Williams's Integrity Objection as a Psychological Problem

Nikhil Venkatesh[1]

Abstract

Utilitarianism is the view that as far as morality goes, one ought to choose the option which will result in the most overall well-being—that is, that maximises the sum of whatever makes life worth living, with each person's life equally weighted. The promise of utilitarianism is to reduce morality to one simple principle, easily incorporated into policy analysis, economics and decision theory. However, utilitarianism is not popular amongst moral philosophers today. This is in large part due to the influence of Williams's 'Integrity Objection' (1973). Though the Integrity Objection has been influential in turning philosophers against utilitarianism, it is also difficult to make precise, evidenced by the myriad interpretations in the literature. In this paper I interpret the objection as holding that agents who accept utilitarianism cannot, as a matter of psychology, be committed to their projects. I explore other interpretations, finding some to be inconsistent with Williams's approach, and others to be relatively easily answerable by the utilitarian. The psychological problem I identify is harder for utilitarians to avoid, though I have begun to offer a response in other work.

1 Introduction

Utilitarianism is the view that as far as morality goes, one ought to choose the option which will result in the most overall well-being—that is, that maximises the sum of whatever makes life worth living, with each person's life equally weighted. The promise of utilitarianism is to reduce morality to one simple principle, easily incorporated into policy analysis, economics and decision theory. However, utilitarianism is not popular amongst moral philosophers today. This is in large part due to the influence of Williams's 'Integrity Objection' (1973).

Though the Integrity Objection has been influential in turning philosophers against utilitarianism, it is also difficult to make precise, evidenced by the myriad interpretations in the literature. In this paper I interpret the objection as holding that agents who accept utilitarianism cannot, as a matter of psychology, be committed to their projects. I explore other interpretations, finding some to be inconsistent

with Williams's approach, and others to be relatively easily answerable by the utilitarian. The psychological problem I identify is harder for utilitarians to avoid, though I have begun to offer a response in other work.

2 Cases, Projects, and Hare's Interpretation

Williams's objection is made through two hypothetical cases (1973, 97–99). In one, a recently graduated chemist, George, is offered a job in a chemical and biological warfare (CBW) laboratory. He decides that he cannot accept, since he is opposed to CBW. He cannot accept even though his unemployment causes him and his family to suffer, and even when he is told that the person who would be hired in his place would pursue the research in such a way that more dangerous chemical weaponry would result. In the second case, Jim, in a foreign land in the aftermath of an uprising, is made an offer by Pedro, an army captain. Pedro will execute twenty innocent prisoners as a warning to dissenters unless Jim agrees to shoot one himself, in which case the other nineteen will be released.

Williams's use of cases invites the interpretation that he aims to dismiss utilitarianism on the grounds that it gives Jim and George the wrong advice. This is Hare's interpretation

✉ Nikhil Venkatesh
    n.venkatesh2@lse.ac.uk

1   Department of Philosophy, Logic and Scientific Method,
    London School of Economics and Political Science, London,
    UK

([1981](), 49, 130–46). Utilitarianism recommends, given some tacit assumptions (that George wouldn't be so depressed by taking the job that he and his family are caused more suffering; that Pedro's prisoners have lives worth living) that George takes the job and Jim shoots the prisoner. Those who believe that one should never assist with CBW, or kill, will disagree. But Williams does not want to persuade only people with such moral beliefs—and is not one of them himself, remarking that 'the utilitarian is probably right' in Jim's case ([1973](), 117). He disclaims Hare's interpretation of the Integrity Objection, writing: 'the objection did not, however, take the form of my trying to disprove a theory by counter-example, as much of the discussion has assumed.' ([1995](), 212) Instead, the cases are meant to make salient a certain feature of moral life, consideration of which reveals utilitarianism to be defective. That feature is integrity.

Williams introduces his discussion of integrity by considering.

> 'the idea, as we might first and simply put it, that each of us is specially responsible for what he does, rather than what other people do. This is an idea closely connected with the value of integrity. It is often suspected that utilitarianism makes integrity as a value more or less unintelligible. I shall try to show that this suspicion is correct.' ([1973](), 99).

He goes on:

> 'the reason why utilitarianism cannot understand integrity is that it cannot coherently describe the relations between a man's projects and his actions.' ([1973](), 100).

We can make out Williams's claim of a 'close connection' between integrity and responsibility, and hence his notion of integrity, in terms of these relations. There is, for Williams, a deep difference between how I relate to my actions and to those of other people, even when I can prevent or encourage the latter. Jim could shoot one person or reject Pedro's offer in which case Pedro will shoot twenty. George could advance CBW or reject the job, allowing his rival to advance it in more dangerous directions. Whether the second state of affairs is realised, in both cases, is in the gift of Jim and George. But it would be misleading, according to Williams, to think of Jim and George as having brought about these states of affairs if they are realised. It would be misleading to think that they will merely have had an 'effect on the world through the medium… of Pedro's [or the unnamed rival chemist's] acts.' ([1973](), 109) In Williams's view their responsibility for those states of affairs is therefore lesser and qualitatively different; morality respects a distinction between my actions, and actions that are not mine but whose occurrence I have control over. What could account for this distinction? Williams points to the relationship between actions and projects. If George takes the job, he adopts the development of chemical weapons as a project, and accordingly conducts the relevant research. If he doesn't, the other chemist would adopt the same project, and pursue the same research, but there would not be the same relationship between George's projects and the research. (Indeed, George could retain his project of opposing CBW.) If Jim were to reject Pedro's offer, twenty people would die. But this would not be because Jim had a project that aimed at their deaths, but because Pedro did. Their deaths in this case would thus be best described as a killing by Pedro, not by Jim, for all the opportunity Jim had to save them. We are 'identified', as Williams says ([1973](), 116), with the actions that 'flow' from our projects.

