Insane Consequentialism: A Pragmatic Objection to Direct Consequentialism

NICK ZANGWILL
University of Hull

I argue that direct consequentialism is not rationally believable. I focus on duties of love. Those feelings are so fundamental to us that believing consequentialism creates insanity. For it entails negative judgements not just about our loyal acts, but also about our deepest feelings. If direct consequentialism is true we should be able to believe it and stay sane. But we cannot, so it is not true.

If consequentialism is true, it must be possible to believe it. This we cannot do – or so I argue. This is not because a statement of consequentialism is contradictory. Rather it cannot be believed in what we might call a 'pragmatic' sense of 'can' – the sense in which we cannot say or think '*p* but I don't believe *p*' or 'I do not exist'. In this pragmatic sense, I argue that it is not possible to believe consequentialism. If not, consequentialism is flawed and should be rejected.

I. CONSEQUENTIALISM, COMMON-SENSE MORALITY AND CONFLICT

The argument will apply only to what is called 'direct consequentialism', which we may say is the thesis that the rightness or obligatoriness of an action is determined by the impartial value of the total consequences of that action (sometimes known as 'utility'). So, of two actions, such as doing something and refraining from doing it, we ought to do the one with better total impartial consequences.

Three comments on this direct consequentialist thesis:

(A) The requirement that it is impartial consequences that determine rightness rules out partial versions of consequentialism, such as ethical egoism, or any view that makes egocentric relations part of what determines rightness. So the fact that it is the agent's brother who is benefited by a person's action, rather than someone else's brother, does not make the total outcome better.

(B) I have cast the view in terms of the rightness or obligatoriness of *actions*, as is most familiar, although the view can in principle be extended to emotions, motives, thoughts, acts of reasoning, institutions, and much more. Despite their differences, all these must be particulars, not kinds, if the view is to count as 'direct' consequentialism.

© Cambridge University Press 2018
Utilitas Vol. 30, No. 3, September 2018
doi:10.1017/S0953820817000267

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

(C) The thesis is that the rightness of an action is determined by the value of that action's outcome, or by what is the best outcome out of a range of alternatives, or on the relative value of alternative outcomes; it is not determined either by what we *think* about this outcome or by what would be *reasonable* for us to think about this outcome. That would be a different theory. Such views are sometimes classified as 'consequentialist', but that classification is odd since the principle that rightness is determined by real consequences is abandoned. On such a theory, rightness would be contingent on the idiosyncrasies of people's causal beliefs. Someone might believe that starting a nuclear war is optimific on the basis of reading tea-leaves. That would not mean that consequentialists should say that starting a nuclear war is the right action for that person. It is the consequences and their value that matter for consequentialism, not believed consequences or believed value.

Let us now consider: how does direct consequentialism fit with ordinary moral thought? By 'Common Sense Morality' (henceforth 'CSM') I mean the informal collection of moral judgements that most people accept, or would accept on reflection. CSM includes the requirement to keep promises, to respect justice, to tell the truth, and to be loyal to loved ones. Moral outlooks, of course, vary quite a lot. But CSM isolates a shared core among otherwise very disparate moral systems.

The question is: to what extent does direct consequentialism depart from CSM in these respects? Famously, quite a long way. There seems to be a clear contradiction between direct consequentialism and many of the demands of CSM. For example, CSM says that in many cases we should be just even when more impartial utility would be generated by injustice; but direct consequentialism denies this and says that, in those cases, we should do what maximizes impartial utility even though it conflicts with what CSM takes to be the demands of justice. So according to direct consequentialism, CSM is wrong in this respect. The same is true of loyalties born of love for friends and family. CSM says that, where acting on bonds of love does not maximize the impartial good, we should nevertheless act in accordance with the demands of love, and not the impartial good, whereas direct consequentialism says the reverse. In this article, I focus on the issue of loyalty to friends and family.

Why, exactly, is there a conflict between CSM and direct consequentialism on this matter? The reason is that feelings for family and friends can hardly exist without generating tendencies to act on them, at least in those who are instrumentally rational. If one loves someone, then one is likely to care about that person, and one will tend to foster that person's welfare. This is not universally true – one may

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

love the dead, for example. But at least on many occasions, one will produce partial benefit, not impartial benefit. Indeed, it is plausible that it is rational (other things being equal) to care about someone whom one loves; and if one cares, it is rational (other things being equal) to act on that care (other things being equal). So, in many cases one will promote the welfare of one's beloved. Thus, feelings of love divert one from the path of impartial goodness and it is instrumentally rational to act partially not impartially. CSM says that our energies, on many occasions, should be diverted to the partial good, away from the impartial good. From the direct consequentialist point of view, therefore, love causes one to deviate from what one should do. Not only does a direct consequentialist think that we should not *act* out of love, they are also committed to thinking that it is bad that one has such *feelings* since they tend to cause us to act such as to produce partial good, not impartial good.

