Introduction

Some years ago, I set out to create/discover the correct normative theory and get at the truth of morality. In this, I would say I was partially successful. I wrote a theory called freedom consequentialism, and it has numerous advantages over other moral theories. It applies to all moral agents, rather than only those capable of experiencing certain emotions; it protects persons' ability to pursue different ends, rather than asserting that everyone is pursuing the same end even if they do not know it; it could be used as a value system to solve the AI control problem with far less risk of tyranny or perverse instantiation than existing moral theories; and it avoids many of the classic objections to consequentialist theories, including the demandingness objection. However, there is a significant problem at the core of freedom consequentialism: the problem of how to weigh freedom. This is the problem I am hoping you can solve or help to solve, and this is why I am asking for your help in the first place.

What follows is a description of freedom consequentialism and the core problem of weighing freedom over different things. A sort of freedom consequentialism primer, as it were. The goal here is to bring you up to speed on the theory and the core problem sufficiently for you to be able to contribute meaningfully to its resolution.

I have attempted to be as brief as possible while still explaining the core features of the theory and the state of the problem as it stands. Because of this, some of my reasoning is not explained as fully as it might be in other circumstances. Most of the material here is also available in my doctoral thesis, *In Defence of Freedom Consequentialism*, and my paper, *Solving Satisficing Consequentialism*. If you want more detail, you can find it in these older works. I have left links in footnotes[1]. However, there are cases where what I wrote in these earlier works is incorrect or incomplete. When this is the case, I have endeavoured to explain the discrepancy in a footnote. I have also assumed a reasonable amount of knowledge about philosophy since this is being sent to philosophers. Non-philosophers are certainly welcome to read this and try to solve the problem of weighing freedom, but they may find this primer a bit esoteric at times.

So, without further ado, let us begin.

[1] In defence of freedom consequentialism is archived and available through the University of Canterbury. Link: https://research.ebsco.com/c/g6lqvo/search/details/oaawob6lvn?q=freedom%20consequentialism
Solving satisficing consequentialism is published in Philosophia. Link: https://link.springer.com/article/10.1007/s11406-021-00382-y

Assumptions and theory-selection criteria

First, I think it is worth being explicit about the assumptions I am making and the theory-selection criteria I think are important to determining which moral theories are most persuasive. It always bothered me that moral theories would treat their core assumptions as if they were obviously the case. For example, the assumption that a great many moral theories make that making an exception of yourself is irrational. Whether or not you agree with this assumption, it would be better to be explicit about making it from the off. So, here are both the assumptions that I am making and the specific theory-selection criteria that I think are relevant to selecting moral theories.

First and most obviously, I wanted this theory to be true. I was aiming at accurately describing the truth of morality. This may not seem worth mentioning, but it is worth noting since it means that any potential solutions to the problem discussed later should not be internally inconsistent or require us to assume the truth of any propositions that are false. This assumption also implies that either moral realism or moral error theory is correct, so solutions rooted in subjectivism or relativism will not be considered.

On a related note, I have also assumed that morality is the way that moral agents ought to be or act, where "ought" is understood in an objective and universal way. This brings me to my next two theory-selection criteria: that freedom consequentialism ought to be universal and objective. The principles it describes should apply at all times and in all places, across all possible worlds, regardless of culture or personal views.

I also assumed that morality is about how moral agents act, and I take moral agents to be free, rational agents. By "agent" I mean a conscious entity capable of taking some action[2]. By "free, rational agent" I mean agents that possess free will and the capacity for rationality. Because of this, the next of my theory-selection criteria is that freedom consequentialism should apply to all free, rational agents[3].

Another assumption that I have made is that ought implies can. I take it to be true that in order for it to be the case that some agent ought to do something, it must be the case that they are able to do that thing. For this reason, freedom consequentialism must not require agents to do things that it is impossible for them to do.

Further, I assumed that morality is intended to be action-guiding. Because of this, freedom consequentialism should be able to provide moral guidance in specific circumstances. This criterion is particularly relevant to the core problem that you are (presumably) attempting to solve.