What if Jim were to accept Pedro's offer? Although Jim would pull the trigger, it would seem wrong to hold him responsible. Pedro's coercion effectively turns Jim into a medium through which Pedro affects the world. This intuitive description of the case is reflected at the level of projects: the killing is the aim of Pedro's projects, not Jim's. Jim's identification with, and moral responsibility for, such an action would be attenuated.

That 'each of us is specially responsible for what he does', then, means that each of us is specially responsible for the actions that flow from our projects. They must flow from our projects in the right way: if my pursuing some innocent project outrages you so much that you lash out at me, I am not responsible for this, even though my project is part of the cause of your lashing out.[1] Rather, actions we are identified with must flow from our projects in a way that is directed by those projects, as Jim's shooting of a prisoner would be directed by Pedro's aim of intimidating dissenters. This—not the fact that we had the opportunity to determine whether the action was performed or not—is what makes those actions ours and not someone else's. To neglect this connection is to attack our integrity. 'Integrity' here is meant in the sense of wholeness or unity—an agent's integrity is the unity between them, their projects and their actions.

Utilitarianism seems to neglect this connection. Utilitarianism provides a criterion of choice between options: what one should do, according to utilitarianism, is determined by the effects on well-being of each option available to you. An agent's options are whatever they are able to realise: this is not limited to actions flowing from their projects. Furthermore, utilitarianism is indifferent between different paths to the same sum of well-being, as reflection on Williams's cases demonstrates.

Jim has two available options: (1) accept the offer and shoot one prisoner; (2) reject the offer and see Pedro shoot twenty. Choosing (1) will lead to more well-being than

---

[1] Thanks to Véronique Munoz-Dardé for this caveat and example.

choosing (2), in normal circumstances (the prisoners will have lives worth living if they survive, bereavement reduces well-being, and so on). So utilitarianism recommends (1). But notice that the very same reasoning would apply if Pedro were not in the picture. Imagine Jim had a choice between (1) shooting one person and (2') shooting twenty people himself. The effects on well-being are equivalent (except perhaps for differences in guilt felt by Jim and Pedro) across (2) and (2'). For utilitarianism, therefore, these choices are equivalent: Jim's choosing (2) is as bad as his choosing (2') would be. This is so even though in (2) the killings would flow from Pedro's projects, and in (2') they would flow from Jim's. Utilitarianism therefore seems to neglect the significance of the agent's identification with their actions through their projects.

What is a project for Williams? He gives no explicit definition. He gives examples (1973, 110–11): desires for oneself, one's family and one's friends to have the basic necessities of life, and for the 'objects of taste'; 'pursuits and interests of an intellectual, cultural or creative character'; political causes such as Zionism; 'projects that flow from some more general disposition towards human conduct and character, such as a hatred of injustice, or of cruelty, or of killing'; the utilitarian project of maximising well-being. A project, to be something from which action may flow, must be capable of motivating the agent who has it. The motivational aspect of projects is reaffirmed by Williams when he says that if we 'step aside' from our projects, we are alienated 'from [our] actions and the source of [our] actions in [our] own convictions.' (1973, 116) But projects cannot be whatever motivates action—a habit or addiction would not be a project. We are conscious that our projects guide our actions (unlike mere habit) and approve of them (unlike addictions).

3 Commitments and Ashford's Interpretation

These two passages form the crux of the Integrity Objection:

> how can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life, just because someone else's projects have so structured the causal scene that that is how the utilitarian sum comes out?

> 'It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires.' (both passages Williams 1973, 116).

In the second quoted passage Williams contrasts the agent's 'own project and decision' with the utilitarian recommendation. This is a false dichotomy. An agent could adopt utilitarianism itself as a project—indeed, Williams himself considers this possibility just one page before. For such an agent, actions performed as utilitarianism requires would flow from one of the agent's projects, and so would really be the agent's own. As long as such an agent is possible, there is no necessary opposition between identified action from one's own project and decision and acknowledging utilitarian recommendations.

However, as the cases of George and Jim show, utilitarian recommendations can conflict with other, non-utilitarian projects. Williams's protagonists cannot simultaneously follow the utilitarian recommendation and their projects of opposing CBW and refraining from killing. Williams alleges an absurdity in demanding that someone step aside from their projects. If to 'step aside from' a project is simply to perform some action antithetical to it, then what seems absurd to demand is that morality never ask one to step aside from one's projects. A project, as we have seen, could be a simple desire or taste. There are surely occasions in which we ought to forego satisfying one of our desires to help someone else from a greater suffering. Williams's objection would be implausible if it rejected 'any morality not based on the accident of the agent's own projects' (Raz 1986, 287).

