

Freedom and Binding Consequentialism

David Braddon-Mitchell*

Abstract

The paper proposes a new version of direct act consequentialism that will provide the same evaluations of the rightness of acts as indirect disposition, motive or character consequentialism, thus reconciling the coherence of direct consequentialism with the plausible results in cases of indirect consequentialism. This is achieved by seeing that adopting certain kinds of moral dispositions causally constrains our future acts, so that the maximizing acts ruled out by the disposition can no longer be chosen. Thus when we act we do the best we can, which is all that is required for rightness according to act consequentialism.

1. Introduction

Direct consequentialism is the view that the right act is the one that will, of all those available to an agent, maximizeⁱ expected value. Indirect consequentialism—in its original form—is the view that the right act is the act produced by a disposition, rule or habit whose inculcation maximizes expected value. Frequently this is supplemented with the thought that the direct consequentialist evaluates dispositions or rules by whether they will produce maximizing acts, and the indirect consequentialist evaluates acts by whether they are produced by maximizing rules or dispositions.ⁱⁱ

Neither of these is remotely satisfactory. Act consequentialism is coherent enough, but the answers it gives in moral puzzle casesⁱⁱⁱ give pause to all but the most ideological consequentialist. We are expected to sacrifice our near and dear when doing so will aid famine relief, we are expected to commit barbaric acts if doing so will result in slightly fewer barbaric acts being committed and so on.

Indirect consequentialism seems to deliver more plausible judgements on the cases. For example, versions that combine rules, characters and dispositions may tell us that it is acceptable to have special bonds with our near and dear inasmuch as the best (consequentially understood) kind of rule or character may be one that favours the near and dear to some extent.

Unfortunately the indirect versions face a trilemma. First, they may accept that the reason for choosing the rule or disposition stems from an ethic of maximizing utility. But then the very same consideration—the overarching morality of consequence maximizing—will tell us to break the rule whenever doing so will maximize. But then we will have the underlying principle that grounds indirect consequentialism telling us to do something contrary to what indirect consequentialism itself requires, which is incoherent^v. The incoherence can be removed by specifying that the best rule must be one which makes us perform the maximizing act on each occasion: but then it seems as though there is a danger of collapse into the act version.^v Finally the indirect consequentialist might claim that the desirability of acting according to the rule with the best consequences is not itself justified consequentially, but is itself a basic and possibly deontological moral commitment. But then we have strayed from consequentialism: and in any case it is hard to see what appeal this would have as a basic non-consequential principle.^{vi}

There are versions of consequentialism that try to have their cake and eat it too: to adopt both rule and act consequentialism as correct in their appropriate domains. A useful term for an overarching position that has this effect is *global consequentialism*.^{vii} Global consequentialism evaluates every kind of entity by the particular consequences of that kind of entity. Certain kinds of act consequentialism evaluate *rules* not by the overall consequences of adopting the rule, but rather by whether the rule produces maximizing acts. Certain kinds of rule consequentialism evaluate *acts* by whether they are produced by the maximizing rule. Global consequentialists on the other hand, evaluate acts and rules, characters, motives and so on (as well as other entities) separately by their *own* consequences. So acts are evaluated by whether they maximize, and rules by whether *they* maximize (it should be remembered that the consequences of a rule may outstrip the acts it produces). This means, then, that such a view has some of the features of both direct and indirect consequentialism. But it is still not a satisfactory reconciliation. When faced with the particular acts that act consequentialism seems to endorse, and which are *prima facie* counterexamples to that view, global consequentialism says that the acts are indeed right: even though the right rule prohibits them, or the right character will not perform them, or the right motive will militate against performing them. So this hybrid view still

accepts that the act is right— but somehow the agent is forgiven for not performing it because this omission is out of virtue (Driver^{viii}) or out of good motives (Adams^{ix}). Some versions of this view see these cases as one where there are inconsistent obligations: one must perform a certain act, but one must instantiate a certain rule character etc that is inconsistent with performing the act. So a serious cost of such views is either accepting that there are all things considered inconsistent obligations (an attractive feature of consequentialism to many is that when all is balanced and you do the best you can you have done the right thing), or else requiring some story about permissible wrongdoing. Plausibly, however, it is worse than that. Our intuitions about some of the cases are that it would not just be permissible wrongdoing to not kill one's spouse so as to save a few strangers, it would be morally required wrongdoing—and surely that is as close to a clear contradiction as we'll get in the murky realm of ethical intuition.

What is needed is an entirely new take on consequentialism: one that reconciles the coherence and action-guidingness of direct act consequentialism with the plausible answers in cases of indirect consequentialism. The solution is a kind of act consequentialism that sees many important moral acts as being concerned with the inculcation of moral dispositions. The indirection comes in only because many of the acts we perform are ones that we know will affect our own psychology and thus our own moral dispositions or rules. It will be crucial for such a view to explain how it is possible to adopt dispositions which are maximizing, even when it is foreseen that they will lead us to perform what look like unmaximizing acts. This paper proposes just such a new consequentialism. It is the view that an important part of moral choice is the adoption of moral dispositions that causally filter our future choices to make us *unable* to take seriously in practical deliberation certain options. But if we choose the best of the *available* options at the time of act selection, we do all that a thoroughgoing act consequentialism demands of us, and thus choose rightly. I call this new kind of consequentialism *binding consequentialism*. Both first order acts, and the acts which affect our dispositions are evaluated and chosen consequentially, and this is made psychologically possible by emphasizing one of the roles of moral dispositions as binding us in a way which reduces our future choice set. What is distinctive about this approach is that I show that acts that affect

moral dispositions have to be seen as causal interventions in our own psychology, which limit our future psychological capacities. What will undergird my account will be a compatibilist take on free will, according to which choices of habit and disposition are in general free, but many acts in the grip of habit and disposition are not. We will be free to choose from the available options, but not choose from those ruled out by moral dispositions.