Of course, actions are complex events. Actions essentially have certain motives. Or they are constituted in part by certain motives. So, the consequences of an action are consequences of the whole complex event that constitutes the action. Even so, the consequentially optimific action will be one that puts the person's loves and loyalties to one side.

II. ATTEMPTS TO AVOID CONFLICT

Consequentialists sometimes try to avoid a clash with CSM. Some direct consequentialists write as if one could happily have such feelings while completely ignoring them in one's practical reasoning.[1] Perhaps perfectly practically rational beings (in the sense of beings who have emotions that dovetail with their moral views), who also happen to be consequentialists, could do so. But people with more usual moral views are bound to be influenced by these feelings, and that is also instrumentally rational. They will sometimes cause people to sway away from promoting the impartial good. Therefore, the consequentialist must say that the feelings as well as the actions are bad.

Perhaps it would not maximize the good for us all simultaneously to get rid of our partial feelings. But what is in question are the feelings of a particular person who has to act. How is it with others is not relevant, at least for a direct consequentialist theory.

A common consequentialist move is to say that, due to limited knowledge, we should let people run their own lives. It is usually better to reason non-consequentially and not to reason consequentially. Maybe so. But that makes no difference to what people should *do*. One can assess modes of reasoning in consequentialist terms and it could

[1] For example, Shelley Kagan, *The Limits of Morality* (Oxford 1989), pp. 367–9.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

be that worshipping local tree gods and divining their wishes is the right mode of deliberation. That does not change the consequentialist assessment of action.

The point can be obscured if morality and rationality are confused. Frank Jackson characterizes consequentialism in terms what 'ought to be done',[2] which is too open about the type of obligation in question. He gives examples of cases where people lack knowledge of what the outcomes of possible actions will be, and he then asks the question: what should they do? The intuition he wants to elicit is that what they should do depends on their state of information. But any theory that makes what people ought to do turn on their state of information is simply not consequentialism, as usually defined, but a different theory. (Jackson calls it 'decision theoretic consequentialism'.) Consequentialism is the view that the moral value of an act depends on its actual consequences, or the comparative value of two acts depends on the comparative values of their actual consequences, and this is independent of the person's knowledge of consequences, subjective probabilities, and so on. The issue of what would it be *rational* to do given a person's beliefs and aims is something else entirely. What morality prescribes and what rationality prescribes are different. Talk of 'what the agents ought to do at a time'[3] obscures this difference. Perhaps there are some agents who have the consequentialist goal of maximizing the good. Even so, what they morally ought to do, for a consequentialist, does not depend either on that goal or on their information concerning that goal.

Some direct consequentialists would have us believe that there is somehow no conflict between loyalties and maximizing the good, because, even though a particular loyal action may not maximize the good, loyal actions spring from loyal motives and character, and having those *does* maximize the good. The idea is to backtrack from the action to the motives and character that produced the action, and then consider the consequences of the whole collection of which the action is a part.[4] However, even if the complex produces more good than the absence of the complex, we can still consider the consequences of the action independently of the complex. Many alternative actions promote the good more than loyal actions; and so the consequentialist should say that we should do those other actions rather than the loyal ones. It is

[2] Frank Jackson, 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection', reprinted in *Mind, Method, and Conditionals: Selected Essays* (London, 1998), pp. 220–46, at 221.

[3] Jackson, 'Decision-theoretic Consequentialism', p. 222.

[4] See, for example, Peter Railton, 'Alienation, Consequentialism and the Demands of Morality', *Philosophy and Public Affairs* 13 (1984), pp. 134–71.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

not at all plausible that the loyal motives and character of a particular individual do produce more impartial good than other motives that that individual might have.

It is true that loyal friends or brothers do tend to produce good for their friends or brothers in an ongoing and systematic way. But they produce less overall impartial good than those people who are less concerned with partial good and more concerned with impartial good. These people will, for example, give vastly more of their resources to global charities than to their friends or brothers.