But not all projects can be so easily put aside. Williams is especially interested in the subset of projects he calls 'commitments'. What distinguishes commitments from other projects is left vague, but has to do with the greater strength of the attitude one has towards them, hinted at by words like 'thorough', 'deep and extensive' and 'serious'. 'One can be committed', Williams writes, 'to such things as a person, a cause, an institution, a career, one's own genius, or the pursuit of danger.' (1973, 112) A commitment is not simply a very strong desire, though; it is a project which in some way defines the person who has it. Consider the desire to eat: when one is very hungry it may be overwhelmingly strong, but it is hardly something that defines one's character and shapes one's life. Williams writes that one could treat a cultural pursuit as a commitment. One's relationship to that pursuit would be 'at once more thoroughgoing and serious than their pursuit of various objects of taste, while it is more individual and permeated with character than the desire for the basic necessities of life.' (1973, 111) Enjoying the tune of some aria does not count as a commitment, even if it motivates you to go to an opera. Being an opera-lover, on the other hand, which involves educating oneself about the history and subtleties of the form, keeping oneself abreast of current productions, watching and listening to opera frequently, defending its value in argument, and so on, could be a commitment. Insofar as there is a distinction between an opera-lover and someone who enjoys the opera, it seems

that for the former their relationship with opera has permeated their character, such as to become partly constitutive of their identity. If being an opera-lover is related to us in this way, and essentially involves certain actions, then performing those actions is essential to our being who we are. This means that a different level of integrity is at stake in the actions flowing from our commitments. Actions flowing from our projects are ours; actions flowing from our commitments are not only ours, they are us.

Is the integrity objection, then, that utilitarianism asks us to step aside from our commitments, and that this is too demanding, given their connection with our identity? This is Elizabeth Ashford's (2000) interpretation of Williams's objection. She responds that any plausible moral theory will ask us to step aside from our commitments in emergencies, and that—as Williams himself suggests elsewhere (1985, 186)—such situations are common in the actual world, given the extremes of poverty and wealth in an ever more closely connected global economy. Thus, she says, though utilitarianism might ask us to put aside our commitments to save lives, this is not an unreasonable demand. I am sympathetic to Ashford's view. However, I think Williams's objection goes beyond it. Utilitarianism is incompatible with commitment in other ways that her response does not fully answer. Chief among these is a problem of the psychology of the utilitarian agent.

4 The Psychological Problem

Williams believes, I think, that if an agent accepts utilitarianism they will be incapable of having commitments. This is because of two facts about the psychology of the utilitarian agent: they regard their projects impartially, and they regard them as dispensable.

To regard projects impartially is to refrain from valuing one project more than another simply because of whose it is. Most importantly, someone who is impartial does not value their own projects more than those of others simply because they are their own. This captures Williams's description that the projects of a utilitarian agent are, to them, 'one satisfaction among others'.

Why does utilitarianism require such impartiality? As Scanlon puts it (1998, 95–100), to value X is to take oneself to have reasons for certain attitudes and actions towards X. In Scanlon's view these attitudes and actions may be several, and may vary depending on what X is. A crucial part of valuing projects is taking oneself to have to pursue them (when they are one's own), and to assist in them (when they belong to others). To value project Y over project Z, then, would be to take oneself to have stronger reasons to pursue/assist in Y than Z. For a utilitarian (as far as morality goes) one has stronger reasons to do one thing than another if and

only if it would result in greater overall well-being. I should do what leads to more well-being, whether that involves acting on my projects or those of a stranger. But if I value my own projects more, simply because they are mine, I take myself to have more reason to pursue them than to assist in others, even when doing the latter would result in greater well-being. Such an action would be wrong, according to utilitarianism.

Insofar as one employs utilitarianism as a decision-procedure, then, one regards one's projects impartially. But if I deliberate in this way, how is my project my project? It seems obvious that if X is my project and Y is not, I must regard X in a different light to Y (typically as more valuable) and be generally disposed to act on X rather than Y. So for the agent who accepts utilitarianism, their projects seem to have a double life: they are both that agent's projects, special to her, and they are, according to utilitarianism 'one satisfaction among others'. The utilitarian agent's actions do not flow from these projects, but rather from well-being calculations that take everyone's projects into account on an equal basis. This, I think, is what Williams means by alleging that utilitarianism 'cannot coherently describe the relations between a man's projects and his actions'.

Williams writes that 'in the case of many sorts of projects' it is 'perfectly reasonable' to weigh the utility gains of your satisfying your project against the gains of someone else satisfying theirs when the two conflict (1973, 115–16). This not only permits a moral theory to ask us to abandon our projects on occasion, it also affirms the utilitarian's impartial method of counting one's projects as 'one satisfaction among others'. But Williams thought such impartiality was impossible with respect to commitments, even if reasonable for other projects. As he puts it concerning a subset of commitments, moral convictions:

> 'we… cannot regard our moral feelings merely as objects of utilitarian value… to come to regard those feelings from a purely utilitarian point of view, that is to say, as happenings outside of one's moral self, is to lose a sense of one's moral identity; to lose, in the most literal way, one's integrity.' (1973, 103–4).

So much for impartiality. What is it to regard a project as dispensable—and why does utilitarianism require it, and commitment preclude it?

It is important to note that regarding is an attitude, not an action or disposition to action. It is not that one *can* never dispense with a project to which one is committed. (Though a commitment may give rise to what Williams terms 'moral incapacity' (1992) or 'practical necessity' (1981b), which is the sense in which it is true that George *cannot* take the job—see next section for more discussion.) Nor is it that one *should* never dispense with one's commitments (Williams thinks Jim should do just that). The modality involved is

about how agents see their possibilities: as Williams says, his cases show 'most importantly of all, what would be implied by certain ways of thinking about the situations'. (1973, 96)

What is it to regard a project as dispensable? For Williams, I think, *to regard a project as dispensable is to entertain as alternatives outcomes in which one dispenses with it.*

What is it to dispense with a project? One might imagine that it involves ridding oneself of positive attitudes towards it and the associated patterns of motivation. But that would not address the cases of George and Jim. They are, Williams thinks, required by utilitarianism to dispense with their commitments. But they are not prevented from continuing to believe in the wrongness of CBW, or of killing, nor from living their lives, beyond these tragic episodes, in accordance with those beliefs. Thus for Williams to dispense with a project does not necessitate fully ceasing to believe in and pursue it. Performing certain one-off actions that are to a sufficient degree at odds with it—such as killing, with respect to the commitment not to kill ('stepping aside' from the project, as Williams puts it)—also counts as dispensing.