The paper is divided into six sections. In section two I consider the general idea of inculcation of dispositions as causal intervention in psychology, both in the case of rationality and morality, and introduce a version of act consequentialism according to which options have been removed by such intervention. In section three I consider the issue of in what sense we are acting freely when we choose from among the available options, and sketch a compatibilist account of free will that will undergird the theory. In section four I contrast my view with some examples of global consequentialism, in particular Adams motive consequentialism and Drivers' consequentialist virtue theory, and argue that my way of reconciling the intuitions behind direct and indirect versions of consequentialism is to be preferred. Before concluding, I examine and solve some problems for my account raised by considering cases where dispositions are inculcated with the purpose of spreading evil.

2. Dispositions as causal interventions

I will start the process of making my view plausible by comparing it with an issue in the philosophy of rationality that has a similar structure. There are interesting cases where it is rational to make oneself irrational (or at least rational to limit the domain of one's future rationality). Simple cases involve things like the beserker drug.^x Suppose that people have access to a drug which will make them disposed to vigorous fighting. And suppose further that it is common knowledge that this drug exists, and that I know that I do not have the acting skills to pretend to be an irrational revenge taker. In a situation in which I am about to be attacked, I know that if I flee I will be chased and badly beaten, but if I fight to the best of my abilities I will inflict very serious damage on my assailant, but in the process will be killed or far more seriously hurt than if I were to flee. What should I do? Plainly the rational strategy is to flee, for that will produce the best outcome. Fighting

would be foolish and result in death or maiming; much worse than a very bad beating. But before I am assailed, I could take the drug in full view of the assailants. Should I do this? Perhaps I should, for if they see that I have taken the drug they will believe that I will fight to the best of my abilities, and thus hurt them more than they think worth while. I take the drug in the hope that I will never in fact act under its influence, though of course I may be unlucky. And if I take the drug often enough, or take a permanent version of it publicly in the hope of general protection against thugs, then I should expect on occasion to get unlucky: I will eventually come across an irrational assailant who will attack anyway and I will, under the influence of the drug, fight back rather than run away. But if this is not expected to happen too often, then it may be rational enough to take the drug.

When I fight back, do I act rationally?^{xi} The drug has causally restricted my options. Running away is not an option, though various forms of fighting or standing my ground are. If I choose from the options available to me the one that among them maximizes, then, modulo my inability to choose from a wider set, it seems I act rationally. For rationality (like morality) does not require that we do what we cannot. Rationality is a theory of choice among options. It cannot require that we consider all options, for no finite mind could even list the options. It does not require that I fly with wings when I cannot, even if that would have the best outcome. And similarly for mental acts: rationality does not require that we solve differential equations so difficult that no human can solve them, even if that would have the best outcome.^{xii} So on my picture, rather than the picture in which the drug makes me act irrationally, I rather act rationally given the constraints the drug places on me.

Of course one attempt to explain why we act rationally in these or similar cases is some extended account of rationality according to which we do indeed choose from a full range of options, but we freely choose the option which is not narrowly maximizingly rational, but is the rationally best option in some special sense of rational. The special sense might be that it is rational to do what it was rational to agree to do, or it is rational to do what it is rational to intend to do. This is the strategy of David Gauthier^{xiii} in *Morals by Agreement* and Joe Mintoff^{xiv} in the case of the Toxin Puzzle. There are well known objections to this solution, and I think what unites them is that it is hard

to see how this extended sense of 'rational' deserves to be called 'rational'. It shares with rule consequentialism the problem that the considerations that recommend breaking the rule in the one case, or breaking the agreement or intention in the other, are the very considerations that justify the rule or agreement. So if these considerations have authority in justifying the rule or agreement, why do they not have authority in recommending a breach?

These are cases where we are considering prudential reason. We will now consider cases where direct causal intervention applies in the domain of ethics. I will only slightly modify the berserker case to make it into an ethical rather than rational example.

Suppose that I am making the same choices about fighting and fleeing as before, but now I am considering the choices from a moral perspective—part of which is in terms of the influence these choices will have on my philanthropic projects. One again, consider the option of fighting when confronted by the assailants. Perhaps it is *morally* wrong to fight in these circumstances (unless we think that fighting back will affect the future actions of the assailants: let us suppose we know this not to be so), for fighting will create more disutility—summed over the assailants and oneself—than running. And yet the act of taking the berserker pill in full view of the assailants may maximise expected utility; for it may be expected to prevent any attack and thus generate none of the disutility of either running or fighting. In that case it would be morally right to take the pill. But is it morally right to fight under the influence of the pill? Well, as the pill has reduced my options, so long as I choose the best option from those available then I do what is right, even if I can see that there is a best option no longer available to me. I do what is right, even if not what is best. Just as when we do the best action to alleviate world poverty, we do the right thing, even if there is an option—writing a brilliant best seller that makes billions of dollars which we would then donate—which is not available to us because of causal limitation on our mental powers.

