Doing Our "Best"?
Utilitarianism, Rationality and the Altruist's Dilemma

Abstract

Utilitarians think that what matters in ethics is making the world a better place. In that case, it might seem that we each rationally ought to *do our best* – perform the actions, out of those open to each of us, with the best expected outcomes. In other words, we should follow *Act-Utilitarian* reasons. But often the result of many altruistic agents following such *individualistic* reasons is *worse* than the result of them following *collectivist* "team-reasons". So Utilitarians should reject Act-Utilitarianism, and accept a Dualist view according to which both individualistic- and team-reasons are fundamental. In order to align these distinct kinds of reason, Utilitarians must focus centrally on questions of political and social reform – as did their historical forebears.

1. Introduction

If what really matters in ethics is making world a better place, then it seems natural to suppose that we each ought to *do our best* – that we each ought to perform the actions, out of those open to us, with the best expected outcomes. In other words, it might seem that we should rationally become *Act-Utilitarians*. But whereas the notion of goodness is absolute, the notion of a *best* option is relative – it is defined relative to a set of options. And determining what count as the options open to a rational agent is not straightforward. After all, what *my* options – as a rational, moral agent – are, depends, in many cases, on what *other people* are going to do. But many of *those* people are *also* rational moral agents. So what *they* think they morally, rationally ought to do will, in part, determine what they *will* do, and so constrains what options *I* have. And, one hopes, in many cases what they *think* they have most reason to do will reflect what they *in fact* have most reason to do.

Thus rational moral guidance cannot just be a matter of *reacting* to the world as we find it, as Act-Utilitarianism supposes – for moral reasons partly *shape* the world to which we are reacting. And there are ways that moral agents could collectively construe their options which diverge from the Act-Utilitarian framing, but which make *better* outcomes available than if they had been strict Act-Utilitarians. So, even if our only premises are fundamentally Utilitarian ones, we should reject Act-Utilitarianism as a complete account of moral reasons.

The case I'm going to focus on concerns political action. I pick this for a reason. Although we now tend to regard Utilitarianism primarily as a view in moral philosophy, this presents a narrow view of Utilitarianism in the history of practical thought. For while the early Utilitarians did articulate versions – criticised by later philosophers for their seeming ambiguity – of the "Principle of Utility", they were not primarily concerned with personal morality. John Stuart Mill, Harriet Taylor, James Mill, William Godwin, even – despite his association with "Act-Utilitarianism" – Jeremy Bentham, were primarily interested in questions of social, legal and political reform. It was to such structural questions that they devoted most of their efforts and writings, seeking radical systemic reforms of society.

By contrast, a tradition running from Henry Sidgwick, through the work of contemporary philosophers such as Derek Parfit and Peter Singer, to the Effective Altruism movement of today, sees Utilitarianism as, primarily, a moral theory of how individuals ought to behave. In this tradition, political questions naturally become questions of applied *personal* morality, questions about the best way for individual Utilitarians to proceed in engaging with the political realm. There has long been a quandary about how political principles (such as those defending liberty) are supposed to relate to the "Principle of Utility" in classical Utilitarian thought. But the tension between Act-Utilitarianism, as a personal morality, and Utilitarianism as an approach to radical political reform, has been made especially clear by the contemporary Effective Altruist movement. For, as Effective Altruists have argued, it is just not clear at all that a rational Utilitarian altruist should concern herself with the pursuit of political reform. Given the reliance of political change on the cooperation of many others, moral rationality should guide the altruist to focus on individual charitable giving, rather than gambling her efforts on the chance of changing society.

This surprising thought is my starting point. I want to suggest that history has been too harsh on the early Utilitarians for their failure to articulate a precise account of the Act-Utilitarian account of reasons (although my argument for this vindication involves the rejection of a certain sort of methodological individualism that the early Utilitarians largely presupposed). For the understanding of Utilitarian moral reasons that I want to advance is one in which moral reasons just *are* ambiguous, and in which questions of personal morality cannot – even at a theoretical level – be decoupled from issues of political and social reform.

But first, I want to tell you a story.

2. The Altruist's Dilemma

Henry wanted, more than anything else, to make the world a better place. He aimed to use rationality and evidence to identify the most effective use of his time, energy, and considerable skills. Not for him a sentimental attachment to lost causes or symbolic protests. Henry, in his love of mankind, wanted to do what was best.

On leaving University, Henry's admirable commitments led him, like many of his most serious and intelligent peers, to seek guidance from the leading lights of *Effective Altruism*. But there he found advice that surprised him. He had always assumed that the forum for those who wanted to change the world for the better was the political realm – perhaps not electoral politics itself, but, at least, fighting for the structural reforms that society desperately needed. Of course, political and economic questions are contentious, and Henry was wise to this; nevertheless, there were reforms – concerning climate change, access to healthcare and education, and foreign aid – which seemed, on careful analysis of the best evidence he could find, clearly such as to benefit people, were they to be enacted.