So CSM and direct consequentialism clash. CSM endorses love and acting out of love on occasions where it does not promote the impartial good. And direct consequentialism has a problem with love and acting out of love in those situations.

Of course, one could cite the quite general benefits of the existence of people with loyal motives and character, but that is not relevant to a direct consequentialist view. Direct consequentialism is often contrasted with several positions called 'indirect consequentialism'. It is characteristic of these theories that particulars are evaluated with respect to some kind under which they fall, and the kind is assessed with respect to the impartial consequences of a general conformity to the kind. One brand of indirect consequentialism says that the rightness of an act depends on the impartial value of the consequences of the adoption of rules under which an act falls.[5] Another brand says that the rightness of an act depends on the impartial consequences of the prevalence of the type of motives that led to the act, where the motive and the intention out of which the act proceeds share contents.[6] The proponents of these theories hope that their indirect consequentialist views will turn out to be relatively conservative doctrines. If so, they can accept CSM. The pragmatic argument I pursue here does not apply to such views. Elsewhere I pursue a different argument against indirect consequentialism.[7] The goal of this article is to argue only that our commitment to duties of love makes direct consequentialism unbelievable. In what follows I shall mean direct consequentialism by 'consequentialism', unless I specify otherwise. Furthermore, we should be vigilant and not allow direct consequentialists to cheat and illegitimately help themselves to the benefits of indirect consequentialism. If there are some benefits from the general existence of love and loyalty for the world, direct consequentialism can

[5] Brad Hooker, *Ideal Code, Real World* (Oxford, 2000).

[6] Robert Adams, 'Motive Consequentialism', *Journal of Philosophy* 73 (1976), pp. 467–81.

[7] See Nick Zangwill, 'Cordelia's Bond and Indirect Consequentialism', *Oxford Studies in Normative Ethics* 1 (2011), pp. 143–65.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

take no advantage from that. For in many if not most particular cases there is an irreconcilable clash between direct consequentialism and the requirements of CSM that spring from love and loyalties.

III. LOYALTIES AND SELF-CRITIQUE

The pragmatic argument takes off from the claim that consequentialists (direct consequentialists) must disown many of their own deep feelings. Consequentialists must think of themselves that their feelings are defective. They have them but must condemn them. Consequentialists say not only that we should act contrary to the duties of love, but also that the feelings of love that generate tendencies to act out of love are not admirable. That is highly revisionary of CSM, *but this is not my complaint*. My argument is that if this revisionary consequentialism is true, it must be possible not just to believe the theory in the abstract, and to think that it binds *other* people, but also to believe it *of oneself*. One must be able to think of oneself that one has or lacks the duties that the theory says one has or lacks. So, it must be possible for consequentialists to think of themselves that they are not bound by ties of loyalty. But is this possible? Is it psychologically possible?

Consequentialists, we may presume, are people with normal feelings, rather than strange and unusual people who only care only for the impartial good. Like everyone else, they have partial feelings for friends and family, which are strong and entrenched. They care deeply for their mothers and fathers, for example. Or if not for their mothers and fathers, then for their children. Or if not for their children, then for their friends or partners. Let us assume that consequentialists love their parents. Since they are consequentialists, they must not just think that they should not act on this partial feeling but also that this partial feeling is problematic and flawed. I shall argue that this combination is crazy, or at least destabilizing. Such a person could not function normally. Believing of oneself that one's personal loyalties are defective and should be ignored in practical reasoning is insane, and believing that would make one insane!

Why so? Well, what is insanity? In a way, this is a crazy question! For insanity is many things. There are many ways to be insane. And some notions of insanity can be called into question on normative grounds. Nevertheless, *one* thing which is in question, in the case that interests us, is a major lack of fit between people's emotions and their judgements about their emotions. In non-insane human beings, for the most part, feelings and endorsement of those feelings go together; they have feelings and they endorse them. These reflective endorsements are what Bernard Williams famously calls 'one

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

thought too many'.[8] However, contrary to Williams, plainly we *should* sometimes have the reflective endorsing thoughts that he says we should not have. We should not have them all the time, but, occasionally, we should have them – as Thomas Nagel rightly observed in a book review of Williams.[9] Having them is a natural and reasonable part of life.