How essential is the pill to our story about the assailant? Suppose that I am unable to go out in the world to pursue my philanthropic interests because of the fear of assailants. If I do go out, it is rational for me to run, and I will. It is also morally right for me to run, for I know that if attacked and hurt badly I can do less good than if I stay at home. But I know that there are

people out there who, on the whole, walk the streets unscathed. There is something about their bearing that makes the street thugs think they will fight back. I invest in videotapes about how to walk in an ape like way that will make assailants fear my tendency to retribution, but it turns out I am no actor. Then I read a book that explains that sadly there is a nomological connexion between seeming like someone who fights back and being one. So I buy more books and tapes (perhaps I enlist a therapist) and by exposing myself to the right kind of violent film, by visualizations and various methods of psychology I make myself into someone who will inevitably fight back. I thus walk the streets doing good largely unscathed because my new disposition is written in my walk and bearing in a way that I was unable to simulate.

Of course this new disposition will eventually get me into trouble. I will be badly maimed when I fight back, and my philanthropy will be in recess for months. Will I be doing the right thing when I fight back? Well if the case of the pill is one where we think that I did not have freedom with respect to the option of fleeing, and thus chose the best action from the restricted set, then it is hard to see how we could come to a different conclusion in this case. The methods of therapy are not so very different from the methods of psychopharmacology. The pill alters my brain structure in a way that makes my choice set limited. The therapy acts on my brain structure in a way that makes my choice set limited. If I do, in some sense, the right thing in the first case, then so do I in the second case. If I did the right thing in undergoing therapy, and it produced a consequentially justified disposition, then indirect consequentialism will judge my act of fighting to be right. On my account direct consequentialism will also judge my act to be right, for it is the best from a set causally limited by the therapeutically induced disposition. Thus there is no clash between the judgements delivered by direct and indirect consequentialism.

To take another example, let us look at the case of our moral strictures against killing. Many think that we act rightly when, in choosing from the available options that do not involve killing, we select the one with the best expected value. We do the right thing, even if it has a somewhat worse outcome than an option that did involve killing. How can this be, if we are consequentialists? Well, if I act as a result of a disposition that typically prevents me from taking seriously the killing options, I have done the best I

can. My disposition may itself be justified by having the best expected outcomes. It is better to put up with some cases where we will not kill when killing may have had a slightly better outcome, than perhaps to be misled by our future judgements and sensibilities into unwarranted killing.^{xv}

These are cases, then, where we bind our future selves to limit the range of choices from which they choose. Why would we perform such binding in these cases? Because we think that with a restriction in place we will err on the side of moderation less frequently than we would err on the side of excess if our choices were unconstrained. Or, in a manner inversely analogous to the case of the berserker, we seek the advantages of co-operation, and we know that someone with a pacific nature is more likely to get such benefits, and all the scope for promotion of the good that comes with them. Of course morally the *best* thing to do would be to give the impression of being non-violent, but reserve the power to kill when calculation told us that, say, killing would have maximizing consequences that outweighed its cost. But, as with thuggery, most of us are unable to fake it. Actually adopting a non-violent disposition is the only way to persuade most people that we are non-violent.

The causal binding feature of this view gains plausibility, I think, when we consider what would be required for a solution to what Robert Frank has called the commitment problem for rational choice theory.^{xvi} The problem here is that whether one is signalling that one will act thuggishly in a narrowly irrational way in the future, or that one will cooperate in a narrowly irrational way, such signals are unbelievable if one is known to be rational.^{xvii} So one must be able to signal that one is not purely decision theoretically rational, and for various reasons deception on this front is not a universally successful strategy. What might work is actually to causally limit one's future capacity to act in a way that is unconstrainedly decision theoretically rational. If one is to preserve one's general rationality into the future the best way to understand this process is as causal limitation on one's future behaviour so as to eliminate the options which, though decision theoretically rational, one must now signal will not be performed if we are to gain trust.

A crucial thing to note is that at least sometimes the good consequences of the best disposition are not produced via the acts that the disposition produces. It might be that a pacific disposition is detected via pacific acts, but equally it may be detected via by-products of the disposition, such as pieces

of non-voluntary behaviour and demeanour. It is this case in general which allows for the possibility that a maximizing disposition could result in quite a lot of non-maximizing behaviour, and thus require that the disposition be causally fettering, since otherwise we will have reason to adopt the disposition but not to perform the individual acts

The upshot of these examples is that we can see how a direct consequentialist evaluation of the act can give the same results as an indirect consequentialist evaluation. The indirect consequentialist account would judge that in the case where I do not kill despite the slightly greater benefit of killing, I do the right thing in virtue of following the rule. My direct consequentialist account also allows that the act is right, since the act is, of those available, the one with the best expected outcome. So my account uses the plausible machinery of direct consequentialism, but tracks the plausible judgements of indirect consequentialism.

2.1 Why is moral training the same as therapy?

The next claim, then, is that the inculcation of moral dispositions and rules is really very much like therapy; by moral training I alter myself (or allow myself to be altered) in such a way as to limit my future choices. When I become sensitised against killing—when I adopt a maxim of not killing—I causally alter my brain so that my future choices do not include killing. Luther's stand against the Roman church—*Hier stehe Ich, Ich kann nicht Anders*—becomes literally true.