One of the more controversial assumptions I am making is that consequentialism is the correct way to approach morality. While I am always happy to discuss consequentialism, solutions to the problem under discussion that require giving up consequentialism are unlikely to be seriously considered.

These seven theory-selection criteria are the most important ones used in the creation of freedom consequentialism. However, there are two more that are worth mentioning as well.

First, in line with Occam's Razor, I assumed that a theory that postulates more entities is generally, all else being equal, worse than one that postulates fewer. Obviously, there are some issues with this

---

[2] Including actions that are wholly mental.

[3] In *In Defence of Freedom Consequentialism*, I also included the criterion that the theory should apply to *only* free, rational agents. I am omitting that here as it is likely to lead people to think I mean that moral patients cannot exist, which is not what I intend to claim at all. In the interest of stating my assumptions though, I am assuming that things that are not moral agents are not subject to moral demands.

idea, but as a general principle, it works well enough in most circumstances. Because of this, simplicity has been sought when possible.

Finally, I have attempted to ensure freedom consequentialism is at least somewhat in line with commonly held moral intuitions. I am *not* assuming that moral intuitions are a good guide to truth and, in fact, I think they are largely unreliable. However, we must work with what we have, and there are at least some pragmatic benefits to aligning with commonly held moral intuitions. So, any proposed solutions should aim to be at least somewhat in line with our moral intuitions, but a reasonable amount of divergence from these intuitions is acceptable.

To summarize, these theory-selection criteria are:

1. Likelihood of truth (including internal consistency and not relying on propositions that we have good reason to believe are false)
2. Universality
3. Objectivity
4. Applicability to all free, rational agents
5. Action-guidingness
6. Achievableness (possible to live up to)
7. Consequentialism
8. Simplicity (in the sense of not postulating entities beyond necessity)
9. Extent to which the theory is in line with commonly held moral intuitions

As well as being in line with freedom consequentialism generally, any solution proposed should fit these criteria.

Measure of value

The "measure of value" is the term I use for the thing that a consequentialist theory treats as valuable. So, for classical utilitarianism, the measure(s) of value would be happiness and lack of unhappiness. For freedom consequentialism, the measure of value is, unsurprisingly, freedom. However, since "freedom" can mean a lot of different things, I should explain what I mean by it here.

When I use the word "freedom" in this context, I mean the ability of free, rational agents to understand and make the choices that belong to them. The reason this is used as the measure of value for freedom consequentialism is that it allows the theory to apply to and take account of every possible moral agent (freedom consequentialism can also take account of moral patients such as infants, as discussed in chapter five of *In Defence of Freedom Consequentialism*[4]). Many normative theories are not capable of this. To use the above example of classical utilitarianism, any free, rational agent that does not experience happiness or unhappiness is presumably not morally relevant according to the classical utilitarian. There might well be an entire planet, or many planets, of free, rational agents that cannot experience these emotions, and classical utilitarianism seems to write the inhabitants of these possible planets off as not morally relevant. However, free, rational agents by definition have free will, so can freely make choices. Rationality can also be understood as the ability to understand one's own choices and the reasons for making one over another, so a free, rational agent also has the capacity to understand their choices. Because of this, the ability to understand and make choices is something that is shared by all free, rational agents, by all *moral* agents, in all possible worlds[5]. So, by using this as the measure of value, freedom consequentialism can take account of all moral agents in all possible worlds.

However, the measure of value here is not merely the ability to understand and make choices. It is specifically the ability to understand and make choices that *belong* to the person in question. This is for a few reasons, such as irresolvable conflict occurring if everyone has an equal claim to all choices. For example, if your choice to keep your car in your driveway were morally equivalent to my choice to steal your car, we would quickly have an irresolvable conflict. This would also heavily conflict with moral intuitions. So, the kind of freedom that freedom consequentialism is concerned with is specifically freedom over those choices that belong to the person in question. The choices that belong to a person, or the choices a person has a "right" to make if you prefer, are the ones over those things that they own, specifically their mind, body, and property. Owning one's own mind and body is fairly easy to establish because this is essentially just self-ownership, especially in the case of the mind. Owning property is a bit harder to establish, and it is a bit of an odd concept generally, but certainly *if* we can own property, then it is something we ought to have freedom over, so we will assume that we can own property and include it on the list of things we can have freedom over.

