Dear Rohan,

Below you will find some comments I made on your essay, some of which we discussed in the tutorial. I'll mention the main substantive points first, and you'll find minor comments in the margins. I would like to emphasise that I was quite impressed by the quality of the essays and the discussions overall. The point of my comments is not to tell you that you have made mistakes or that your essay is bad – I send every student roughly the same number of comments, all of them critical. They are meant to help you think further about integrity and alienation and to write essays that are even better than they already are. I have limited time to write these comments, and I'm sorry if there are typos or errors. Please read the comments in a critical spirit and check them against this week's readings. It was a pleasure to teach you; I wish you the very best for your studies!

All the best,
Jonas

- There are different ways in which one might interpret Williams's "integrity objection". It is important to engage charitably with the literature. You might be attacking a strawman. You write:
  - according to Williams, "an agent must be permitted to make decisions on the basis of their own projects and commitments, and not morally obligated to take a particular action as a result of other individuals' choices (p116-7)"
  - Note that it would be an extreme view to hold that even for minor commitments one could not be obligated to give them up. (Suppose A could give up his commitment to playing tennis in order to save millions of lives.)
- You also seem to imply later that the integrity objection is an "extensional" one: utilitarianism does not get the extension of action-guiding verdicts right.
- Roger Crisp (1997) would argue that this is not the point of Williams's objection:
  - "First, [the examples of Jim and George] are not meant to be 'counterexamples' to utilitarianism. […]. It is not in the end what utilitarians say that worries him, but how they reach their conclusions: 'The first question for philosophy is not "do you agree with utilitarianism's answer?" but "do you really accept utilitarianism's way of looking at the question?"' (Williams 1973b: 78).
- What else might Williams be concerned with if not the extension of action-guiding verdicts?
  - According to one reading of Williams's objection, the problem is that
    (1) utilitarianism is incompatible with regarding one's projects as more than "just one satisfaction among others" and as not "dispensable" (see Williams 1973, p. 116). (What does "not dispensable" mean? Imagine a pianist who says at the dinner table, "Playing music is not dispensable to me!" Under normal circumstances, we would not assume that she is suggesting that she would not give up music even if she could thereby prevent a war. She does <u>not</u> mean that there is no conceivable situation in which she would give up this project).
  - The multilevel utilitarian might reply to (1): utilitarian encourages people to adopt the rule of thumb to regard their own projects <u>as if</u> they were not just some satisfactions among others – although they actually are.
  - What would you say if someone wanted to defend Williams as follows? (a) To regard, for example, one's project of painting <u>as if</u> it were more than just one satisfaction among others it not the same as to regard it <u>as</u> more than just satisfaction among others. (b) Even if giving up one's project would lead to greater utility, it is appropriate to regard, for example, one's project of painting as more than just one satisfaction among others.
  - Relatedly, Crisp (1997) argues that
    (2) utilitarianism fails to properly account for certain agent-relative reasons.
    - "There are reasons, revealed to us in our emotional reactions to imaginary cases and to circumstances in the lives we live, that run counter to impartial maximization. This is what is true in the integrity objection." "We might call these 'agent-relative' reasons, since they make essential reference in their formulation to

the agent who has them (see Nagel 1986:164–75).” “This is what is true in the integrity objection.” (Agent-neutral consequentialism – whether sophisticated or not – cannot account for these reasons.)

- Perhaps the multilevel utilitarian would reply:
    - You have the *intuition* that this is the case! However, your intuition is the result of an inculcated decision procedure. It does not reveal what you think it reveals.” (Similarly, she might reply to Crisp: “It seems to you that there are these special reasons, but in fact there are no such reasons.”)
    - It’s worth thinking about how forceful this rejoinder is. It is worth thinking about not only whether this is a plausible reply, but how plausible it is compared to Williams’s/Crisp’s view. If both are, say, equally plausible, it would not be a knock-down argument. But it might still make utilitarianism significantly less compelling.

What is ‘integrity’? Does utilitarianism threaten it? Is that a bad thing?

