Other Essays*, ed. J. Gray. Oxford: Oxford University Press.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. 2017. *On What Matters*, Vol. 3, 1st edn. Oxford: Oxford University Press.

Postow, B.C. 1977. Generalized act utilitarianism. *Analysis* 37: 49–52.

Railton, P. 1984. Alienation, consequentialism, and the demands of morality. *Philosophy & Public Affairs* 13: 134–71.

Ross, J. 2006. *Acceptance and Practical Reason*. PhD thesis, Rutgers University.

Schmitt, F.F. 1994. The justification of group beliefs. In *Socializing Epistemology: The Social Dimensions of Knowledge*, ed. Frederick F. Schmitt, 257–88. Lanham, MD: Rowman & Littlefield.

Smith, T. H. 2009. Non-Distributive Blameworthiness. *Proceedings of the Aristotelian Society* 109: 31–60. https://doi.org/10.1111/j.1467-9264.2009.00257.x.

Sidgwick, H. 1962. *Methods of Ethics*. London: Macmillan.

Tuomela, R. 1992. Group beliefs. *Synthese* 91: 285–318.

Van Fraassen, B.C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Williams, B. 1973. A critique of utilitarianism. In *Utilitarianism: For and Against*, eds. J.J.C. Smart and B. Williams, 77–150. Cambridge: Cambridge University Press.

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021

Abstract

Act-consequentialism (C) is self-effacing for an agent iff that agent's not accepting C would produce the best outcome. The question of whether C is self-effacing is important for evaluating C. Some hold that if C is self-effacing that would be a mark against it (Williams 1973: 134); however, the claim that C is self-effacing is also used to defend C against certain objections (Parfit 1984: Ch. 1, Railton 1984).

Keywords: consequentialism, moral philosophy, self-effacingness

Downloaded from https://academic.oup.com/analysis/advance-article/doi/10.1093/analys/anab042/6377185 by UCL, London user on 28 September 2021