But there, he discovered, was the catch. The organisation to which he had turned for guidance, *Doing the Best We Can* (DBWC), agreed with him that these reforms *would* be exceptionally valuable, *if* they could be brought about. But, they pointed out, achieving such reform would take far more than the efforts of any one Henry, no matter how intelligent and committed he should be. It would take the cooperation of many people for any of the campaigns Henry dreamt of to have any chance of succeeding. And the evidence suggested that this was unlikely. Given this, Henry concluded, regretfully, that seeking political change was not, after all, his best option in his search to change the world for the better. DBWC advised him that he would do better to seek a conventional job, make plenty of money, and donate it to highly effective charities. This seemed to Henry to be the only reasonable conclusion.

But then something struck him. After all, *a great many of Henry's peers from university had also turned to DBWC for guidance*. These were morally motivated, intelligent people just like Henry, who would have pursued political reform if they had thought that this was recommended by rational morality. In convincing them to eschew politics in favour of lucrative employment and charitable giving, DBWC had *reduced* the number of potential co-operators Henry might have found in his political ventures, and thus made it *even more unlikely* that any such political campaigns might succeed. After all, unlike some avenues for improving the world that offered a low probability of a great reward – things like nuclear fusion research – there was nothing *intrinsic* to political change that made it improbable. Its probability of success was almost entirely a function of the willingness of agents to pursue it.

Thus DBWC's pessimism about political change, which once struck Henry as impeccably rational, now appeared as a dangerously self-fulfilling prophecy. DBWC had encouraged Henry and his peers to think rationally about how best to use their resources to change the world for the better. But the effect of their following this advice, he worried, was to change the world for the worse. On the other hand, if Henry and his peers had pursued politics, then, given their numbers and talents, the probability of effecting change would have been quite high. Given the immense value of such change, this would, Henry judged, have been a better result.

At first, Henry suspected that the problem lay in the role of DBWC. Even if the best option for Henry and each of his peers was to follow the pessimistic path to apolitical altruism, perhaps, he reasoned, DBWC should have thought more carefully about its *own* causal effects. Knowing that it might influence a great number of highly motivated and intelligent people should have informed the advice that DBWC decided to give, and so perhaps it really ought to have told Henry and his peers to pursue politics, even if that was not really the thing that any of them individually had most reason to do. In a sense, DBWC should have hidden the truth about moral reasons from its audience in service of the greater good.

But this idea, Henry concluded, was a red herring. After all, *the only thing DBWC had done was to tell each of its followers what they had most reason to do*; it had no power beyond the power of rational advice! And neither Henry nor his peers would have followed that advice had they not *agreed* that it was rational. Any of them could have reached the same conclusions on their own. And if it would have been best for DBWC to try to *get* each of them to pursue politics despite the poor individual odds of success, this was also something that they might have worked out by themselves – as Henry had done. But how could it be that the reasons that it would be best for all of them to follow were not reasons for each of them to do what was best?

On the one hand, it did seem rational to think as DBWC had recommended, and eschew politics in favour of the more reliable option of charitable giving. But if it was rational for Henry to act that way, then it was equally rational for others to do the same, which would make political reform even more unlikely.

On the other hand, reflecting on this fact made it look rational to think in a more collectivist manner, and pursue political reform despite the individually poor odds of success. But that would only be a good thing if other people *did* in fact follow this reasoning – and Henry had no way of knowing that they would do this. It would be best for any of them to pursue politics if and only if *the others* were also going to pursue politics, and it would be best to *treat this fact as a reason* to pursue politics if and only if *others* were also going to treat this fact as a reason.

Henry's dilemma ran deep. It wasn't just that he wasn't sure which option to pick in this situation. He no longer felt sure what it meant to be rationally altruistic in the first place. He wanted to do his best. But what *was* his best option?

3. And You May Ask Yourself – How Did I Get Here?

Henry arrives at Act-Utilitarianism because he thinks it is the best expression of his deeper commitments. He thinks that what matters most, in moral terms, is how good, or happy, the world is. And he thinks that the moral reasons that apply to him and other moral agents derive from the ultimate moral goal of promoting the good. It is the apparent clash between the desire to see the world become a better place and Act-Utilitarianism that generates Henry's dilemma.

Here are Henry's basic principles:

> *Axiological Utilitarianism:* What matters most, morally, is welfare; welfare is what is good, and the more welfare people experience, the better.
>
> *Meta-Deontic Utilitarianism:* There are moral reasons which apply to agents, and the contents of these reasons is fixed by facts about promoting welfare.

Meta-Deontic Utilitarianism claims that, *whatever* reasons we have, these are *somehow* derivative of the goal of promoting welfare. But that "somehow" is hard to interpret. Here is an interpretation:

> *Optimific Reasons:* Morality assigns to agents the reasons that it is best for them to follow; the optimific moral reasons are the reasons that will guide agents in such a way that leads to the best outcomes.

Indeed, even those who aren't full-blown Utilitarians might agree that these are plausible principles *for cases like this* – where we are trying to make the world a better place. After all, "side-constraints" regarding rights and principles of justice arguably won't settle whether we should pursue political reform or charitable giving. So even deontologists who allow some role for utilitarian-style thought in their moral worldviews should feel the pinch of Henry's dilemma.