Of course, feeling and reflective endorsement can come apart. We are unlikely to approve of *all* our feelings. We may think that some things we feel are petty. We may even think that some of them are sick or depraved. We can be down on ourselves. We can and should be critical of ourselves. But what is important in the present context is that we must endorse *some* of our feelings if we are to be sane enough to critique others. The appropriate picture is the Neurathean one of a raft that we must rebuild at sea. We can condemn some feelings but only if we approve others. The feelings we approve are the firm basis on which we stand when we critique other feelings. What we cannot do is to condemn them all.

This thought needs refining. What is in question is not the sheer *number* of feelings, which cannot be condemned; the claim is not simply that we cannot condemn *most* of our feelings. For one thing, it may make no sense to count their number. If there is no determinate number of feelings that we have, then it makes no sense to talk of most of our feelings. The claim should not be that we cannot be critical of ourselves in *most* respects.

Some feelings are more essential to a person or deeper than others. Perhaps one could condemn a great number of one's feelings if only the most central and more important ones are endorsed. Perhaps those with an overriding religious or political goal in their lives, who may be labelled religious or political 'fanatics', could reject their previous loyalties and everything that they previously cared about, including their love for their friends and family, without going crazy. For the (partial?) love of God or the political cause may be so deep and central in their psychology that it can be the firm basis for the rejection of all their other desires, without losing sanity. Thus, although those who condemn *all* their feelings and commitments are insane, we may condemn *very many* feelings and commitments without insanity.

However, partial commitments born of love are *not* eccentric, unusual or peripheral, but central, deep and significant commitments of most or all people. To condemn *these* feelings is to condemn something central enough and deep enough in us – central enough to what is sometimes

[8] Bernard Williams, *Moral Luck* (Cambridge, 1981), p. 18

[9] Thomas Nagel, review of William's *Moral Luck*, reprinted in *Other Minds* (Oxford, 1995), pp. 167–73.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

called our 'identity' – to amount to insanity. Very few people are like religious or political fanatics, with one overriding passion that could provide stability and be a source of sanity when the rest of a person's loving commitments are thrown overboard.[10]

The argument is not that it is impossible to be a consequentialist, only that it is impossible without insanity. The notion of insanity I am considering is surely a matter of degree. So the consequentialist could reply that we should do consequentialist actions to the extent to which it does not make you crazy. In principle, I suppose that might be feasible. But love for family, in particular for our parents or children, and by contrast with most friendships, is central enough in our psychology to create a psychological disturbance in any consequentialist. Of course, feelings for family may cool, for whatever reason. But that is not the same as having deep feelings but wishing one did not, and judging that it would be better to be without them. That is insane.

This means that we can argue that those who actively reject their deep or central partial feelings of love (not merely those who fail to reflectively endorse them) have *disintegrated*, since such partial commitments make up the bulk of a person's central concerns in life. Such people are not merely alienated, like juvenile delinquents in 1950s Rock 'n' Roll movies: they are insane. If consequentialists are insane, let us be kind and caring towards them; but we need not take what they say seriously.

IV. NIHILISM, CONSEQUENTIALISM, MADNESS AND PRACTICALITY

It might be replied – why is insanity a problem? Cannot a madman tell the truth? Can a madman not know something important that the rest of us know not?[11]

Compare a radical nihilistic view, which in one way is far from impartial consequentialism, but which in another way shares quite a lot with it. On this nihilistic view, *all* positive normative claims are false, and the truth about life is utterly and unrelentingly bleak. Those who appreciate this fact may suffer motivational breakdown, or even want to kill themselves, or at any rate not strive to stay alive. Nevertheless, such people may have a better grasp of the negative normative facts, and the true awful bleakness or absurdity

[10] Many of Rabindranath Tagore's short stories and novels explore with great insight the psychological and moral costs of political and religious fanaticism. For instance, see his *The Home and the World* (Harmondsworth, 2005) and *Selected Short Stories* (Harmondsworth, 2005).

[11] Compare Friedrich Nietzsche, *The Gay Science*, trans. Josefine Nauckhoff (Cambridge, 2012), bk. III, sect. 125.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

of human existence, than the rest of us who are sane. Perhaps nothing matters even though believing anything, including believing that nothing matters, presupposes that something matters. If so, the doctrine that nothing matters is pragmatically self-undermining. But someone might say that it could still be true. Perhaps it is not possible sanely to believe that nothing matters, for in believing anything we are aiming to get things right, or believe things on reasonable grounds, which presupposes that this is a worthwhile endeavour. But someone who believes normative nihilism might still be right. So, it might be argued that the idea that the normative truth must be believable without insanity is questionable. A madman may have something to teach us. And consequentialists might be such seer-madmen who bear uncomfortable truths.