But it might be objected that moral training, or inculcation of dispositions, is not the same as therapy. Therapy is about manipulation in a causal way, moral training is about improving moral sensibility. We need here to distinguish between two kinds of moral dispositions: weak and strong.

A weak moral disposition^{xviii} is one where we operate on rules of thumb, but our deliberative powers rest in the background ready to intervene should the disposition be about to cause an action that the agent then judges to be less than optimal.

Strong moral dispositions bind our future selves. In inculcating a strong disposition, we adopt a disposition that ensures that our future choices are causally constrained. We do this precisely because this will maximize overall, either because we cannot trust our deliberative powers in certain

circumstances, whether because they would be too slow, too unreliable or where the fact that others see that we are not casually bound may have bad consequences^{xix}. Thus we can make a morally justifiable choice to adopt these dispositions even knowing that sometime in the future we will act with worse consequences than we would have if we had deliberated.

Both kinds of moral dispositions are part of our lives, but it is the second kind that we are concerned with here. And this is the kind of moral disposition that it is hard to see in terms other than causal intervention. What we do when we teach our children how to be moral is to give examples that we think will rub off, and when we engage in programs of moral self-improvement we engage in exercises which affect our sensibilities in a way we think will affect our future choices. How can we make sense of the adoption of a moral disposition unless we think that it will causally restrict our future range of choices in some situations?

To answer this question, we need to distinguish between two kinds of moral training. One sort of moral training is what we engage in to improve our judgements as to what makes actions right, either by improving the decision making process, or by improving the values built into it. Suppose that we were to change our assessment of what makes actions the best: it would make perfect sense to build that new assessment into our decision making system without restricting in some sense our future range of choices. We would be modifying our future choice behaviour, in light of changed views about what is the best basis on which to choose, and then leaving our future selves free to choose what they think will be all things considered best.

But there is another kind of moral training: one where we have not altered our opinions about what the best calculational formula is, nor what the right values are. In this kind of training we think that it is best to be such that we will *not* choose according to correct principles of decision making at some point in the future. This is something, which can only be achieved by genuine causal restriction of options. For we can foresee that, if we were perfectly rational moral and free, we would act in accord with the correct principles in future situations unless we do something to prevent us from choosing freely, rationally and morally.

Strong moral dispositions are required, then, in this latter case: where we have reason to think that there is a case to be made for overriding what would

be calculated on a case by case basis – much as when there is a case for overriding what would be rational to decide on each occasion in the berserker case above. And if that is so then it only makes sense to train morally in this way on the assumption that it really will limit our choices. Just as there would be no point in taking the pill if what it did was to enhance rationality (for then it would ensure that we ran rather than fought) there would be no point in moral training if all it did was slightly enhance the accuracy of consequential calculation. The point is that strong moral dispositions are supposed to circumvent the deliberations that a free rational and moral agent would make.

Some might object that this second kind of moral training is not the correct account of moral dispositions in cases where we have not changed our opinions about the values or rationality. Suppose instead, they might argue, that what we do when we engage in moral training is not to affect our outcomes by constraining choice, but rather to limit the influences on our choices. Perhaps moral training is about making better choices by cutting off influences for ill. It's about reducing our selfish desires so they are not inputs to future decisions, or alerting ourselves to salient features of future situations we may encounter which will be important to correct decisions and which may otherwise be overlooked.

I do not disagree with this account of some kinds of moral training, but this is simply a third kind of moral training. When this is the right way to describe the training, then moral training is indeed just like taking therapy to improve one's implementation of an unchanged conception of rationality. In the particular case of the berserker pill, an alternative that enhanced narrow maximizing rationality would not have served the same purpose. But there are many situations in which our rational decisions would be improved by better calculation, or more care in choosing relevant data. So it is with moral decisions—moral training that enhances our ability to calculate and evaluate consequences and attend to morally salient factors in situations is no doubt a good thing. When this kind of moral disposition is what we are talking about, then on my picture a pure act consequentialist account of right action should be at work. The disposition improves our capacity to calculate, and we do the right thing when we calculate correctly and act accordingly. But these are exactly not the kind of dispositions that indirect consequentialists are concerned with. Indirect consequentialists concentrate on rules of thumb,

moral maxims and dispositions that make one act in some cases other than how one would by merely doing the calculations and then acting in response to them. And only in these cases is there any issue about the way in which the direct and the indirect evaluations stay in track. Thus they cannot be mere fodder for the making of better decisions.

So the idea, then, is that when we take on strong dispositions we limit the range of our future choices causally. In the future we act freely, but only with respect to the available options. In some cases the best act will no longer be available to us, so that in choosing the best available act we do all that consequentialism demands. All of this demands that I say something about freedom, which I do in the next section.

Before I do that however, I should deal with an objection that might forestall the point of investigating issues of freedom. It might be argued that the right act is not the act that is available to the agent and that has the best consequences. It is rather the act that is available to a *reasonable agent in those circumstances*. This would rule out the physically impossible acts, and those beyond the reasonable mental powers of an agent, but it would include as genuine options things that the agent is unable to do insofar as he is unreasonable. Plausibly, having taken a berserker drug, for example, makes me unreasonable, however reasonable it may have been to take it. So the option of not fighting is available to the reasonable agent. Similarly in the case of moral dispositions, if someone has limited his capacity to morally reason, and act with practical wisdom on those deliberations, and as a result is in the grip of a disposition not to kill (except perhaps if the consequences are extreme), then the option of killing is not available to him. But if the right act is the one with best consequences available to the reasonable agent, and the reasonable agent does not have these causal strictures on their deliberative powers and practical reason, then he still acts wrongly by not selecting the option that involves killing.