Why should optimific reasons be Act-Utilitarian? The early Utilitarians often formulated the "Principle of Utility", in ways that seem ambiguous between Act-Utilitarianism and "Indirect" Utilitarian views, such as Rule-Utilitarianism:

> *Act-Utilitarianism*: Each agent has reasons to choose the act, out of those open to her considered individually, that has the highest (expected) utility.
>
> *Indirect-Utilitarianism*: Each agent has reasons to perform acts selected by rules/principles/virtues/motives that (would) lead to the best (expected) outcomes if (all/most) agents (obeyed them/tried to obey them/promulgated them).

Obviously, this construal of Indirect-Utilitarianism is ambiguous. But that doesn't matter for our purposes. Because Act-Utilitarians (following Lyons 1965) can pose a dilemma to any form of Indirect-Utilitarianism. If an alternative to Act-Utilitarianism gives the same advice, then the difference is chimerical. If they diverge, they appeal to the following principle:

> *No Rule Worship*: If any account of reasons for action tells us to choose the options with (predictably) worse outcomes than some alternative account of reasons for action, then we should reject this account of reasons for action.

If we act in a way that leads to predictable net harm to others, *just because* some rules, principles or standards of virtue seem to recommend this, then it seems that we have put these rules above the goals that they were supposed to serve. And that seems incompatible with *Optimific Reasons*.[1]

But this assumes that Act-Utilitarian reasons *are* the *Optimific Reasons*. Call this the:

> *Equivalence Thesis:* The reasons that are best for agents to follow are, uniquely, reasons for each agent to act upon the option with the best (expected) consequences, out of those open to her. The *Optimific Reasons* just are *Act-Utilitarian Reasons*.

Henry's dilemma should lead us to question the *Equivalence Thesis*. DBWC thought that political reform was unlikely to succeed. If that's right, and if Henry and his peers followed Act-Utilitarian reasons, they would each eschew politics in favour of individual charitable giving.

[1] Some Rule-Utilitarians, such as Hooker (eg Hooker 2000), argue that their theory better coheres with our moral intuitions or common sense; but, to that extent, their motivation is *not* the sparse one articulated so far. Though such theories may make reference to the promotion of utility in determining which rules or norms they accept, their ultimate motivation is *not* just to find whatever account of moral reasons is optimific. Rather, the desire to respect common sense serves as an independent normative goal. Thus, these theories are not truly *Meta-Deontically* Utilitarian – they do not fully subordinate reasons to the ultimate goal of promoting the good.

But had they been guided by different principles, perhaps they would have pursued politics instead. And this would have been a better result.

We might, then, be tempted by the following two thoughts:

1) In *The Altruist's Dilemma*, the Optimific Reasons are reasons to pursue politics.
2) In *The Altruist's Dilemma*, Act-Utilitarian Reasons are not reasons to pursue politics.

If we grant 1) and 2), then the *Equivalence Thesis* seems false, so the argument from the *Equivalence Thesis* to Act-Utilitarianism fails.

4. Objective Reasons and Hidden Principles

Act-Utilitarians may protest. After all, as Railton argued, Utilitarianism is a theory of *objective reasons* – a claim about how people should act – not a theory of *subjective reasons* – a claim about how people ought to deliberate. Indeed some, like Singer and Lazari-Radek (2014), following Sidgwick, accept that Act-Utilitarianism might be an *Esoteric Morality* – people might act better if they did not even believe in Act-Utilitarianism. So Act-Utilitarians needn't try to persuade others to *become* Act-Utilitarians – they should merely give other people whatever moral advice would likely bring about the best outcomes.

So Act-Utilitarians might argue that DBWC should have *told* its audience to pursue politics:

i) The expectedly-best outcome would be brought about if all of DBWC's audience were to pursue political reform.
ii) If DBWC were to tell its audience that they each had most reason to pursue political reform, then they would indeed each pursue political reform.
iii) The option open to [whoever is in charge of] DBWC with the highest expected utility is to *tell DBWC's audience to pursue political reform*.

It is not morally important that agents subjectively *deliberate* in Act-Utilitarian terms. So, if the best outcome is one where Henry's peers pursue politics, then Act-Utilitarian reasons recommend that *DBWC* should give them whatever advice will bring this about – even if that advice does not state the reasons that Act-Utilitarianism gives them!

This might make sense if DBWC's audience were blindly obedient. But, as I have told the story, ii) is false. As Henry observed, they would only follow the advice *if they could see that it was rational*. So DBWC cannot solve the dilemma just by *making* its followers do the best thing. More importantly, nothing in *The Altruist's Dilemma* rides on how Henry and his peers deliberate subjectively – what matters is what actions they choose. If 1) is true, then the correct *action* for them to choose is to pursue political reform, but if 2) is true, then Act-Utilitarianism does not attribute to them reasons to pursue political reform. It may be consistent with Act-Utilitarianism to say that we should hide the truth of Act-Utilitarianism from people in order to get them to do the actions that Act-Utilitarian reasons objectively recommend; but could it be consistent with Act-Utilitarianism to say that we should manipulate others in order to get them to do something *other* than what Act-Utilitarian reasons objectively recommend?