However, the madness that is consequential on believing consequentialism is particularly troublesome and unlike the madness generated by normative nihilism. Things are worse for consequentialists than for normative nihilists. It is true that normative nihilists critique fundamental psychological dispositions, all of which turn out not only not to be rational but also to lack any other positive value, and this is like the consequentialist assault on our emotions. Both attack the mind at its core. But the consequentialist is still in the normative game, unlike the normative nihilist. The problem with consequentialism, to repeat, is the huge gap between centrally important feelings of love and loyalty and self-reflective judgements about those feelings. This gap constitutes a specific kind of madness, and in *this* respect the situation is the same as it is with our judgements for the normative nihilist. However, normative nihilists make no positive normative judgements, only negative ones – they judge only that things do *not* have this or that value. By contrast, consequentialists make the judgement not just that it is not the case that we should have partial emotions, but that we definitely should lack them. This is a different matter. They make the positive judgement about themselves that their feelings possess negative characteristics. CSM allows that this or that aspect of ourselves might be detestable. Some of our characteristics might be flawed or corrupt or decadent or sinful or wrong. But consequentialism goes a long way beyond that. By contrast, normative nihilists do not condemn themselves. In fact, they are indifferent to themselves along with everything else. But for consequentialists, it is not that their feelings lack value but that they have negative value. Thus consequentialists are at war with themselves in a way that normative nihilists are not. The consequentialist (unlike the normative nihilist) is therefore someone who normatively assaults themselves – like someone with a large tattoo on their face. Consequentialism is a kind of violence against the self!

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

Compare those who condemn their own sexual or eating desires. They have their desires; but they also think that their feelings concerning sex or food are wrong or somehow traitorous. Thus they are at war with themselves in that they disapprove of their own desires and of their acting on them. Consequentialists are similar, except that their self-fragmentation, the inner civil war, runs deeper.

It is this *kind* of insanity that consequentialism invites. Consequentialism and normative nihilism are similar in that neither is liveable in the sense that both cannot live any recognizable kind of life while believing and acting in accordance with the theory. But there is also a difference. Only consequentialism says that human nature is flawed, that we are built wrongly. Normative nihilism does not say that. For normative nihilism, our (positive) normative *judgements* are all false; and in that respect, *they* are flawed. But our psychological *dispositions* are not *flawed*. They are just *not virtuous*. By contrast, for consequentialism, our deep emotions are flawed. Our emotions do not just lack the value that CSM takes them to have; instead they have a negative value, since they conflict with and draw us away from maximizing the good. For consequentialists, human nature has something like the Christian idea of original sin, except worse. Our deep emotions are major flaws in our nature, and this is unlike our psychological dispositions for a normative nihilist.

Despite all this, could consequentialism be true? There are certain precedents in the history of thought for the idea that there are some truths that only a madman can know: the ancient Greeks had a notion of divine madness; the surrealists celebrated insanity as a source of insight;[12] and many have thought that drugs could open 'the doors of perception' yielding insight into a higher or deeper reality.[13] The idea would be not merely that there are some things that one can only come to know if one is mad, but the stronger idea that only a madman can know those things, and one cannot take this knowledge back with one as one regains sanity. One loses knowledge as one regains sanity.

This does seem possible. However, at this point it is not too controversial, I believe, to assume some kind of weak relation between morality and life. Morality is 'practical' in the sense that it prescribes how to live. So, the true moral theory must be one that we *could* obey, or at least that we *could* believe about ourselves. (This is one sense in which 'ought' implies 'can'.) This is not to say that consequentialism implies that we *should* believe consequentialism about ourselves, in addition to doing what consequentialism prescribes,

[12] André Breton, *What is Surrealism? Selected Writings*, trans. Franklin Rosemont (New York, 1969).

[13] See, for example, Aldous Huxley, *The Doors of Perception* (London, 1954).

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

since, for consequentialism, it is consequences not truth that dictate what we should believe. The point is just that it must be *possible* both to believe it about ourselves and to act on it – without going insane by disintegrating as a person. But we cannot do that. So how could such a moral view be true? If a person would have to be mad to both believe it and act on it, how can it be true? It cannot, unless consequentialism is a totally abstract doctrine with absolutely no relevance to life.