Now, if agents were unable to conceive of circumstances in which they would dispense with their commitments, commitments would be unattractive. For any project you have, I can ask you to imagine that Satan has promised to wreak untold suffering on humanity if you do not dispense with it. If you held that even in such circumstances you would not dispense with their commitments, you would be not principled but dangerously fanatical.

Williams's position was not as implausible as this. By 'entertaining as alternatives' he does not mean merely conceiving of them. He writes:

> 'it could be a feature of a man's moral outlook that he regarded certain courses of action as unthinkable, in the sense that he would not entertain the idea of doing them… Entertaining certain alternatives, regarding them even as alternatives, is itself something he regards as dishonourable or morally absurd. But, further, he might equally find it unacceptable to consider what to do in certain conceivable situations. Logically, or indeed empirically conceivable they may be, but they are not to him morally conceivable, meaning by that that their occurrence as situations would represent not a special problem in his moral world, but something that lay beyond its limits.' (1973, 92).

If Jim and George are committed to their respective projects of not killing and of opposing CBW, Williams suggests, the situations in which they find themselves require a very different way of thinking to that which they employ in other situations. It is not that they cannot, or don't like to, think about situations in which they have to kill, or do military research. It is that their commitments circumscribe a set of alternatives that they are willing to entertain, and this is partly constitutive of their outlooks on life. Their commitments are usually inputs, or constraints, on their deliberation. In the kind of situations in which Jim and George are placed, they find themselves required to deliberate without them, as the commitments themselves are up for debate. What was solid in their thinking melts into air; they are compelled to question what was previously bedrock. This is what makes taking the job, for George, seem 'absurd'.[2]

So for Williams, entertaining some outcome as an alternative is not merely conceiving of it. It is being willing to conceive of it within the constraints set by one's outlook on life. Commitments set such constraints: a committed agent is unwilling to conceive of outcomes in which they dispense with their projects. This does not mean they never do, but that when they do, a novel and (to them) unsettling mode of deliberation is required.

This is not the case for the utilitarian agent. It is a distinctive (and, to some, attractive) feature of the utilitarian outlook on life that it does not shirk difficult decisions, applying one simple formula to all moral choices. The cases of Jim and George are to utilitarians, like all cases, cost-benefit problems—with the sad fact that one of the costs is the agent's dispensing with a project. The only inputs to utilitarian deliberation are facts about the well-being that a course of action will produce, and, as the cases of Jim and George suggest, there will always be possible alternatives in which dispensing with a project maximises well-being. Therefore, the utilitarian must entertain alternatives in which they abandon any of their projects.[3,4]

---

[2] For a similar (though more general) account of absurdity, see Nagel (1971).

[3] Frankfurt (1988, 180–81) argues that a utilitarian may be so sure that a project of theirs will never be inimical to well-being that they do not entertain such outcomes, and that even if they did they may be sure that in such circumstances they would not be able to bring themselves to dispense with it. Though both phenomena are possible, I don't think they save utilitarianism. An agent who is as Frankfurt describes would probably not be complying with utilitarianism (because such surety is unlikely to be warranted).

[4] Other interpretations of Williams that are similar to mine include Harcourt (1998), Mulgan (2001, 15–16) and Tanyi (2015, 502–5). Harcourt also offers a second interpretation of the Integrity Objection, too, which applies only to preference-satisfaction utilitarianism. Mulgan and Tanyi only mention the Integrity Objection briefly, as they focus on the demandingness objection from which they (correctly) distinguish it. Another interpretation in the neighbourhood is that of Raz (1986, Chaps. 11–13). On Raz's view, it is constitutive of commitment that one does not regard that to which one is committed as commensurable with other goods. Commensurability and dispensability may be related. However, I see no evidence that Williams's argument depends on incommensurability, or that he thinks that it is the commensurability, rather than the dispensability, of projects which condemns utilitarianism.

The psychological problem, then, is this: if we accept utilitarianism then we regard our projects in a way (impartially and as dispensable) that is impossible for us insofar as we are committed to them. Schematically:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

There is a well-worn response to such arguments, which is that accepting utilitarianism is not optimific (either for the reasons above or for others). In this case, the response goes, utilitarianism would direct agents not to accept utilitarianism: it would be 'self-effacing'. Therefore, it is claimed that utilitarianism makes no recommendation that is incompatible with commitment, even if *accepting* utilitarianism is incompatible with having commitments. In other work I cast doubt on this response (Venkatesh 2022). I think that utilitarians would be better off confronting the psychological problem head-on, arguing that utilitarianism does endorse a psychology that is incompatible with commitment, but that it is right to do so.

5 The Normative Problem

The psychological problem reconstructs the Integrity Objection in terms of thought: how utilitarianism requires us to regard our commitments, and how commitments, according to Williams, must be regarded. But commitments require action as well as thought (one cannot be an opera-lover without going or listening to the opera, or a friend to someone without ever lifting a finger to help them). What I will call 'the normative problem' is the allegation that utilitarianism is in tension with the reasons that committed people have for action.