5. The Agent-Neutrality of Utilitarian Reasons

What has gone wrong here? Act-Utilitarian reasons are agent-neutral. This seems to imply:

> *No Immoral Morality:* If the objectively morally correct thing is that I do P, it cannot be that I act objectively immorally in doing P.

And I think it follows from this that:

> *Reasons Transmission:* If the best thing possible is that A *brings it about that* B performs P, then (other things being equal[2]) B already has most objective moral reason to P.

So if the best thing is for DBWC to *get* its followers to pursue politics, then *they* have most reason to pursue politics. If we think of rationality in terms of instantiating particular patterns of decision-making, then familiar examples such as Parfit's (1984) *Robber* scenario show that there can be cases of *rational irrationality* – cases where it is rational for an agent to make herself think subjectively irrationally. But given the Utilitarian commitment to agent-neutrality, there cannot likewise be cases of *moral immorality* – if it is objectively moral for me to make myself or someone else act in a particular way, then it is objectively moral for them to act this way.

Here is an exception to the *Transmission* principle:

> *Buridanic Coordination*: When B and C need to coordinate, and both do *either* P *or* Q, such that [B and C both doing P] is just as good as [B and C both doing Q], but no good results if they each choose different options, then A's advice can make a difference to what B and C have most reason to do.

In a Buridanic case, B and C each have sufficient reason to do either P or Q, but no decisive reason to choose between the two. Likewise, A has sufficient reason to advise both B and C to do either P or Q, but no decisive reason to choose between the two. But they all have decisive

[2] There might be extraneous factors complicating this inference – for example, if there is some independent benefit caused by A's act of trying to influence B, or if some third party will punish B if she performs P without being compelled. These *ceteris paribus* riders are not relevant for the argument I'm making.

reason to bring it about that B and C coordinate. If they cannot do this on their own (for example, if they cannot communicate), then A can pick at random and break the tie for them. However, A's advice changes the normative landscape only by providing B and C with an extra piece of *information*: A marks out one of the options as a mutually obvious point of coordination – that is to say, A makes that option *salient*, or, as Schelling (1960) would call it, a *focal point*. Thus we might accept the following as a principle of moral rationality:

> *Rational Salience*: If an agent must perform either P or Q, and has equally strong reasons to perform both P and Q, she may rationally pick whichever is more salient.

If A tells B and C to do P, and both and B and C know that A has done this, then they should both do P, not because it is intrinsically a better option, but because a good outcome requires coordination, and they now each have evidence that the other will plump for P over Q, since P is now mutually salient: each knows that the other is rational, and that rational people will use salience to break Buridanic ties.

But this principle isn't relevant to *The Altruist's Dilemma*. If 2) is true, and if Act-Utilitarianism is true, then Henry and his peers *do not* have equally strong reasons to pursue both options. Rather, they each have strongest reason to pursue charitable giving. They cannot use salience to break the tie between the two, since there *is* no tie. So none of them can expect the others to pursue politics just because DBWC tells them to, and no-one has reason to pursue politics in response.

Conversely, if DBWC were to give its audience advice, it would not be a matter of *Buridanic Coordination*. DBWC shouldn't just toss a coin to decide which choice to direct its audience to. Rather, since political reform is objectively better than collective charitable giving, DBWC has *determinate* reason to advise its audience to pursue reform. But then, given *Reasons Transmission*, it must have been true that they already had determinate reason to pursue political reform.

So we can see why Henry's first thought about DBWC's role was incorrect. If there is a problem for Act-Utilitarianism, it cannot be solved by appeal to the causal, coordinating role of DBWC.

6. What Does Act-Utilitarianism really advise?

The Altruist's Dilemma is inspired by the "Institutional Critique", advanced by Srinavasan (2015) and others, which charges that Effective Altruists have, in prioritising charitable giving, wrongly ignored possibilities for valuable systemic change. Effective Altruists, like Berkey (2018), have responded that this is unfair. If seeking systemic change isn't the most effective way to make the world a better place, then they don't see why they should endorse it. But if it *is*, then surely it already follows from their principles that they should endorse it, and so the Institutional Critique is not a critique of Effective Altruism or its Act-Utilitarian underpinnings, but merely of the misapplication of their principles. In other words, either 1) or 2) must be false – either the Optimific Reasons *are not* reasons to pursue political reform, or Act-Utilitarian reasons, properly understood, recommend the pursuit of politics.

Maybe 2) is false. Perhaps DBWC was wrong, and Act-Utilitarian Reasons really do direct each of Henry and his peers to pursue politics. After all, political reform is the outcome with the greatest utility, if it can be brought about, and if enough of DBWC's audience choose to pursue it, then it *can* be brought about. But if they are orthodox Act-Utilitarians, we face the following puzzle:

i) If Act-Utilitarian reasons recommend that DBWC's audience should pursue politics, then politics has the highest expected utility.

ii) If Act-Utilitarian reasons recommend that DBWC's audience should not pursue politics, then politics does not have the highest expected utility.

iii) Act-Utilitarian reasons recommend that DBWC's audience should pursue politics if and only if it has the highest expected utility.