1. No, utilitarianism does not threaten integrity, though careless definitions of either term might lead one to mistakenly conclude that it does. In this essay, I will show that integrity is entirely consistent with correctly practised act-utilitarian moral frameworks when it is understood as it ought to be: the quality of acting in line with one’s deeply-held beliefs and values. As part of this, I demonstrate that purported alternative notions of integrity which give an absolute priority to autonomy and independence are incompatible with any kind of social morality and should be set aside entirely. Finally, I argue there is a false assumption behind the titular question – that one can non-circularly assess the truth of moral propositions – and that this leads to more fundamental difficulties with attempts to evaluate ethical theories.

2. Before arriving at my definition of integrity, we will first examine another view, to bring out the contrast between the two and highlight why the alternative is a flawed characterisation. Williams argues that integrity is about autonomy: an agent must be permitted to make decisions on the basis of their own projects and commitments, and not morally obligated to take a particular action as a result of other individuals’ choices (p116-7). On this view, integrity of the self is tied up with personal identity. Each agent is uniquely themself precisely because they have their own projects and commitments which are not to be externally interfered with. This concern is also seen in the importance Rawls places on the separateness of persons, arguing that some facets of identity are so important that we cannot ask one individual to sacrifice them for the greater good (p24). However, this proves too much: if agents must always be permitted to remain committed to their own ground projects, then all notions of universal moral obligations disappear completely. For any such posited duty, one could always come up with an agent whose ground projects conflict with it, thereby (according to Williams) rendering the duty invalid. An insistence on this kind of integrity must therefore wrongly restrict us only to relativist or egoistic forms of morality. Defining integrity as the quality of generally acting in line with one’s deeply-held beliefs and values allows us to avoid this conclusion. There is no reason to insist that agents must always act in accordance with their values in order to possess identity, and integrity therefore does not have a veto over moral obligations.

3. Williams might reply that utilitarianism is incompatible even with this modified conception of integrity, because its joint doctrine of impartiality and negative responsibility requires so much of agents that they would constantly be obligated to perform welfare-maximising acts and would never have the opportunity to fulfill their deeply-held beliefs and values. But the most deeply-held belief for a utilitarian is that they have a moral obligation to impartially maximise utility. Acting towards that goal is therefore clearly not a challenge to their integrity. This analysis helps us interpret our intuitions in response to Williams’s famous thought experiments more clearly. Take the case of George, a pacifist chemist who cannot decide whether to take up a job overseeing weapons production in order that he could slow the process down. If George were a utilitarian then his pacifism would be merely instrumental, and he could accept the post without threatening his integrity (indeed, not accepting the post would run counter to it). If George were not a utilitarian, then of course the utilitarian course of action might clash with his beliefs and values – but on its own, this tells us nothing about whether or not George *ought* to be a utilitarian, let alone if utilitarianism is correct.

Commented [A1]: Be careful with labelling your opponents definitions „careless“!

Commented [A2]: Style: can *frameworks* be practiced?

Commented [A3]: The following seems non-circular.

What about:
A presents the proposition: "Every action is permissible."
B assesses it as wrong based on the judgment that "It is impermissible to torture a baby, all else being equal."

Maybe you have certain kind of circularity in mind here? Is circularity the right term for your worry?

Commented [A4]: Make sure it is clear what „this“ refers to, it wasn’t clear to me here.

Commented [A5]: why is that? How exactly does does relativism enter in here?

Commented [A6]: …without acting wrongly?

4. Perhaps George and others like him would be catastrophically disadvantaged in some way if they were all utilitarian, showing that the moral theory cannot be universally correct. As part of his argument that utilitarianism is "absurd" as a sole fundamental value (p116), Williams claims that a world comprising only of utilitarians would be devoid of value even on a utilitarian account, as there would be no way to endogenously induce welfare amongst agents whose individual wellbeing is determined only by the total utility.

5. However, if any possible worlds contain moral value, then those inhabited solely by perfect utilitarians must contain value – otherwise welfare would not be being maximised. More specifically, if an absence

of non-utilitarian ground-level projects really would lead to less happiness or, as Williams outlandishly claims, no agents at all (p110-112), our utilitarian actors would simply adopt some such projects to avoid that outcome. Railton (p143) terms this "sophisticated" hedonism, but it in reality it is simply non-naïve hedonism, and directly analogous to how actual utilitarians adopt rules and heuristics where useful, through multilevel consequentialism (Crisp p143).