So, freedom consequentialism's measure of value is the ability of persons to understand and make their own choices, specifically those choices regarding what to do with their mind, body, and property. It is generally best to think of this kind of freedom as to be *protected* rather than *promoted*. So long as a person is able to understand and make their own choices, they have their freedom. It is

---

[4] This is done by using consciousness as the basis for personal identity and granting moral considerability to infants on the basis that they are the same individuals who will be free, rational agents in the future. This could have all been covered in a footnote, but I am aware that not everyone reads them, and I think it is worth mentioning that freedom consequentialism does not treat small children as morally irrelevant.

[5] It might be objected that freedom consequentialism cannot take account of entities incapable of action but, since "capable of acting" includes actions that are wholly mental, an entity would only count as incapable of action if it could not perform even a voluntary mental action, in which case taking account of it morally seems strange.

only the freedom over things that already belong to a person that matters, so getting more stuff over which a person can have freedom is not morally valuable. Things are bad, on this measure of value, when they prevent a person from being able to understand and make their own choices. Doing good—which I will discuss more in the following section—using this measure of value, is just a matter of preventing or reducing bad things from happening. This will be important in the next section, as it allows freedom consequentialism to avoid the demandingness objection.

It is also worth noting that there is sometimes a distinction drawn between positive and negative freedom, where positive freedom is freedom *to* do, have, or be something, whereas negative freedom is freedom *from* some external constraint. I do not think this distinction is particularly helpful, and personally prefer thinking in terms of Gerald MacCallum's triadic relationship of freedom, but I will say that the freedom being used as the measure of value here can certainly be limited/violated/reduced in a morally relevant way by not just the actions of other people but also by many other things. If a person's choice to continue living, which is theirs to make as a choice about one's own mind and body, is taken away from them as a result of murder, that is a morally bad thing. However, it is morally bad in the same way and for the same reasons for that choice to be taken away from that person due to a tiger, or a virus, or simply ignorance (for example, the person in question believes they are able to fly and jumps off a building because of that belief, but they cannot, and they fall to their death). What is important is that the person in question is able to understand the choice they are making such that they are able to apply their rationality to it[6], and are able to make it for themselves. It is also worth noting that coercion can take someone's choice away from them in a morally relevant way if the threat they are being coerced with would itself take their choice away. The classic example of this is robbing someone at gunpoint. The person's choice to keep their money is being taken away from them by force in that they are presented with a choice to lose their money or lose their life. This coercion would still be morally bad in the same way if the gun in question was fake (so long as the person being threatened does not know this), as the person is still giving up their money under threat of losing their life, even if that threat turns out to be hollow.

[6] This is intentionally quite a low bar to set for understanding, and means that people need not be well-informed in order to be free. They only need to understand what choices they are making and what it means to make those choices, such that they are able to apply their rationality to them.

Determining the right thing to do

While I have now explained the measure of value that freedom consequentialism uses, that does not tell us how we ought to act. We still need to know how to determine the right thing to do. In this, freedom consequentialism is a kind of satisficing consequentialism[7], but one that avoids the objections that Ben Bradley raises against existing forms of satisficing consequentialism as well as the demandingness objection. This form of satisficing consequentialism, which I call counterfactual agent-central satisficing consequentialism, or CASC, is somewhat complicated, but I will attempt to explain it as clearly as possible.

So, when determining how we ought to act, first we must determine which of our potential actions are good, bad, or neutral. The way we do this is by comparing two scenarios to determine whether your actions were better than if you were not "in play" as a moral agent.

Scenario one: You perform the action.