In other words, it seems like we need to know what Act-Utilitarian reasons recommend before we can work out what Act-Utilitarian reasons recommend!

To get clearer on this, we can turn to work on the *Hi-Lo Game*, described by game-theorists and economists such as Bacharach (2006), Gold & Sugden (2007), Sugden (2015) and Colman and Gold (2020), and theorists of Utilitarianism, including Gibbard (1965), Regan (1980) and Woodard (2017). In the Hi-Lo game, Agents A and B must select between Actions 1 and 2, with the following payoff matrix, with payoffs defined as arbitrary units of impartial utility:

HI-LO (moral)	A chooses 1	A chooses 2
B chooses 1	20 units of utility	0 units of utility
B chooses 2	0 units of utility	10 units of utility

The value of the options open to each agent depends on what the other agents choose. There are two "Nash Equilibria" – outcomes where no agent can improve the situation by changing what she does, given what the other does. But one Nash Equilibrium is better than the other. It is best if the agents converge upon this "Hi" outcome. But can each agent expect the other to do this? It's not clear they can. If each agent plays her part in bringing about the suboptimal Nash Equilibrium ("Lo"), they have each done the best act open to them individually. Since each agent knows that the other can see this, neither can rationally expect that the other agents *will* do their part in collectively enacting "Hi". And so, picking 1 is not determinately rational.

Of course, our agents would prefer the Hi result to the Lo result. But *this* outcome is not an option for either of them individually – it is not an outcome that either can bring about alone. And Act-Utilitarianism tells agents to choose only between options that are open to them individually. The only options available to each agent are to play 1 or 2, with the value of each dependent on what the other chooses. If all they each know about each other is that they are rational Act-Utilitarians, they cannot know what the other will choose.

You might think that Act-Utilitarians can solve this problem by applying the:

> *Principle of Indifference*: Where A has no idea what B is going to do, she can assign equal probabilities to each of B's options – as though B is going to choose at random.

In that case, A will assign a 50% chance to B choosing 1, and a 50% chance to B choosing 2. So now A can assign an expected utility score to each choice of hers – Option 1 has an expected utility of 10, and Option 2 has a score of 5. And so, as an Act-Utilitarian she should rationally choose Option 1. And by the same rationale, B will also choose 1.[3]

But this yields the opposite conclusion in *The Altruist's Dilemma*. The Hi outcome requires *a great many* of the relevant agents choose to pursue political reform; if that will not happen, then it is better for all of them to devote themselves to charity instead. If each agent treated the others as having a 50% chance of pursuing either politics or charity, then each would expect that not

[3] Colman and Gold (2020) and Bacharach (2006), argue that this argument is fallacious, because it involves a self-contradiction – it starts by assigning B a 50% chance of playing each of 1 and 2, and then concludes that B has a 100% chance of playing 1. However, I think this problem is properly philosophical, not logical. Contradiction only arises if A attributes symmetrical reasoning to B, and concludes that B will in fact choose 1. But A is not *forced* to perform that last step – she could just assume that B will act randomly and leave it there. Now, this would imply an asymmetry between A and B – as though A were a true reasoner, and B a mere force of nature, and one might object to this way of thinking. But this is a more general problem for Act-Utilitarianism, in its insistence that we try to predict the actions of others before deliberating for ourselves – I return to this in §10.

In any case, if we follow Colman, Gold and Bacharach in rejecting indifference reasoning here, we are then left with the conclusion that Act-Utilitarian reasons do not determinately recommend either option in such cases, and hence, *a fortiori*, that they do not determinately recommend pursuing politics in *The Altruist's Dilemma*.

enough others would pursue politics to make it worthwhile, and so would rationally conclude that the best thing would be to devote themselves to charity instead. If each agent applies the principle of indifference, then they will, in *The Altruist's Dilemma*, coordinate on the Lo outcome.

Thus, if all the relevant actors are rational Act-Utilitarians, DBWC's advice is sound: Act-Utilitarian reasons really do *not* recommend that participants in *The Altruist's Dilemma* pursue political reform. And if the Optimific Reasons are indeed reasons to pursue political reform, then the Optimific Reasons are *not* Act-Utilitarian Reasons – so the *Equivalence Thesis* is false.

7. Optimific Reasons and Team Reasons

But in real life, DBWC's audience are not likely to be idealised Act-Utilitarians. Perhaps this a good thing! Indeed, DBWC should *want* its audience to pursue politics, and not follow Act-Utilitarian reasons. But the *Transmission* principle implies that if it's best for DBWC to get its audience to pursue politics, then they have reasons to do so. If those reasons are not Act-Utilitarian, then what are they?

Game theorists like Bacharach solve the Hi-Lo puzzle by appealing to the theory of "Team Reasoning". This is a theory of subjective rational decision-making, and so is not directly applicable here. But we can translate this idea into the language of objective moral reasons:

> *Team Reasons:* In coordination problems where multiple rational, moral agents can coordinate to bring about an optimific outcome, each agent has reasons to play her part in bringing about the best outcome that the collective of agents can jointly enact.