It might be replied that since self-fragmentation is a matter of degree, a consequentialist could reply that we should maximize the good up to some limit of psychological possibility. That is, we should do optimific acts to the extent to which it does not make us crazy – after which the ought-implies-can principle means that we are no longer obligated to promote the impartial good. The objection is not that any people who are under consequentialist obligations and who also act on them are made crazy; the problem is just for consequentialists who ought to be able to self-apply their theory. Even if so, the reply could be that *they* can say of themselves that they are under consequentialist obligations only up to some psychological limit. However, the general problem is that those who have normal partial concerns face a clash between rationality and morality, which is somewhat ironic considering that consequentialists often try to sell consequentialism as a rational doctrine. If consequentialists have any human emotions and thus partial concerns, indeed if they even have prudential concerns, they must condemn themselves, and these emotions and concerns must be ignored in practical reasoning, or at least treated as a drop in the ocean of the good of seven billion humans, not to mention all the non-human beneficiaries. Thus the point of psychological insanity will arrive very soon indeed, not during some leisurely moment after a morning shower and a cappuccino.

By saying that the right moral theory must have some connection with life, I do not wish to imply that we necessarily or always care about morality. People care more or less about morality. One can even, to a great extent, be indifferent to morality.[14] But when we *do* care about morality to a significant degree, we must be able to make it 'practical', where that means factoring morality into our practical reasoning. We ought to be able to do that. But we cannot make consequentialism practical in the sense of believing it and factoring it into our practical reasoning, because that would involve generating a huge dislocation between our emotional life and evaluation. We might, I suppose, will

[14] Nick Zangwill, 'The Indifference Argument', *Philosophical Studies* 138 (2008), pp. 91–124; Nick Zangwill, 'Motivational Externalism: Formulation, Methodology, Rationality and Indifference', *Motivational Internalism*, ed. Gunnar Björnsson et al. (Oxford, 2015), pp. 44–60.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

ourselves to be irrational as a side-effect of believing the right theory. But making consequentialism practical would mean embracing and sustaining the consequentialist moral belief as part of our regime of practical reason. It would mean believing consequentialism and resolving to act on it and intending to act on it. But reasoning with a theory that one knows requires insanity is like fighting for peace or fornicating for virginity! Hence consequentialism cannot be true despite the fact that it generates insanity, which means that it will be hard to factor consequentialism into our practical reasoning. And if we cannot do that then there is no connection at all between morality and life. But there is, so consequentialism is false.[15]

What about indirect consequentialism – the cousin of direct consequentialism? Is indirect consequentialism psychologically sustainable? Indirect consequentialism may be *intellectually* problematic if there turns out to be too much of a distance between impartial consequences and the partial grounds of loyalties. But I cannot see that it is *pragmatically* impossible to accept it. It does not create a major fissure between feeling and judgement that is psychologically destabilizing, as it does for the direct consequentialist. It is hard to see how, in the pragmatic Moore-paradoxical sense or cogito-negation sense of impossibility, it is impossible to accept the proposed indirect consequentialist account, since it is an account designed to vindicate the loyalties endorsed in CSM. So, indirect consequentialism is not threatened by this argument.

V. COMPARISON WITH OTHER WRITERS

How does what I say here relate to what others have said, along similar lines?

Henry Sidgwick and Derek Parfit consider whether consequentialism (or 'utilitarianism') generates the paradox that for consequentialist reasons the theory itself should not be believed by most people.[16] So, consequentialism is an 'esoteric' morality that most people should not believe. While there is perhaps a certain tension here, the issue is not a distinctively first-personal one, as my issue is. Sidgwick worries that consequentialists must say that most people should believe non-consequentialist morality, but, like Parfit, he thinks that the problem is not fatal. My problem, by contrast, is with the remainder of enlightened

[15] For a different argument but with some parallels, see James Lenman, 'Consequentialism and Cluelessness', *Philosophy & Public Affairs*, 29 (1999), pp. 342–70.

[16] Henry Sidgwick, *The Methods of Ethics* (London, 1874), pp. 489–90; Derek Parfit, *Reasons and Persons* (Oxford, 1984), sects. 10–11 and 17.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

consequentialists who believe the theory. I worry about *them*. Sidgwick does not see a difficulty there.