Scenario two: You have a mental blank, as if you had briefly stopped existing as a moral agent, rather than perform that specific action[8].

An action is bad if scenario one has worse consequences than scenario two.

An action is good if scenario one has better consequences than scenario two. And it is good to the extent that it either causes no bad consequences ("causes" in the sense that they do not happen in scenario two) or the good consequences it produces could not have occurred without producing at least that much bad[9].

All other actions are morally neutral[10].

This method allows us to demarcate good, bad, and neutral actions from one another, which then allows us to determine which actions are permissible, impermissible, obligatory, and supererogatory.

---

[7] In *In Defence of Freedom Consequentialism* I resisted using the term "satisficing consequentialism," largely because I wanted to separate freedom consequentialism from existing types of satisficing consequentialism. In *Solving Satisficing Consequentialism* I articulate the method of determining what actions are permissible, impermissible, obligatory, and supererogatory more clearly and properly refer to this as a form of satisficing consequentialism.

[8] This can either be a mental blank that only applies to whether you perform the action in question or a complete mental blank where you are entirely not a moral agent until after you would have performed that action. If the latter, then an additional criterion is required in determining the goodness, badness, or neutrality of the action in question. Specifically, that the action is bad if: it causes a net bad consequence that would not have occurred if you continued being a moral agent, stipulating that your stopping being a moral agent would cause no bad consequences that would not have occurred if you continued to be a moral agent.

[9] "Good to the extent that" introduces the concept of goodness as a matter of degree, but this is not relevant to the recommendation of CASC and is barely relevant to the evaluation of actions by CASC. It is just a matter of calling some actions that produce net good consequences but do so imperfectly in a way that produces some unnecessary bad consequences somewhat good, rather than neutral.

[10] This method of defining good, bad, and neutral actions is a significant departure from what I have written previously. In the past, I used the counterfactual of the agent not existing at all, rather than simply not existing *mentally*. This idea of using not existing mentally as the relevant counterfactual was developed after a lot of discussion with Doug Campbell, and it seems like it avoids some weird features of the original counterfactuals. The original version had already undergone a lot of revisions to avoid various problems, but I think this version avoids those problems and also some stranger ones, like how scientists might react to a person suddenly and temporarily ceasing to exist. In both cases, the core idea is to consider what the world would be like if the agent in question were not "in play" as a moral agent, but that agent briefly ceasing to be a moral agent appears to be a cleaner way of doing this than the agent briefly ceasing to exist entirely.

To start with, performing bad actions is generally impermissible. There are some exceptions, and discussions to be had on how exactly we should think about small risks to others or the moral cost of just living but, suffice to say, one should generally not perform bad actions if one has the option. Of course, one may find themselves in a situation where only bad actions are available. In situations like this, one should choose the least bad option. So, in situations where a person can only perform bad actions, that person is required to perform the best available action(s). All non-optimal actions in situations like this are impermissible, and the optimal action (or performing one of the optimal actions) is obligatory. In short, when all options are bad, the agent must act in order to maximize value[11].

When one has performed a bad action in the past, they ought to rectify that action to the extent that they can. If it is possible to alleviate the violations of freedom one caused in the past, one has a moral obligation to do so as long as this causes no other bad consequences or those consequences are outweighed by the good produced and that good could not be produced without producing at least that much bad. A person has this moral obligation because the world is worse off due to their existence as a moral agent and they are morally responsible for this state of affairs[12].

Determining which good or neutral actions are "good enough" to be permissible and when actions are obligatory rather than supererogatory is more difficult. Just because an action is "good" in the sense described above does not mean it is morally permissible to perform that action, as it may be obligatory to perform a better one. What we need to determine is the minimum amount of good an agent needs to do.

To do this, we use three factors. The first is the amount of good that needs doing, by which I mean the amount of good that is required before the world is in such a state that no one has any moral obligation to improve it.