But this depends on an account of 'reasonable agent' that is tendentious. Is the reasonable agent one who has failed to adopt a disposition that a reasonable agent would in fact have adopted? That is what would be required. We would be claiming that the reasonable agent is someone who has not acted reasonably in the past. Of course there could be an account of 'reasonable' according to which what makes one reasonable is unconstrained

decision theoretic rationality. But it does not seem to be relevant to the question of evaluation of acts—for it makes the right act the act that would be performed by someone *who is not as they ought to be had they been reasonable*. It thus would not give us an account of right that would help in the business of choosing acts given the kinds of agents that we are and ought to be.

2.2 Free Consideration Rejected

Why must we sometimes see the adoption of dispositions as a causal intervention that reduces our choices? An objection might go like this: if you can have a pill which will eliminate some options, why not have a pill which allows that you consider all options, but which ensures that you will freely choose to decline all the (for example) killing options?

The brief answer is that if the pill left us free and unfettered, how could it guarantee that we would always decline a killing option? The longer answer is this. Consider a case where killing saves a few lives. Suppose that we are free of will, unfettered in our psychological capacity, rational and knowledgeable, and moral. To the extent that we are free of will we have the power to select a killing option. To the extent that we are moral we will choose the maximizing act (even if we have chosen some maximizing rule that rules it out: we now see that we can create more utility by breaking it in this case). To the extent that we are unfettered we still have the psychological capacity to kill, or can take the option seriously, to the extent that we are rational and knowledgeable we understand the expected outcomes, and have the practical reason to bring about what our morals require. So we can be sure that we will choose the killing option.

If the pill then (or the moral training) prevents us from choosing this option, it must make us immoral, fettered, ignorant or irrational. The kind of moral disposition I am considering here makes one psychologically fettered: it takes away one's ability to kill, or to take seriously the killing option, while leaving us unaffected with respect to the remaining options. Of course this is not the only way the pill could work. The pill or training could work by, for example, giving us false beliefs and making us into deontologists. Or it could work by making us systematically miscalculate the expected value of the outcomes. All of these perhaps happen. But to the extent that there are act consequentialists who can inculcate these moral dispositions – and there are –

the option that is sometimes taken is fettering. We become people who are unable, without reprogramming, to kill.

3. A need for a theory of free will

In the previous section I said that strong moral dispositions circumvent the deliberations that an agent who is free of will, unfettered, rational, knowledgeable and moral make. This paper is about the kind of dispositions that circumvent these deliberations by fettering agents' psychological capacities, so as to limit the range of choices.

What does it take to count as a limitation on a range of choices? It means that some of the choices are ones that we no longer have because we are unable to choose them. Of course this is not an inability of the kind that we have when we cannot choose to sprout wings and fly. It is rather a kind of psychological inability. It is an inability such that, while we are free to choose from among the remaining options—in some sense we retain our free will—we are not free with respect to the missing options.

So in what sense are we not free? Someone who has a disposition against killing, or a disposition not to lie, might not have any kind of obvious impediment that prevents her from killing or lying. She may, even as she does not lie, think to herself that she could lie if she wished, but chooses not to because she has internalised the non-lying disposition.

Of course there is a full causal story that explains why she will not lie, and it may make sense from that perspective to describe her not lying as a product of her not being free. But of course there is equally a full causal story about how she chooses when she deliberates, or how she chooses when she takes on the dispositions. There is, plausibly, a full causal story about everything from the perspective of which it looks like there is no free will.

If we are not to be eliminativists about free will, or chain ourselves to the questionable metaphysical presuppositions of libertarianism about the will, then some kind of compatibilist account is required. A compatibilist account of free will is one that accepts that an action can be freely chosen if it is fully causally determined by factors prior to the agent. Thus a compatibilist about free will must admit that we can have a range of choices open to us from which we choose a particular one, even though there were predetermined causal factors which eliminated all the other options. So merely knowing that

a range of options has been causally limited does not guarantee that we were unfree with respect to them. So, *a fortiori*, knowing that a moral disposition eliminates certain choices from being actualised does not mean we are unfree with respect to those options.

Thus we might still take an act consequentialist view according to which we have done the wrong thing, because the right option was one of those that we did not perform, even if the fact that we did not perform it was causally determined. Of course this does not mean that the compatibilist need deny the principle that ought implies can (and the contrapositive that if one can't do something, it is not the case that one ought do it). But it means that there must be some compatibilist reading of 'can' with respect to an option, which is compatible with that option's being causally ruled out.

What the present paper requires is that there is some compatibilist reading of 'can', and of 'free will', according to which when one chooses from options left open to one by one's moral dispositions, one chooses freely from among the things one can do, but according to which the options ruled out by the dispositions are ones one can't select, and with respect to which one is not free. But in its bare abstract form, compatibilism does not tell us that the options ruled out by the dispositions are ones we can't select, nor does it tell us that the options ruled in are ones we can select. Compatibilism only claims that there is some basis for choosing among determined actions those which are determined but freely chosen, and those which are not. To motivate the claim that the genuine options are only those that are consistent with our moral dispositions, we will need to put a little more flesh on the theoretical bones of compatibilism.