If agents follow their Team Reasons, then each of them will choose to pursue political reform, collectively bringing about the best outcome. So perhaps Optimific Reasons just *are* Team Reasons – at least in cases like *The Altruist's Dilemma*.

8. The Problem of Higher-Order Coordination

But, as Henry saw, this conclusion is also unsatisfactory. It would be best for each agent to pursue politics *only if others were to do so*. Act-Utilitarian reasons tell agents to cooperate only *conditionally*, and strictly conditional cooperators lack a trigger to initiate cooperation. Team Reasons were introduced to break that impasse. They determinately tell agents to cooperate to bring about the best option that can be brought about in a situation. But the existence of *reasons* to pursue politics doesn't *force* people – even rational, moral people – to pursue politics! If enough of Henry's peers didn't realise – or didn't agree – that they had Team Reasons to pursue politics, then it would be sub-optimal if Henry pursued politics regardless.

So we face a higher-order coordination problem. Just as the question of whether Henry should pursue political reforms depends on whether others are going to do the same, now we can see that the question of whether Team Reasons are Optimific Reasons depends on whether other people are *in fact* going to follow Team Reasons. The problem can be stated as follows:

> *Higher-Order Coordination:* In a coordination situation, Team Reasons are Optimific Reasons if enough people are going to follow Team Reasons; if not enough people are going to follow Team Reasons, then Act-Utilitarian Reasons are Optimific Reasons.

You might think we should solve this problem by reference to what we can *predict* each agent will do. And this does seem apt when a single agent be quite sure that others, due to immorality, stupidity or sheer pig-headedness, are not going to behave cooperatively. But when several morally-motivated rational agents are on the stage, this would just reprise the original problem of *The Altruist's Dilemma*. If each agent will follow Team Reasons only if it is predictable that others are going to, then they will not follow Team Reasons, and so will not pursue politics. Conversely, if any of them follow Team Reasons *in deciding whether to follow Team Reasons*, then they will pursue politics *whether or not the others are going to*.

Thus, the *Higher Order Coordination Problem* seems just as intractable as the initial coordination problem. Neither kind of reason seems to be determinately optimific until *after* the agents have acted. But then the principle that we should act according to optimific reasons *could not guide agents in acting*. And that is precisely when we need reasons! Indeed, as I have presented it, the only Utilitarian interest in reasons is in helping us to bring about better outcomes – a theory of reasons that could only be used to assess acts after the fact could be of no intrinsic interest.

9. The Dualism of Moral Reasons

So we need to revisit the construal of Optimific Reasons. My original formulation was:

> *Optimific Reasons:* Morality assigns to agents the reasons that it is best for them to follow; the optimific moral reasons are the reasons that will guide agents in such a way that leads to the best outcomes.

But this is ambiguous. In the Altruist's Dilemma, it might mean that morality assigns the optimific reasons for *all* the relevant agents – the entire audience of DBWC – to follow together. In that case, the Optimific Reasons are Team Reasons. But it could also mean that morality assigns the optimific reasons for *whichever agent is asking the question*. If Henry is wondering what he should do, and he predicts that not enough of his peers will follow Team Reasons, then the Optimific Reasons for *him* are to respond to the predictable behaviour of his peers by pursuing charitable giving – that's to say, the Optimific Reasons for him are Act-Utilitarian.

I think both readings are legitimate. There are two quite different senses in which reasons might be optimific.

> *Outcome-Optimific Reasons*: When there is a best outcome that can be collectively brought about by moral, rational agents, the optimific reasons are the ones that direct these agents to bring it about.
>
> *Prediction-Optimific Reasons*: The optimific reasons for each agent are the ones that lead them to take the options with the greatest expected utility, given what can rationally be predicted about the actions of all the other agents.

Both Act-Utilitarian and Team-Reasons are optimific *in one sense of optimific*. In other words, I think there is a dualism of Utilitarian reasons, between Team Reasons that tell us to choose between options that are collectively open to groups of moral agents, and Act-Utilitarian reasons that tell each agent only to choose between options open to *her*.

10. What Explains the Dualism?

How could this be true? You might think that *Prediction-Optimific Reasons* are the only reasons that a Utilitarian should concern herself with. After all, our interest is in *making this world a better place*, not in performing acts that *would* be useful in an imaginary world of rational perfection.

Henry articulated a simple version of a response:

> *Why Can't We Do What's Best?* If there is a best outcome that can be collectively brought about by moral, rational agents, the only thing that would stop them from achieving this best outcome is their own decision not to cooperate. And, so far as they are morally rational, the only thing that would make them decide this is if there were no reasons to cooperate. Since it would be best if they cooperated, there must be reasons to do so.

It seems hard to square with *Meta-Deontic Utilitarianism* that there might be cases where the only obstacle to the best outcome obtaining is morality!