This is an area where the freedom that freedom consequentialism aims to protect has the advantage over other measures of value, in that there is a theoretical world in which no one has any moral obligations to improve the state of that world. In the case of, for example, happiness, there is no clear limit on how much happiness we could create, so the amount of good that needs doing is essentially infinite. Only theories with finite amounts of good that need doing can really use CASC, which in the consequentialist landscape amounts to freedom consequentialism and perhaps negative utilitarianism.

The next factor we need to consider when determining an agent's minimum goodness threshold is their own ability to do good. Because of the connection between ability to do good and obligation to do good, in that the latter cannot exist without the former if we assume that ought implies can, we can reason that the amount of good a person is obligated to do is dependent upon the amount of good they are capable of doing. Consequentialist theories, at least consequentialist theories that make demands, assume or directly claim that the amount of good one is morally obligated to bring about is related to that person's ability to bring about good. For example, according to classical utilitarianism, the amount of good a person is obligated to bring about is the maximum that person is able to bring about. This factor will help us to "parcel out" the amount of good that is required of

[11] This sort of scenario is much less likely to occur using the new method of defining actions as good, bad, or neutral than it was using the old, so this point may only be relevant in very strange, niche circumstances.
[12] This is another addition from previous work. Originally, I compared actions to the counterfactual of the agent never having existed as well, but this caused problems, such as considering seemingly good actions to be bad because of past events, even if those events were not the fault of the actor. So that was cut in favour of an obligation to right past wrongs, which seems a more sensible approach.

each person, as it will be a function of how much good is required in total and their particular ability to bring about good.

The third and final factor in this process is the uniqueness of a person's position to bring about some particular good. It seems reasonable to assume that those who are in a unique position to do some particular good have a greater obligation to bring about that good, all else being equal, than those in a relatively common position to do a particular good, as it seems that those in a relatively common position to do a particular good share in that responsibility and thereby lessen their portion of it. For example, it seems intuitively obvious that if a person comes across a man dying in the street, they have the training necessary to save him, and no one else is around to assist, then they have a greater obligation to do so than they do to donate the amount of money required to save a person on the other side of the world.

So, when determining how much good one is obligated to do, one needs to consider how much good needs doing, one's ability to do good, and how unique one's position to do that good is (or how many people are in a similar position to do that good). We can use these three factors like X, Y, and Z coordinates to locate the minimum required threshold of good a person is obligated to do. Performing at least that much good and remedying past wrong actions to the extent that you can, so long as this does not create any bad consequences (or any bad consequences are outweighed by the good produced and that good could not be produced without at least that much bad), is obligatory; performing more is supererogatory; performing bad actions (so long as good or neutral actions are available) or failing to perform the minimum required amount of good is impermissible; and performing neutral actions (so long as the minimum goodness threshold is met) is permissible.

CASC, as I have just described it, does not fall prey to the demandingness objection because it rarely requires people to maximize the good. It also does not allow for what Ben Bradley calls the "gratuitous prevention of goodness" because it does not allow people to perform supposedly good actions that actively prevent more good from occurring, as Bradley is concerned satisficing consequentialism would allow for. It is through this method that freedom consequentialism determines the right thing for agents to do in specific circumstances.

The problem in question

So, hopefully you now have a reasonable understanding of freedom consequentialism. You know that it is a form of satisficing consequentialism that treats the ability of persons to understand and make their own choices as the measure of consequences' moral value. Which brings me to the core problem: that of weighing freedom.

Freedom consequentialism includes freedom over multiple things, our minds, bodies, and property, but it is not clear how much weight we should give each of those things in our moral decision-making. When faced with a decision between funding a drug that saves one life every ten years and a drug that restores eyesight to five blind people within the same timeframe, how do we know which to fund? The obvious answer is "whichever protects the most freedom," but how do we know which that is? We might reason that the freedom over one's eyes is less important than the freedom to live, but the question is *how much* less important. Five times? A hundred times? A thousand? Presumably, there is a number, and it seems unlikely that it would be morally correct to blind the world to save one life, but it is not clear what that number is. This is further complicated when we take into account that in many cases that violate someone's freedom, there are also issues of duration and intensity to consider. For example, imprisoning someone for a year is much worse than imprisoning them for a day, burning someone's house to the ground is worse than smashing their window. So this is the problem: how do we determine the value of freedom over different things?