The normative problem is not simply that utilitarianism sometimes requires action at odds with other commitments. Rather, it is that commitments alter 'the normative landscape' (Owens 2012) in a distinctive way that utilitarianism is unable to acknowledge. One way of thinking about such alteration is along the lines of obligation. That one is obliged to do something, many believe, does not merely mean one has a very strong reason to do it. Rather, obligations affect the force of other reasons: they might exclude certain kinds of reasons from one's consideration, for instance. As Scanlon puts it:

'The fact that it would be slightly inconvenient for me to keep a promise should be excluded as a reason for

[not] doing so. Even if I am in great need of money to complete my life project, this gives me no reason to hasten the death of my rich uncle or even to hope that, flourishing and happy at the age of seventy-three, he will soon be felled by a heart attack. Against this, it might be claimed that I do have such reasons and that what happens in these cases is that I conclude that an action (breaking the promise or hiding my uncle's medicine) would be wrong and that the normative consequences of this conclusion then outweigh the very real reasons I have to do it. But this does not seem to me, intuitively, to be correct.' (1998, 156–57).

Utilitarianism, it might be thought, is committed to the position that Scanlon rejects. For the good my uncle's money could do, or the minor inconvenience of keeping a promise, may well affect the sum total of well-being and are thus weighed against the harms of death and of promise-breaking respectively in the utilitarian calculus. On Scanlon's view, the fact that we are obliged not to kill our uncles or break our promises makes such weighing inappropriate.[5] The Scanlonian claim is not that there could never be any reasons to breach an obligation. It is that some considerations do not count as reasons, in the light of obligation. Utilitarianism, it seems, does count such considerations as reasons, so long as they bear on the sum of well-being.

Something similar could be true of commitment. When some action is required by a commitment, this is not simply a strong reason for one to do it. It affects the normative landscape. Other reasons—reasons that utilitarianism takes to be relevant—become irrelevant. Thus, the actions of a committed person are robust: that is, they are performed across a range of circumstances (Pettit 2015). This is because they are not sensitive to small changes in the balance of reasons—commitment makes countervailing considerations that would ordinarily be reasons irrelevant. This accords with a second meaning of integrity, roughly synonymous with incorruptibility; someone is 'a person of integrity' insofar as they stick to their principles in a range of circumstances. A judge with integrity, for example, will deliver fair trials, however much money she is offered to do otherwise. That the money would be nice is not a reason that figures in her deliberation (nor even is the good it could do in famine relief). Similarly, someone committed to a romantic relationship is prepared to stick with their beloved 'for richer, for poorer; in sickness and in health'—that their beloved becomes sick or poor is not a reason for them to end the relationship.

This line of thought relates to what Williams called 'moral incapacity'. This is 'the kind of incapacity that is in

---

5 For slightly different views that accord with this general point, see (Raz 1975, 1999; Owens 2012).

question when we say of someone, usually in commendation of him, that he could not act or was not capable of acting in certain ways.' (1992, 59) George, as Williams puts it, 'cannot' take the job. The idea is not simply that he should not, nor that he will not in these circumstances—although both of these are true—but that it is not possible for him. The impossibility is not unbounded. If George signed the contract when a gun was put to his wife's head and a pen in his hand, one could not say that this undermines his claim to be committed to opposing CBW. Williams writes:

> 'It is plausible to say, with the pessimist, that if having a moral incapacity implies that there are no circumstances at all in which the agent would knowingly do the thing in question, then there are no moral incapacities. Ingenious coercion or brutal extremity can almost always produce such circumstance.' (1992, 69).

For Williams, a moral incapacity is at least 'proof against rewards' (1992, 69)—if a greater salary were offered, George would still refuse the job. Just as inconvenience and money were no reasons, for Scanlon, to breach obligations, rewards are no reason, for Williams, to breach commitments. The normative problem is that utilitarianism recognises considerations that, in the light of commitment, are irrelevant—and is thus incompatible with the robust action that commitment demands.

I believe that utilitarianism can, contrary to appearances, accommodate the normative landscaping of commitment (or at least, the most attractive aspects of it). Utilitarianism acknowledges that commitment can make some facts that would otherwise provide reasons irrelevant. This is the case when commitment makes responding to those facts impossible. Commitments cannot always be disposed of at will. It may be that George really cannot, not simply as a matter of moral incapacity, but physical incapacity too, bring himself to accept the CBW job simply for the promise of a higher salary. That is not to say that he could not physically accept the job, in situations of 'ingenious coercion or brutal extremity', and perhaps in situations like the one from Williams's original example, where it is unemployment and the prospect of his rival conducting more dangerous research that George is asked to regard as reasons to accept. Rather, it is that he could not accept in some range of circumstances short of that, where the putative reasons for doing so are things such as salary. In these cases it may well be impossible for George to accept, and this would be because he has a commitment to opposing CBW. As we saw above, utilitarianism does not ask us to do the impossible—and it need not acknowledge any seeming considerations in favour of the impossible as reasons.

In the case of such a commitment, George's refusal would also be robust: he would refuse in any circumstances except those of coercion and extremity. How about cases in which

George *can* dispense with his commitment? Then, accepting the job is possible for him. However, it is only possible at the cost of his commitment. Fully spelt out, accepting the job involves George not only signing a contract, clocking in, performing experiments and collecting his salary, but also dropping his commitment.