More deeply, Henry's thought shows the complex relationship of reasons to the causal structure of the world. The argument I gave against *Outcome-Optimific Reasons* assumes what we can call, following Bernard Williams (Smart & Williams, 1973), the:

> *Reactive Concept of Reasons*: What reasons a decision-maker has is fixed by all the other facts in the causal nexus. Questions about what can be predicted are *prior* to questions about what choices would be best in fixing what an agent has most reason to do.

This view treats each individual decision point as somehow special, detached from the nexus, whereas every other decision is just a brute fact of nature. But if reasons can guide *my* actions, then they can also guide *other people's* actions, and so they partially determine, if defeasibly, what happens at other points in the causal nexus. So we should instead accept:

> *The Reciprocity of Reasons and Facts:* Optimising reasons attempt to guide agents to bring about the best outcomes, reacting to the facts in the causal nexus about what is going to happen; but they also guide other points in the causal nexus at the same time. Neither deliberation nor prediction can be strictly prior to the other.

Judgements about what is best attempt to *react* to the causal nexus, but they also *determine* parts of the causal nexus. In a world where there are many agents trying to do what is best, there is often no single determinate fact of the matter about whether their beliefs are true or not, because they are all simultaneously and reacting to the world and shaping it at the same time.

Thus, I think that the Dualism of Moral Reasons follows from the nature of optimising thought in a world with many agents attempting to optimise. When we are all simultaneously reacting to and shaping the world, there is often no unique fact of the matter as to what option is best.

11. Humean Eliminativism

We might reach an eliminativist conclusion here. I suggested that *if* Axiological Utilitarians accept a theory of reasons for action, they should hold that agents have optimific reasons. But now I have argued that it is often ambiguous which reasons are optimific. We want reasons in order to guide us towards the best outcomes – but if the guidance of reasons is ambiguous, what use is it? Perhaps Axiological Utilitarians should *abandon* the search for an account of reasons.

I think this is one way of understanding the normative ethics of Hume. Hume can be seen as an Axiological Utilitarian – he thinks that what ultimately matters, in ethics, is utility. He appeals to utility in the justification of various norms of virtue and approbation. But, as I read him, Hume offers no theory of reasons.[4] He does not, for example, say how conflicting norms are to be weighed to yield all-things-considered judgements about what an agent has most reason to do. In this spirit, Axiological Utilitarians might adopt:

> *Humean Utilitarianism*: We can assess social norms in terms of their utility: there are more and less useful patterns of thought and feeling, strategies for deliberation, and norms of praise and blame. But we cannot apply Utilitarian principles directly to actions – there is no such thing as the best action, and no such thing as the action that agents have most reason to perform.

If Humean Utilitarianism is true, there is just no question of entering the debates between Act- and Indirect- Utilitarians – all of these views are asking a question that cannot be answered. Of course, in deciding how to act or how to apportion praise and blame, we will be guided by the norms and standards we have adopted, and hopefully these will be useful ones – but there is no sense in which these are "merely" rules of thumb, for there is no deeper fact about what reasons we "really" have by reference to which we might make this invidious comparison. Rules of thumb are the only action-guiding principles a Humean Utilitarian can admit, so there is no available sense in which they are "mere".

Moreover, this view leads naturally to the idea that Utilitarianism should be more interested in questions of social reform than in casuistical questions of individual morality. We can't ask whether agents behave rightly or wrongly in artificial counterfactual scenarios like trolley cases, for there is no deep fact about moral reasons to determine such judgements. All we can do is assess the norms and practices of our society for their utility, and try to reform them where we find them lacking. Of course, Hume himself was a political conservative – but in his social views, for example his condemnation of the ascetic "monkish virtues", he showed the reformist potential for this kind of approach.

12. Overcoming the Dualism through Social and Political Change

However, I think there is another option for responding to the Dualism of Utilitarian reasons, without endorsing eliminativism. For Act-Utilitarian Reasons and Team Reasons can *align*.

[4] Like Millgram 1995, I do not think that Hume is an instrumentalist, who thinks that what we have most reason to do is to satisfy our desires. On my reading, Hume does not believe in "reasons", in the sense that contemporary ethical theorists speak of them – as discreet practical considerations that can be added up to yield a verdict about what an agent *must* rationally do. But whether or not this is in fact Hume's view, I think it offers an interesting possibility that contemporary ethical theorists might otherwise overlook.

Suppose DBWC *were* able to make its audience *default cooperators*. They would be doing what they had Team Reasons to do when they cooperated. But each agent would *also* have *Act-Utilitarian* reasons to cooperate. If Henry can *predict* that his peers will pursue politics, then, for him, pursuing politics is also the option with the greatest expected utility. So, in this case, pursuing politics is optimific *in both senses*. It is *unambiguously* what everyone should do.

But *becoming* a default cooperator is more than just deliberating by reference to Team Reasons on a case-by-case basis. In the case I just gave, it is true that each agent has both Team- and Act- reasons to cooperate; so if any one of them were to start being guided *only* by Act-Utilitarian reasons, she would still cooperate. But if more than a handful of agents started to behave this way, the alignment would break down: it would no longer be determinately true that the option with the best expected utility was the cooperative one. In such situations, agents guided *only* by Act-Utilitarian reasons would be, in a sense, *free-riders* on their cooperative peers.