It could be suggested that the freedom over different things is incommensurable in the sense of not being able to be weighed against each other at all, but it seems like that is not the case. After all, it is all the same thing: the ability of free, rational agents to understand and make their own choices. So it seems that this should be solvable. I have considered several possible solutions, of which one seems to be clearly the best, but I will detail each so that you know what ground has already been trod.

Perhaps the most obvious method for determining what to do in cases where we must weigh freedom over one thing against freedom over another is to simply claim that each choice is worth the same as each other choice. We could say that for each choice a person ought to be able to make that they are denied, we can count that as one unit of value. Then, we could determine what the right thing to do is by the number of choices a person ought to have that have been denied them. I call this the Choice Counting Method (CCM). There are some significant problems with CCM. One is that it is not clear how to "count" choices, as it is not clear what qualifies as "one choice." This may make it hard to be precise enough to be action-guiding. Also there is the issue of duration and intensity, which both complicate the idea of counting choices further. Another problem is that it seems, at least intuitively, that the freedom to make some choices is more important than others. It would be deeply counterintuitive to say that someone's freedom over their arm is of equal importance to their freedom over their cheese grater.

Another possible solution to this problem is to rank freedom over different things by objective importance. Then, when there is a choice to be made between violating freedom over two different things, we can say that one ought to breach the less important one. I call this position the Objective Order Method (OOM). OOM has the advantage of resolving dilemmas relatively easily. Once one has determined the objective order of importance of different freedoms, one can easily determine which freedoms one should choose to protect at the cost of which others in would-be moral dilemmas. Because freedom consequentialism aims to be objective and universal, it seems that it should rank freedom over different things in an objective manner. For this reason, OOM seems to be an attractive position. However, there is a significant problem with OOM. It is not clear how we would determine

the objective order of importance of different things that a person ought to have freedom over. All we could do is guess at a method to do so. Worse, it is not clear that even were we to stumble upon the correct method, we would recognize it. Another problem with OOM is that even if we were to know the order of importance of all the different things we should have freedom over, that would not necessarily tell us how much of one is worth how much of another. If, for example, we knew that freedom over whether we live or not is more important than freedom over whether we keep our eyesight, we could not tell from that how many persons' eyesight are worth one person's life. An objective order, even if we could determine one, is not enough. What we need is a method of weighing freedom over these different things.

A method that is something of a hybrid of these two approaches is the Freedom Subsumption Method (FSM). This method suggests that freedom over some things subsumes freedom over other, less important things. For example, the freedom to live might be said to subsume all other choices a person ought to have freedom over. If this is the case, then these lesser choices could be thought of as a portion of these greater choices, and be weighed by reference to how big of a portion they are. One problem with FSM is that there appear to be plenty of choices that are more important than others without subsuming them in any way. For example, the choice to refuse sexual consent appears to be much more important than the choice to not have your cutlery stolen, but the one does not subsume the other. We could potentially avoid that problem by reference to some larger choice that subsumes them both, such as the choice to continue living, but this may cause more problems than it solves. First, it is not clear how great a "proportion" of the choice to continue living any individual choice can be. The choice to continue living seems to subsume almost infinitely many other choices and, while it seems some of them are obviously more important than others, it is not clear that those choices necessarily make up a greater proportion of the choice to continue living. This has the potential to become unacceptably subjective and also faces the issue of how we determine what proportion of one choice another choice is with enough specificity for FSM to be action-guiding.

There is also the possibility of using our moral intuitions to guide our weighing of the freedom over different things. We could use these intuitions to determine the importance of different freedoms. In circumstances where we need to choose between breaching one kind of freedom or another, we could fall back on what is best supported by commonly held moral intuitions. I call this method the Intuitive Order Method (IOM). However, IOM has significant problems. The most obvious of these is that we do not have good reason to think our moral intuitions are a good guide to the truth, so we do not have good reason to rely on them.