3.1 Compatibilism and moral dispositions

The full account of compatibilist freedom is not something that needs to be settled here. Nor is it something remotely uncontroversial. But I will assume that out of the various accounts available there are principles which render it sufficient for freedom that one is acting and choosing on the basis of a well functioning deliberating device that is reliably connected to action. One acts freely just so long as this device is efficacious and functioning normally, regardless of the causal determination of the device. This extremely abstract

formulation captures something in common with higher order desire accounts,^{xx} bio-functional accounts,^{xxi} ideal desire or ideal higher order desire accounts and so on.

So the thought is that the way to characterize freedom of the will is that for an action to be performed with free will, it must be under the control of one's own decision-making apparatus in some way. In many cases our day-to-day actions are settled by individual deliberations of this kind, and thus are free. Perhaps even more commonly, actions are under the control of simple habits or rules of thumb—i.e. dispositions to produce behaviour in a more or less reflex way—but crucially where this is mere calculative convenience: if the decision centre is presented with information that the current circumstances are ones where the habituated behaviour is inappropriate, the habit is overridden. These dispositions are a generalized version of weak moral dispositions, and cover all cases where deliberation is moved to a back-up role. These weak dispositions produce freely chosen behaviour, since they are still sensitive on a case-by-case basis to the deliberative control unit, if in a slightly indirect way.

The kinds of moral or rational dispositions we are concerned with here, however, are strong dispositions, where even when we realize that the behaviour we are about to engage in does not achieve the very goals that justified inculcating the disposition in the first place, we proceed anyway. We have made the decision to constrain our future behaviour through inculcating a disposition to limit the range of options considered by our decision-making module^{xxii}. We have treated ourselves—or at least our future selves—as mechanisms to be manipulated by causal intervention, just as we can imagine doing to one another. The future behaviour, insofar as it is produced by our decision making system without external interference is free, though it is not *free with respect to* the eliminated options; just as our decisions are in general not free with respect to what we cannot do. Our causal intervention has changed what we are able to do in the future.

Compatibilist accounts of freedom of this kind are accounts of when actions are performed with free will. Freedom of the will is a connexion between action and the control device that is the will; freedom of action is having one's acts under the control of a control device working in the appropriate way. The behaviour of the control device is determined, but on

this view it is some kind of category mistake to ask about freedom of the control device insofar as it is working correctly. What counts as incorrect working is part of the difficult task of coming up with a substantive compatibilist theory of free will, but everyone who believes in rationality and is not a metaphysical libertarian has that task. Some plausible candidates for sufficient conditions for improper working include being under the control of a further control device external to the self—normal environmental causal impacts are irrelevant. Thus actions can be free or unfree. There can be a lack of freedom due to the control device being improperly connected to another control device, as in brainwashing, hypnotism and so on. There can be lack of freedom due to the actions not being caused by the control device. So free action is action that is:

(a) Under the control of the correct control device, working correctly (and no other).

Of course the act that I performed when I took the pill or began the inculcation of the disposition or moral habit is free on such an account, because it is under the control of the control device. The actions performed under the influence of the pill are also free, but the range of options has been causally limited by the pill. So these considerations give us the following taxonomy of freedom of action:

- (1) Ordinary actions under the control of the control device are free when the device is not under the control of a further device (regardless of other environmental influences on the device) and the device is functioning normally.
- (2) Actions under the control of weak dispositions are free.
- (3) Actions under the control of strong dispositions are free when the control device exerts a synchronic effect in choosing between the available options, but the actions are not free with respect to the ruled out options.

I do not expect these constraints on an account of freedom to be completely intuitively satisfactory. But this is because of the nature of the debate about free will. I do think that there is a kind of conceptual priority to libertarianism. If it were coherent, and if the properties the libertarian believes in were ever realized in actions, then those and only those acts would be free.^{xiii} A compatibilist account is a second best account; it is an account of

what we should call 'free' given that there are tensions among the core ideas of freedom of the will. As such it cannot hope to be completely intuitively satisfactory. In the next section I address those for whom second best is not good enough.

3.2 What if the error theory is true?

Many are not convinced that any compatibilist account of free will is correct. Some of these are libertarians about free will, and hope that determinism (or determinism plus chance) is false, or that there is direct intervention by some faculty of will that is not naturalistically causally determined or some such. I have nothing to say to such folk. I very much doubt that their view is right or even coherent, though if it is both then I agree that freedom of the will would track the operations of such a faculty and thus whether actions under the control of strong dispositions would be free would perhaps depend on whether the special faculty was in operations in each of the actions.

There are however another group of non-compatibilists with whom I have much more sympathy: error theorists about free will. The error theorist agrees with the libertarian on conceptual matters: she agrees that is *a priori* about free will that if there is any, there must be an uncaused and undetermined effect on options, which causes one out of a range of actions to be performed, where there was no prior determination of the outcome. The error theorist, however, disagrees with the libertarian on an *a posteriori* or perhaps logical level. She holds that this condition that is necessary for free will is either incoherent, or at least empirically ungrounded, and thus that true free will is either logically impossible, or at least not actual. In either event there is, for her, no free will.