It's a familiar idea that social and political norms are supposed to mark out domains of cooperation, and, through both enculturation and external sanctions, prevent free-riding. The usual assumption is that this is supposed to align *altruistic* and *egoistic* motives. But I think we also need to align *individualistically altruistic* and *collectively altruistic* motives. Social and political norms are needed to mark out the places where we should be default cooperators, to bind us to *be* default cooperators reliably, and to distinguish such situations from situations where, due to the absence of social organisation, it is best for us to follow individualistially altruistic – Act-Utilitarian – reasons.

There is thus a far more robust role for norms and rules in a Dualistic Utilitarian view than there is within an Act-Utilitarian framework. The Act-Utilitarian sees such principles as, at best, mere "rules of thumb" – heuristic devices to help imperfect agents do what Act-Utilitarianism recommends. If possible, it would be preferable to do away with them, and simply apply Act-Utilitarianism directly. But on my view, creating a norm-bound social world is part of what makes the very best outcomes rationally and morally chooseable. In other words, these structures are *deeply* indispensable – political and social reform is not an overlay, to make up for the failings of agents – it is what allows morality to speak with a univocal voice in the first place.

There are, I think, at least three aspects to the kind of social and political project I envisage here. The first part is one of organisation – establishing contexts in which groups of agents become default cooperators. The second is psychological – instilling a virtue of solidarity. An important lesson of the Trade Union movement and other radical political movements is that political collectives fighting towards a shared goal may do better when individual members refrain from individualistic strategic reasoning – that's to say, from being motivated by distinctively Act-Utilitarian Reasons. And the third aspect of this project involves the use of social and legal sanctions to deter agents from deviating from cooperative norms as Act-Utilitarian freeriders.

In other words, I think the early Utilitarians were right to emphasise social and political reform over questions of individual morality – even if not for the reasons that they supposed.

13. Conclusion

I have argued that Act-Utilitarianism is not the best version of Utilitarianism, even given purely Utilitarian premises. Rather, there is a Dualism of Utilitarian Reasons – both Team Reasons and Act-Utilitarian Reasons deserve equal footing. There is more than one way to do our best.

This vindicates two features of the early Utilitarian tradition. One is the failure to state the "Principle of Utility" as an unambiguous theory of reasons, and the apparently excessive stress the early Utilitarians placed on "secondary principles". If my arguments are correct, then Utilitarian reasons *are* ambiguous, and principles deserve a greater status than contemporary Act-Utilitarians give them. The second is the early Utilitarian emphasis on social and political reform, rather than on personal morality. If my view makes sense, then such reform may be the best way to *make* the recommendations of morality unambiguous.

As for poor Henry – what should he do? In one sense, the answer is genuinely unclear, and my theory aims to explain why this is so. But perhaps one thought might tip the scale in favour of political reform. For, with luck, perhaps political reforms can help to create a world in which Henry's successors in altruism need not face the kind of dilemma that bedevils him.

Bibliography

Bacharach, Michael. 2006: *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ: Princeton University Press.

Berkey, Brian. 2018: 'The Institutional Critique of Effective Altruism'. *Utilitas* 30 (2):143–171.

Colman, Andrew & Gold, Natalie. 2020: 'Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games'. *Topoi* 39 (2):305-316.

Gibbard, Alan. 1965: 'Rule-Utilitarianism: Merely an Illusory Alternative?'. *Australasian Journal of Philosophy* 43.

Gold, Natalie & Sugden, Robert. 2007: 'Collective Intentions and Team Agency'. *Journal of Philosophy* 104 (3):109-137.

Hooker, Brad. 2000: *Ideal code, real world: a rule-consequentialist theory of morality*. New York: Oxford University Press.

Lazari-Radek, Katarzyna de & Singer, Peter. 2014: *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. New York: Oxford University Press.

Lyons, David. 1965: *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press.

Millgram, Elijah. 1995: 'Was Hume a Humean?'. *Hume Studies* 21 (1):75-94.

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

Railton, Peter. 1984: 'Alienation, Consequentialism, and the Demands of Morality'. *Philosophy and Public Affairs* Vol. 13, No. 2. 134–171.

Regan, Donald. 1980: *Utilitarianism and Co-operation*. Oxford: Clarendon Press.

Schelling, Thomas. 1960: *The Strategy of Conflict*. Cambridge MA: Harvard University Press

Smart, J. J. C. & Williams, Bernard. 1973: *Utilitarianism: For and Against*. Cambridge: Cambridge University Press. Edited by Bernard Williams.

Srinivasan, Amia. 2015: 'Stop the Robot Apocalypse'. *London Review of Books* 37: 3–6

Sugden, Robert. 2015: 'Team Reasoning and Intentional Cooperation for Mutual Benefit'. *Journal of Social Ontology* 1 (1), 143 – 166

Woodard, Christopher. 2017: 'Three conceptions of group-based reasons'. *Journal of Social Ontology* 3 (1):102-127.