Which brings me to the method that I think is most promising, that of determining the relative importance of freedom over different things by reference to a *preferred order of wrongs*[13]. That is, we could rank freedom over different things by the order in which the individual in question would choose to have that freedom violated. For example, we can imagine that one person would prefer to be imprisoned for a few days than to receive a beating, while another person would prefer the opposite. In the case of the former person, that person's freedom to not be beaten would be more important than their freedom to not be imprisoned for a few days. In the case of the latter person, the reverse would be true. I call this the Preferential Order Method (POM). POM has the advantage of determining the value of freedom to make different choices by reference to the choice of the individual which, given freedom consequentialism is all about the ability of persons to understand and make their own choices, seems appropriate.

[13] I am not suggesting anyone make an actual list, as that list would presumably be infinite and in constant flux. Rather, we would compare different options by reference to those affected's preferences.

One might object that POM introduces an additional measure of value, that of preferences, and so does not align with the goal of simplicity, but this is not the case. Preferences need not be valuable in order to use them to rank the order of freedom over different things. Rather, preferences are being used to determine which of the things a person should have freedom over they value the most. As the morally relevant kind of freedom here is the ability of persons to understand and make the choices that belong to them, persons ought to be able to choose which of their freedoms are breached. POM allows persons to rank their own freedoms in order of value, and uses their preferences as a way of determining the order they are in. So, as it allows persons to choose the order of importance of their own freedoms, POM seems like an attractive method[14].

POM even allows us to resolve some conflicts between different groups[15], because we can use preference for a chance of something occurring as a proxy for preference for a percentage of that thing. If we use the example of funding a drug that saves one life every ten years or a drug that restores five people's eyesight during the same period, we could simply ask all those involved whether they would accept a 20% chance of death to restore their eyesight (this can be asked hypothetically of the person who is not blind). If everyone involved agrees that restoring one's eyesight is worth a 20% risk of death, but not a 21% risk of death, then we can treat one life as equivalent to five people's eyesight for that group. If they would all only accept a 10% risk of death to restore their eyesight, then we can treat one life as equivalent to ten people's eyesight for that group.

But, and it is a rather large but, this does not tell us what to do when preferences conflict. If, for example, the people who want their eyesight restored and the people who want their lives saved do not agree on what chance of death it is worth risking to restore one's eyesight, then POM does not give us a way of determining what we should do. Remember that the preferences themselves are not important, so we cannot simply determine this by majority. The strength of one person's preferences, or indeed many persons' preferences, is not relevant to the moral importance of the freedom to make a specific choice that belongs to a different person. Also, it is persons' *actual* choices that matter, not what their choices might be if they were smarter or considered the issue more carefully, so referring to ideal preferences does not help either. We simply do not have a way to resolve cases where preferences conflict regarding which choices we ought to have freedom over are most important.

So, I think POM is the way forward to resolving the problem of weighing freedom over different things, but I do not have a good answer to how to resolve conflicting preferences within the context of freedom consequentialism. Hopefully, someone reading this is able to solve this last piece of the problem where I have failed (or else is able to find some other method that works even better). I think freedom consequentialism is the best normative theory we have available, the closest we have to a representation of moral truth. It has the potential to resolve all moral dilemmas and provide the ideal value system to solve the AI control problem. However, until this problem is solved, it is a tiger without teeth, and not a fully functional, action-guiding normative theory. So please, help me.

---

[14] It is worth noting here that POM is *not* preference utilitarianism. The preferences of those involved are not themselves morally important. They are only used to order the importance of freedom over choices that belong to the persons in question. For example, if the whole world would prefer that a great artist did not set fire to a work they had just painted (and owned), that would not count against them burning it at all, as the choice would belong to that person alone.

[15] This is different from what I said in *In Defence of Freedom Consequentialism* because of further work I have done on this since.