It is a familiar point that modes of deliberation can be assessed in terms of truth or in terms of benefits, and consequentialists can say that it is good when people, other people, accept and deliberate in non-consequentialist terms, even though the non-consequentialism that those people accept is not true. By contrast, my problem is with the self-application of consequentialism. Those maintaining consequentialism have deep feelings that they must condemn. Consequentialists say of themselves that they should act contrary to their deep loyalties, and that they must also condemn those loyalties, since they encourage wrong actions. It is irrelevant what consequentialists say about *other* people. It is what they say about *themselves* that is the problem. The problem for consequentialists arises from the negative judgements that they make about their own feelings. The fact that consequentialists can deny that *other* folk should believe consequentialism, and say that *they*, the other folk, should deliberate in non-consequentialist ways, is irrelevant.

Jackson's 'decision theoretic consequentialism' (which we encountered in section I) posits odd people who have the goal of maximizing the impartial good, and Jackson then argues that given their states of information and facts about their commitments it could be rational for them to give special attention to their friends and family.[17] This theory is not consequentialist, but it is relevant to the my complaint against consequentialism, since I ask what the situation would be if consequentialists were to apply their theory to themselves. I have said that they must condemn their own feelings because they must condemn their own actions. Jackson denies this. He claims that we are justified in aiming to produce benefit for friends and family because of our greater knowledge of how to produce benefits for them.[18] This is in some sense possible, but it is incredibly unlikely that an act causing modest partial benefit in fact maximizes the good more than many alternatives, given that many people are starving in need of medicine overseas, and many people locally, who the agent could easily find out about, would be benefited far more. Given our resources and our state of information it is very unlikely to be the case that the impartially optimific act conveniently happens to be helping my friends or family. Hence my problem arises, since a consequentialist should condemn both their own partial actions as well as their own partial emotions as desires. The insanity problem cannot be avoided.

[17] Jackson, 'Decision-theoretic Consequentialism'.
[18] Jackson, 'Decision-theoretic Consequentialism', p. 235.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

The writer closest to my own thoughts is Bernard Williams. Nevertheless, there are several significant differences between his arguments and mine. His 'integrity' objection is that if consequentialism is true, agents must *act* contrary to their own deep moral beliefs and the moral feelings (their 'projects'). This worry is that the consequentialist conception of human *action* is incoherent, since, somehow, we must act on other people's desires, not on our own desires. This is, Williams says, 'absurd', which an interesting choice of words.[19] Only if I am a weird agent who only desires the impartial good would action make sense. This *is* a first-personal worry (unlike what Sidgwick and Parfit discuss), but it concerns action. I have great respect for this argument. Indeed, I think it is profound. But it is not the argument that I make here. What I worry about is a tension between consequentialist first-personal judgements about their own feelings of love and loyalty, and those very feelings of love and loyalty. This is not the distance between moral judgement and *action*.

At another point, Williams complains: 'such things as deep attachments to other persons will express themselves in the world in ways which cannot at the same time embody the impartial view, and that they also run the risk of offending against it'.[20] But given the most likely reading of 'express', this is, again, a conflict between the actions that the feeling tends to produce with the actions recommended by consequentialism, not a conflict between consequentialism and reflective judgements about feelings.

Another argument in Williams is that we cannot combine intuitive and reflective modes of thought (or 'first-' and 'second-order' thoughts).[21] We cannot combine the intuitive judgement that I should act out of love and loyalty, with the reflective consequentialist judgement that acting according to love and loyalties is right only as a means to the consequentialist goal. But this is a collision between two judgements, not between judgement and feeling.

There are other passages where Williams does address the feelings themselves and their conflict with consequentialism. But there he seems to want to separate feeling from *all* judgement; he thinks that *any* judgements about feelings for one's wife are 'one thought too many', not just consequentialist judgements.[22] I previously agreed with Nagel, against Williams, that Williams is wrong to suggest that we

[19] Bernard Williams, 'A Critique of Utilitarianism', in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge, 1973), pp. 77–150, at 116–17.

[20] Bernard Williams, *Moral Luck* (Cambridge, 1981), p. 18.

[21] Bernard Williams, 'The Structure of Hare's Theory', *Hare and his Critics*, ed. Douglas Seanor and Nicholas Fotion (Oxford, 1988), sects. 14.4 and 14.6.

[22] He wrote: 'his motivating thought, *fully* spelled out, would be the thought that it was his wife' (*Moral Luck*, p. 18; my emphasis).