This significantly affects the utilitarian calculation regarding the case. The utility costs of dropping a commitment are likely to be very high. Firstly, consider the disutility to the agent of breaching their commitments. As Railton puts it: 'Commitments… may be very closely linked to the self, and a hedonist who knows what he's about will not be one who turns on his self at the slightest provocation.' (1984, 142) One aspect of well-being, or precondition for it, is a secure sense of identity. The questioning of one's identity—being told that one is not a real philosopher, learning that one was adopted, being misgendered—is distressing. In abandoning a commitment, one calls one's own identity into question, which may be similarly painful. Secondly, consider the disutility to the world, given that the commitment is conducive to well-being. Commitments determine repeated actions across a long period of time. Some commitments (say, to a spouse) lead us to do things (say, to help her in her projects, please her, lessen her burdens, and so on) every day, for the foreseeable future. George's commitment to opposing chemical warfare might not be so frequently active in determining what he does. But it will regularly influence him—every election, every protest march, every conversation he has in which the subject is brought up—for as long as he has the commitment, which could be many decades. If his commitment is conducive to well-being, then it is highly likely that the actions it leads him to perform are themselves conducive to well-being. And if each of these actions is conducive to well-being, then the sum of well-being generated by the commitment that leads George to perform them over and over again is likely to be very large.[6]

If utilitarianism were to recommend that he drop the commitment for the sake of performing just one action (taking the job), then that action must have an even larger positive impact on well-being, outweighing the sum of all the possible actions dependent on the commitment.

---

[6] In the previous section I noted that it is possible to dispense with a commitment, in Williams's terms, whilst retaining some of the attitudes and patterns of action associated with it. Thus George could take the job and still oppose CBW. That this is conceptually possible is important; however, it may not be in actual fact likely. Once he takes a CBW job, George will probably be exposed to more arguments for CBW, be barred from engaging in anti-CBW activism, and above all come to have a vested interested in CBW. This will likely at least temper his opposition to it, and he could foresee this likelihood at the point of deciding whether to accept. As Upton Sinclair found: 'It is difficult to get a man to understand something, when his salary depends upon his not understanding it!' (1994, 109).

So utilitarianism will only recommend that George take the job when (a) he is able to give up his commitment, and (b) taking the job has very high benefits for overall well-being. This may mean that George can, consistent with utilitarianism, refuse the job in a wider range of possibilities than it seems: his action can still be fairly robust. (Recall that Williams's incapacities are not fully robust: they except 'brutal extremity'. Whether the committed utilitarian agent acts more or less robustly than the one with a Williamsian moral incapacity depends on what counts as 'brutal extremity', and (as it should) on the value of the commitment in question.)

Because utilitarianism can accommodate the ways that commitment changes the normative landscape: in making some reasons irrelevant, and in making some actions robust—it can also accommodate the notion of normative powers, that is, that agents can voluntarily change the reasons they have. For instance, when one makes a promise, one places oneself under an obligation, altering the normative landscape accordingly. Likewise, for those commitments we choose to take on, when we take them on we change the reasons we have according to utilitarianism. If George had no commitment to opposing chemical warfare he would be directed by utilitarianism to accept the job in a much wider range of circumstances, because the costs of dropping a commitment would not feature in the utilitarian calculation. And if we can take on commitments that become impossible for us to dispense with, this excludes reasons that would otherwise apply to do things that would involve dispensing with the commitment.

It is worth mentioning the view that, in the light of some commitments, moral considerations of any kind should not figure as reasons. This is Williams's famous 'one thought too many' case (1981a). This is not part of the Integrity Objection—though it is related. Williams presents the case as a problem not just for utilitarianism but for moral theory in general. There are at least two responses a utilitarian can make to this. Firstly, that morality is not the whole of normativity—this is the lesson Wolf (2012) takes from Williams, and one with which the utilitarian can wholeheartedly agree. It may be thought that utilitarianism is less open to such a pluralism than many other moral theories; after all, it is supposed to ground all reasons on one thing, overall well-being. However, utilitarians must qualify this as all *moral* reasons, and recognise other kinds of reasons. Utilitarians tend to assert that it is rational for individuals to maximise their own well-being—indeed, anti-utilitarians often criticise them for doing so[7]—and use this claim in

arguments for utilitarianism, making morality a generalised version of prudential rationality (Harsanyi 1977). So they must recognise reasons that are not about maximising overall well-being, and are not moral. They may also endorse reasons that are neither moral nor prudential, such as reasons stemming from roles, identities and projects. The utilitarian need not deny that a footballer has a reason to shoot at goal, even if doing so would not maximise overall well-being (or even her own). They simply deny that this is a moral reason, or even a prudential one, but one grounded in the sport they are playing. Similarly, those involved in a particular project or relationship might have non-moral, non-prudential reason to do certain things. From the perspective of that project or relationship it might be inappropriate for moral concerns to be considered.

Secondly, that there is good utilitarian reason for the agent to sometimes not consider morality when acting – doing so could impede other aspects of the act which are conducive to overall well-being, such as the joy of someone acting or being acted upon for no reason but love. Now we are back into the psychological realm: discussing not what reasons agents have, but how they ought to deliberate.

6 The Coherence Problem

The psychological problem was this:

1. Having a project as a commitment is incompatible with regarding that project impartially or as dispensable.
2. If one accepts utilitarianism, one regards all projects impartially and as dispensable.
3. By 1 and 2, those who accept utilitarianism cannot have commitments.