What use are my considerations about the nature of freedom to an error theorist? More than you might think, I suspect. Although the error theorist thinks that there is no free will, this does not mean that there is no responsible action. So one might give an account of responsible action (in the sense of actions for which one is in some way responsible) that denies an analytic connexion between attribution of responsibility and the non-existent freedom. Instead of taking the sketch I give above to be an adequate account of freedom, the error theorist might take it to be an account of an important component of responsibility. My account of responsibility would then look

much the same as an account that does insist on the analytic connexion with (compatibilist) freedom, except that the component which was labelled 'freedom' is now labelled as a psychological precondition of responsibility.

Even if the error theorist about freedom insists on an analytic connexion between freedom and responsibility, and is thus also an error theorist about responsibility, this does not rule out an account of the properties in virtue of which we hold people responsible. Of course this may end up being a revival of Sidgwick.^{xxiv} we might promulgate theories about when we should hold people responsible on consequentialist grounds. That is, the best theory or theories about responsibility might be those that have the best consequences.

I think that the account I give here might be recommended on those grounds as well.^{xxv} My sketch of responsibility focuses us on the right level of intervention. When a bad action is the result of a strong disposition, this account encourages us to focus on the disposition, and attempt to change it. When your friend is routinely late, and innumerable instances of annoyance have not changed the pattern, it becomes futile to insist that he simply pull his socks up and exercise more willpower on a case-by-case basis. Instead the focus of disapproval should move to his not having instigated methods of changing the disposition that causes the lateness. When someone is involved in a car crash when drunk, as a result of recklessness that would not happen when sober, the focus of disapproval should be on the disposition to drive when drunk. If that disposition is also under the control of a further disposition to drive when drunk, the disapproval and intervention should focus on their not having done what it takes to undermine that disposition, or to remove themselves from situations in which they will exemplify the disposition. The controlling disposition, whichever it is, is the one that it will have best consequences to modify.

4. Direct Consequentialism Reconfigured

We have, then, a reconciliation of direct and indirect consequentialism. In fact it is formally a purely direct consequentialist theory, since acts are right iff they are the available act that has the best consequences. The reconciliation consists in tweaking the account of 'available' so that the evaluations that direct act consequentialism gives, track the evaluations that indirect consequentialism would give.

How does this contrast with other attempts to reconcile the intuitions underlying direct and indirect consequentialism? There is a number of such theories, all of which I think are species of what Pettit and Smith call global consequentialism. In this section I will focus on two of these views—Adams' motive consequentialism, and Driver's consequentialist virtue theory^{xxvi}—but the basic structure of all such views is quite similar. On Adams' view we evaluate motive sets consequentially based on their overall consequences, and we independently evaluate our acts consequentially. Driver offers a consequentialist virtue theory, where virtues are sets of psychological traits and dispositions (perhaps including motives) that are satisficingly consequentially good, but the rightness of acts is separately evaluated according to a satisficing consequential calculation.

Both these views treat the evaluation of acts, and the evaluation of some indirect property that relates to acts—in Adams' case motives, and in Driver's virtues—separately. The right motive/virtue is the one that has best or good enough consequences, the right act is the one which has best or good enough consequences. On these views if we have adopted a disposition that causes us to act in a way that is non-maximizing^{xxvii} then we have done the wrong thing. But nonetheless, if the disposition is still one that will generally have good consequences, we can approve of the agent for having acted out of virtue. This is why they are both species of global consequentialism. I will call such views *non-binding global consequentialisms*.

This is not the principal point of difference with my view; although I present my view here as a pure act consequentialism—the evaluation of dispositions, motives and so forth is limited to *acts of disposition inculcation or retention*—there might well be a version of my view which evaluated dispositions themselves, rather than acts of inculcation, consequentially^{xxviii}. The principal point of difference, rather, is that these forms of global consequentialism allow there to be a clash between the two components of the theory. Sometimes the right act will not be the one produced by the virtuous person, or the right motives. On my view, there is a connexion between the right disposition and the process that produces (or monitors) the act: the right disposition rules out causally the options that might otherwise have been right, so that the best available act at the time of performance—and thus the right act—will always be the one recommended by the disposition.

Why should we prefer binding consequentialism to non-binding global consequentialisms like Adams' or Driver's? Their views have a number of costs. First there will need to be an account of permissible wrongdoing. Where binding consequentialism says that when you, say, fail to kill your spouse and thus save a few lives you do the right thing because you do the best available thing, non-binding global consequentialism says that you do the wrong thing but permissibly: permissibly because it is done out of virtue or good motives. Alternatively non-binding versions might allow that there can be all things considered conflicting obligations—such as to be virtuous and to do the right thing. It is not appropriate here to argue that such consequences are intolerable costs; some are certainly prepared to embrace them. But many are not: and to them my account may recommend itself. Another worry is that rather than permissible wrongdoing, perhaps it may look like there is *required* wrongdoing: if the evaluations of the dispositions and the acts are indeed separate, and generate their own independent obligations, there may nonetheless appear to be cases where one obligation trumps another without eliminating the trumped obligation. Thus it might be that we are required to prefer to retain virtue or correct motivation and not to kill our spouse so as to save two strangers, but at the same time the obligation to kill remains in force. So we have done wrong in not killing our spouse, but we were morally required to do so. Morally required wrongdoing is surely a significant price to pay for a view.