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

should not make reflective judgements about ourselves, judgements that endorse our loves and loyalties. We should have the extra thoughts that Williams thinks are too many.[23] And if we have them, why should they not play some role in motivation? My worry is about negative evaluations of central feelings. I am not worried about the grounds for such evaluations (although I think this is relevant to indirect consequentialism).[24] It is the misfit between feelings and judgements about these feeling that concerns me. Too large a misfit constitutes insanity. My problem arises only given the extra reflective thought that Williams thinks we should not have.

Williams's focus is not on the fit between reflective judgements about feelings and consequentialism, which is what concerns me – except perhaps in one passage, where he says: 'These feelings represent something he in general stands by'.[25] A contrast is drawn with the feelings of the lapsed Catholic, who presumably does not 'stand by' (endorse) his feelings. Only in this one sentence does Williams hint at a problem that has its source in second-order reflective endorsements of deep feelings. If so, this article is a spelling out and defence of the objection suggested in this one sentence.

While there are affinities between Williams's various objections and mine (they are distinctively first-personal, for example, unlike what Sidgwick and Parfit discuss), the arguments are, except perhaps for one sentence, substantially different.

VI. CONCLUSION

The conclusion is this. Direct consequentialism is a claim that can be understood, even by those who reject it. The problem is to believe it. Perhaps a purely intellectual being could believe it. But believing direct consequentialism about *ourselves* means retaining our feelings of loyalty while rejecting them intellectually; and that is a problem if we are to stay sane. It is not pragmatically possible. There is surely something strange about a doctrine that we cannot accept without our partial commitments themselves being undermined. If it is not possible to believe the deeper moral truth about our partial commitments without undermining them, then it is not possible to believe the moral truth without setting up a major psychological disturbance, one that arises from the gap between deep feelings and consequentialist judgements.

[23] Williams, *Moral Luck*, pp. 17–19.
[24] See Zangwill, 'Cordelia's Bond and Indirect Consequentialism'.
[25] Williams, *Moral Luck*, p. 52.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press

If direct consequentialism is true, it must be possible to believe it. If it is possible to believe direct consequentialism, it must be possible to condemn one's partial feelings of love. But there is a problem about condemning deep and central feelings while remaining sane. By contrast, indirect consequentialism does not generate this difficulty, since, although we must endorse our deep feelings for surprising reasons, it does not open a chasm between our feelings and our reflective intellectual life. It is this chasm that makes it difficult for us to accept direct consequentialism about ourselves. Indirect consequentialism is self-believable because it does not contain the requirement to condemn our own feelings, only our grounds for acting on those feelings, and that might create intellectual discomfort, but no grand lunacy – no desecration of the soul.

The direct consequentialist cannot reply that such self-immolating insanity is fine and something we must learn to live with, as the price of believing the true moral theory; for direct consequentialism is not a doctrine that anyone can really succeed in believing about themselves and survive. So severe is the dislocation that believing direct consequentialism yields that only madmen, lunatics, crazy people, could really believe it about themselves. But the true moral theory must be sanely self-believable.

It is a controversial principle in metaphysics and epistemology that [*p* implies that it is possible to know that *p*]. It is also controversial that [*p* implies that it is possible to believe that *p*]. Perhaps there are true propositions that are beyond our power to comprehend. Perhaps it is even controversial whether if we can grasp a proposition then it is possible to believe it (for example, 'I do not exist'). But there is surely something dubious about a proposition that if believed makes one crazy!

Monty Python's Flying Circus featured the idea of a lethal joke, which is so funny that anyone who hears it dies of laughter. And Douglas Adams's *Hitch-hiker's Guide to the Galaxy* featured the idea of Vogon poetry, which is so terrible that if one hears it one dies in aesthetic agony. Similarly, one cannot believe direct consequentialism about oneself, without going crazy. Direct consequentialism is like a lethally funny joke or Vogon poetry. Believing direct consequentialism is the road to the madhouse! Direct consequentialism is an insane doctrine.[26]

nick.zangwill@hull.ac.uk

[26] This article was presented at Torquad Di Tella, in Buenos Aires. Thanks for helpful questions from Horacio Spector and Eduardo Rivera Lopez. Thanks also for comments from Brad Hooker, Geoffrey Scarre and a referee for this journal.

https://doi.org/10.1017/S0953820817000267 Published online by Cambridge University Press