One response to this problem is to argue that utilitarianism is self-effacing; it does not require acceptance of itself. But if this is false:

4. Utilitarianism requires individuals to accept utilitarianism.

At points, Williams suggests a further problem for utilitarianism. This is that commitments are so important for well-being that:

5. Utilitarianism requires agents to have commitments.

If this were the case, then, given 3 and 4, utilitarianism would be incoherent. It would require agents to both have and not have commitments. Call this the 'coherence problem'.

---

[7] Consider Marx's identification of the marketplace of classical economics with 'Bentham, because each looks only to his own advantage. The only force bringing them together, and putting them into relation with each other, is the selfishness, the gain and the private interest of each.' (1990, 280).

How could 5 be supported? In one section of his *Critique*, Williams suggests that utilitarianism would be nonsensical if people did not have projects, of which commitments are a subset. Rejecting hedonism, he takes it 'that in talking of happiness or utility one is talking about people's desires and preferences and their getting what they want or prefer, rather than about some sensation of pleasure or happiness.' (1973, 80) For utilitarianism to be meaningful, therefore, there must be preferences to satisfy. Williams assumes that this requires people to have projects from which those preferences arise. And this must include some projects that are not the utilitarian project itself, since conceived as a project of maximising preference-satisfaction, it is 'vacuous' unless there are 'other more basic or lower-order projects.' (1973, 110) Commitments are one class of those projects.

This argument fails to establish that utilitarianism requires us to have commitments. For one thing, the 'requirement' in question is more like a presupposition than a moral prescription. If utilitarianism would be vacuous without commitments, then the defender of utilitarianism might be glad that some people have them, for this makes her theory meaningful, but it does not follow that utilitarianism says that agents *should* adopt commitments. For another, even if we grant that utilitarianism requires us to have preferences, it does not follow that utilitarianism requires us to have commitments. Imagine (with Parfit 2016, 118) a world whose inhabitants had only the drabbest of things in their lives—muzak and potatoes. Suppose that they only have two preferences: that there should be muzak rather than silence, and potatoes rather than gruel. It is not that they have any deep affection for muzak or potatoes: they desire and enjoy them no more than we do. They simply lack the means to form preferences for anything else. They do not have commitments, in Williams's sense. Yet it is obvious what they prefer and therefore what utilitarianism recommends for this world: more muzak and more potatoes. Lastly, Williams's argument here, would only apply to preference-satisfaction utilitarianism (or closely related alternatives), which is problematic on other grounds.

Williams makes a better argument for 5. He proposes, as an empirical hypothesis, 'that many of those with commitments, who have really identified themselves with objects outside themselves, who are thoroughly involved with other persons, or institutions, or activities or causes, are actually happier than those whose projects and wants are not like that.' (1973, 113–14) Whether this hypothesis is true or not is an interesting and important question, and the answer is not obvious, although the prevailing view is that it is (see Calhoun 2009 for dissent). Williams's hypothesis does not depend on any particular account of well-being, but some accounts of well-being—for instance, Joseph Raz's—make it closer to tautological than empirical. Raz writes that that our typical notion of a good life.

'is of a life well spent, of a life of achievement, of handicaps overcome, talents wisely used, of good judgment in the conduct of one's affairs, of warm and trusting relations with family and friends, stormy and enthusiastic involvement with other people, many hours spent having fun in good company, and so on.' (1986, 306).

Trusting familial relationships and friendships paradigmatically involve commitments, and achieving things, overcoming adversity and wisely using one's talents may also do so. A life without these things might include good company and sound judgment (as well as sensory pleasures), but we might resist calling a truly happy life—or at least think that it would have been better with respect to well-being had it involved commitments. This tells in favour of Williams's hypothesis. On the other hand, some of the worst lives tend to involve commitments as well: loving marriages and friendships break up, dreams are unfulfilled, martyrs are made in defence of lost causes. It is likely that those with commitments that are fulfilled have happier lives than those without commitments, but this group is sadly only a subset of those with commitments. Nevertheless, I will grant that it seems likely there are some kinds of commitment such that for most of us, if we adopted them, our lives would be happier.

However, this still does not entail 5. Utilitarianism does not require agents to make their lives go as well as possible. It requires agents to maximise overall well-being—that is, to make the lives of everyone, taken as an equally weighted sum, go as well as possible. And it could be that whilst it would reduce *my* well-being for me to reject commitment, doing so would free me to do things for others that would maximise well-being *overall*.

At this point one might object that if we all rejected commitment, there would be no commitment in the world, and this would reduce overall well-being (to zero, perhaps, if something like Raz's account of well-being is true). But as far as utilitarianism is concerned, what matters is not what would happen if we all did something, but the marginal effect of my doing so. And these things can come apart. Consider this case from Feldman (who used it to make a different point):

'Suppose a group of adults has taken a group of children out to do some ice skating. The adults have assured the children and their parents that, in case of accident, they will do everything in their power to protect the children. Each adult in the party is a good skater and swimmer. Suppose, finally, that, while 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 that you are one of the adults and you know that no other adult will go to save the child. Utilitarianism then directs you to save the child: this would result in the best outcome your actions could produce. This is true even though if every adult did the same, a disaster would occur.