6. We can generalise this response to deal with other extensional criticisms. A utilitarian would argue that, by definition, nothing utilitarianism recommends (or more precisely, no action performed by a rational perfectly utilitarian agent) can be "a bad thing":

> **P1** A moral theory *T* says that an agent should take action *A* (e.g. having only utilitarianism as a deeply-held value) in a context *C*.
>
> **P2** It would have been bad in a utilitarian sense for the agent to have taken this action *A* given *C*.
>
> **P2*** There existed another action *A'* (e.g. having additional deeply-held values) available in *C* which, if taken, would have produced more aggregate welfare.
>
> **C** T cannot be an act-utilitarian theory, because it does recommend agents take actions in line with the act-utilitarian criterion of rightness.

7. The above argument demonstrates that utilitarianism can, and will, absorb any other theory if doing so would increase welfare. However, this should not be conflated with an ability of utilitarianism to accommodate multiple sources of ultimate value, as Railton does (p148-50). Valuing something only in virtue of its producing another desirable object is the very definition of instrumentalism (Crisp p144): if the additional deeply-held values did not in actuality lead to greater welfare, then the utilitarian would abandon them. This exposes a broader problem in ethics. By what lights can we judge some effect to be a bad thing, except the lights that we trying to show are the correct, undistorting ones that ought to function as our criterion of rightness? No adherent of any moral theory would accept that their framework omits what matters or promotes what is wrong, as otherwise they would subscribe to another. It is perfectly possible for several internally consistent theories to exist, and for us to have no way of judging between them besides our subjective intuitions.

8. To conclude, integrity is the quality of acting in line with one's deeply-held beliefs and values, and I do not believe it is threatened by utilitarianism. As I have shown above, we should reject absolutist definitions of integrity as identity, as they are not compatible with any commonsense notions of morality. There is ample space for integrity as I define it within act-utilitarianism – and indeed, that theory would actively embrace it if necessary to achieve utilitarian goals. However, the fact that it remains an open question whether threatening integrity is a bad thing or not illustrates perfectly a more fundamental difficulty in ethics: our inability to agree on axiomatic propositions upon which moral theories can build.

Commented [A7]: This representation of Williams is unclear. Can you think of another way to read this passage:

"let us grant to utilitarianism that all worthwhile human projects must conduce, one way or another, to happiness. The point is that even if that is true, it does not follow, nor could it possibly be true, that those projects are themselves project of pursuing happiness. One has to believe in, or at least want or quite minimally, be content with, other things, for her to be anywhere that happiness can come from." (Williams 1973)

Commented [A8]: Any world? Would a possible world which contained only stones also contain moral value?

Commented [A9]: Similarly to your use of „careless" above: Before calling Williams's view outlandish, you might want to explain how he understands this claim, and make sure to give it a charitable interpretation.

Commented [A10]: A citation convention that seems to be more widely used in philosophy has the form (Crisp 1997, 143).

Commented [A11]: is „having" values an action? Do you mean „inculcating", or „acquiring", …., values?

Commented [A12]: Here is a parallel claim:

"No person can accept that her belief that p is false. So convincing people that they are wrong to believe p is impossible."

Why? „Once you convince them, they don't believe p anymore."

The „anymore" is crucial. They once believed it, in similar adherent of moral theories can accept that their past frameworks were wrong. Quite a few philosophers have accepted that.

Commented [A13]: How do we choose between different internally consistent theories in science?

Commented [A14]: You did not mention before that you were concerned with „axiomatic" propositions. It migth help to give an example for an axiomatic proposition in ethics, and introduce this term earlier, if you need it.

References

Roger Crisp, "Routledge Philosophy Guidebook to Mill on Utilitarianism" (Routledge, 1997): 95-124.

Peter Railton, "Alienation, Consequentialism, and the Demands of Morality", *Philosophy & Public Affairs*, vol. 13, no. 2 (1984): 134-171.

John Rawls, "A Theory of Justice" (Harvard, 1971 [1999 ed.]).

Bernard Williams, "A Critique of Utilitarianism" in Smart & Williams, "Utilitarianism: For and Against" (Cambridge, 1993).