Rejecting commitment could be like going to save the child. That is, it could be that both (a) if we all rejected commitment, this would be suboptimal, and (b) that the expected effects of my rejecting commitment, given the likely behaviour of others, would be optimal. In such cases, utilitarianism would direct me to reject commitment. So even if commitment is necessary for well-being, this does not mean that utilitarianism requires commitment from agents.[8]

7 Conclusion

I have explored Williams's Integrity Objection, and the diverse range of problems it is alleged to pose for utilitarianism. I rejected two interpretations of the objection: that it offers counterexamples in which utilitarianism's recommendations are at odds with moral intuition, and that it aims to show that utilitarianism is too demanding by making agents abandon commitments. Williams's objection has to do with the way utilitarianism treats the relationship between agents and their commitments. The most significant problem that the objection poses is what I called 'the psychological problem', that agents who accept utilitarianism cannot have the attitudes towards their projects that they must if they are to be properly committed to them. There are other problems in the neighbourhood: that utilitarianism does not acknowledge that commitments can change the reasons agents have, and that it incoherently asks agents to both hold and not hold commitments. I briefly showed how utilitarians might respond to these three problems.

Responding to the psychological problem is more difficult. I have begun to do so in (Venkatesh 2022). My response suggests that utilitarians 'bite the bullet', and stand by the conclusion that we ought not to be committed to our projects, at least in Williams's sense of commitment. Drawing on the social nature of individuals—in particular, the ways in which our projects are interdependent with those of others, I suggest that refusing to regard commitments as dispensable for the sake of others, given this mutual interdependence, risks proper engagement with both our own projects and our social relations. Though Williams may be right that utilitarianism precludes commitment, then, commitment might be less attractive than he took it to be.

Declarations

References

Ashford E (2000) Utilitarianism, integrity, and partiality. J Philos 97(8):421–439. https://doi.org/10.5840/jphil200097834

Calhoun C (2009) What good is commitment? Ethics 119(4):813–841. https://doi.org/10.1086/605564

Feldman F (1980) The principle of moral harmony. J Philos 77(3):166–179. https://doi.org/10.2307/2025668

Frankfurt HG (1988) Rationality and the unthinkable. The importance of what we care about: philosophical essays. Cambridge University Press, pp 177–190. https://doi.org/10.1017/CBO9780511818172.014.

Harcourt E (1998) Integrity, practical deliberation and utilitarianism. Philos Q 48(191):189–198. https://doi.org/10.1111/1467-9213.00091

Hare RM (1981) Moral thinking: its levels, method, and point. Oxford University Press, Oxford

Harsanyi JC (1977) Morality and the theory of rational behavior. Soc Res 44(4):623–656

Marx K (1990) Capital volume I (1867). Penguin Classics, London

---

[8] Wolf expresses a similar thought: 'if the utilitarian wants to influence more people to achieve more good, then he would do better to encourage them to pursue happiness-producing goals that are more attractive and more within a normal person's reach. These considerations still leave open, however, the question of what kind of an ideal the committed utilitarian should privately aspire to himself. Utilitarianism requires him to want to achieve the greatest general happiness, and this would seem to commit him to the ideal of the moral saint.' (1982, 427–28).

Mulgan T (2001) The demands of Consequentialism. Clarendon, Oxford

Nagel T (1971) The absurd. J Philos 68(20):716–727. https://doi.org/10.2307/2024942

Owens D (2012) Shaping the normative landscape. Oxford University Press, Oxford

Parfit D (2016) Can we avoid the repugnant conclusion? Theoria 82(2):110–127. https://doi.org/10.1111/theo.12097

Pettit P (2015) The robust demands of the good: ethics with attachment, virtue, and respect. Oxford University Press, Oxford

Railton P (1984) Alienation, consequentialism, and the demands of morality. Philos Public Affairs 13(2):134–171

Raz J (1975) Practical reason and norms. Hutchinson, Paris

Raz J (1986) The morality of freedom. Oxford University Press, Oxford

Raz J (1999) Postscript to the second edition: rethinking exclusionary reasons. practical reason and norms. Oxford University Press, Oxford

Scanlon TM (1998) What we owe to each other. Belknap Press of Harvard University Press, Cambridge, Mass. ; London

Sinclair U (1994) I, candidate for governor: and how i got licked (1934). University of California Press, California

Tanyi A (2015) Moral demands and ethical theory the case of consequentialism. In: Dainton B, Robinson H (eds) Bloomsbury companion to analytic philosophy. Bloomsbury Academic, London

Venkatesh N (2022) Against Commitment. Philos Stud 179(12):3511–3534

Williams B (1973) A critique of Utilitarianism. In: Smart JJC, Williams B (eds) Utilitarianism: for and against. Cambridge University Press, Cambridge, pp 77–150

Williams B (1981a) Persons, character, and morality. In: Rachels J (ed) Moral luck: philosophical papers 1973–1980. Cambridge University Press, Cambridge

Williams B (1981b) Practical necessity. moral luck: philosophical papers 1973–1980. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139165860

Williams B (1985) Ethics and the limits of philosophy. Harvard University Press, Cambridge

Williams B (1992) Moral incapacity. Proc Aristot Soc 93:59–70

Williams B (1995) Replies. In: Altham JEJ, Harrison R (eds) World, mind, and ethics: essays on the ethical philosophy of bernard williams. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511621086.011

Wolf S (1982) Moral Saints. J Philos 79(8):419–439. https://doi.org/10.2307/2026228

Wolf S (2012) One thought too many: love, morality, and the ordering of commitment. In: Heuer U, Lang G (eds) Luck, value, and commitment: themes from the ethics of Bernard Williams. Oxford University Press, New York