Perhaps the greatest cost of the non-binding global consequentialisms is that it removes the point of 'right' as a reasonable evaluative concept distinct from 'good' or 'best'. Consequentialism promotes a connexion between the right action and the best action. But it does not identify them for good reason. Rightness is an evaluative notion connected directly or indirectly with choice. The best outcome is almost always one quite outside of our control on absolutely anyone's conception of control. The option according to which one clicks one's fingers and rights all wrong, the option according to which one sells rights to one's image and banks the huge profits in the name of some ideal charity, the option according to which one prevents starvation by proving Goldbach's conjecture and using the resultant prize money, are all options which are better than any that are usually available to us, and yet they

are not what we ought do because, being impossible to us because of physical or psychological deficiencies, they are not really options at all.

This much is agreed territory: Adams' and Driver's version of direct consequentialism respects the idea that the right act is not one of the acts that it is impossible for us to perform. However it turns out that in the case of moral dispositions or virtues, the maximizing acts prohibited by the virtue or motive still turn out to be right. To justify this we would need a story about rightness, which explains how we can rule out acts prohibited by physical constraints and (perhaps) mental illnesses, but not those ruled out by the virtues or motives. The deliberative function of rightness does not seem to support such a story. We need the idea of rightness to give us an account of best available option. We need this idea to choose or evaluate the best available action. But the whole point of inculcating motive sets or virtues is to rule out certain options from serious consideration and deliberation. If the justification for not regarding as right any action that is physically impossible is that it is excluded from deliberation because it is pointless to deliberate over actions that we *know* will not be performed, then the very same justification applies to actions that will not be performed because of the motives or virtues that we have inculcated in ourselves. For an account of rightness to go outside of the range of the possible would merely be to give an account of best action, which would then require us to have a new notion – the best action which we can perform – to do the evaluative work for us. It would be terminologically idle to use 'right' for the action we already can call best, and to have to create a new term to play the rightness role.

Of course this only establishes that there are *some* constraints on options. It establishes that 'ought implies can' is true in general, but not what the correct account of 'can' is. Driver's or Adams' act consequentialist components are presumably not designed to tell us that we ought to do these superhuman things. But focusing on them makes salient the features of acting contrary to virtue or inculcated motives. We treated ourselves as bindable when we inculcated the virtues: *there would have been no point in doing so otherwise*. What justifies a change in perspective so that we no longer see ourselves as bound when we come to act? It is not as though we repeat the calculations that we performed when selecting the disposition, and then act on our current judgement. The central point is that to see the process of strong

disposition forming as rational, requires that we see ourselves as setting in place a causal change in our psychology that binds us, and removes as options things outside what we are morally disposed to do. When we act on dispositions, the field of our deliberation is the field allowed by the disposition, and the connexion between deliberative evaluation and rightness thus guarantees, on my view, that the right action is not outside that realm.

5. The best choices of evil people

What should we say about the choices of people who have deliberately inculcated dispositions to do bad things; to limit their future choices in such a way as to eliminate as options many acts which otherwise would have been real options and which then it would have been right to choose.

We can imagine two ways these dispositions could be adopted. There could be a figure like Milton's Satan who, working under the slogan 'Evil be thou my good!', chooses to adopt dispositions to make him behave in ways that will on balance have bad consequences. Or we can imagine someone out of self interest adopting moral dispositions that will have on balance worse expected consequences, but better selfish maximizing consequences. Such a person might, for example, deliberately engage in a hardening of the heart that will prevent him from being able to give to the poor, by focusing on tales of fecklessness that make poverty seem like despicable weakness.

The problem is that in either case, once in the grip of the dispositions, if the agents nevertheless choose the best of the limited options left to them, then on my account they have done the right thing, for they have chosen from among the genuine options that one with the best expected outcome. So having eliminated the option of giving to the charity collector, an agent reflects and decides that it would be better to refuse politely than hurl abuse at the collector. They decline politely, and it turns out on my account they have done the right thing, even though there is a very perspicuous action – giving the money – that they have not performed.

The first thing to say is that my story inherits from the general consequentialist tradition a distinction between the rightness of actions and the rightness of blame. And the second thing is that such a person *has* performed relevant actions that are wrong: the actions of inculcation of the wrong disposition.

The combination of these makes acceptable this somewhat peculiar consequence. For there is a focus for attributions of rightness and wrongness, which is at the level of the adoption of the dispositions. And from the perspective of where we should focus our annoyance, anger and disapprobation, it should (consequentially) be at that level. For complaining that the instances are morally poor choices on a case-by-case basis will do little good. The point is that the outcomes on a case-by-case basis are bad ones: and the moral problem lies in a culpable *lack* of deliberation, or limited scope of deliberation, and the moral critic needs to insist on the wrongness of the decision to impose these limits and demand that it be re-addressed. It is these limits that should be the focus of our moral concern, not the deliberation and choice within them.

It should also be noted that there is something to be said here about the difference between the agent who chooses subject to a strong disposition and chooses the best available option, and the agent who chooses subject to the disposition and chooses a poor one. It is possible for an agent who has morally reformed, and has begun the process of eroding their previously inculcated disposition, to at least conscientiously do the best they can (the right thing), and they are in marked distinction to the agent for whom the procedure of choosing the best available option plays no role, even in choosing between the limited options left by the evil disposition. We can at least see that the person who politely refuses the collector is morally superior to the one who abuses the collector. Perhaps it is this that makes moral sense of drama where we see someone thoroughly gripped by evil dispositions trying to do the (morally) best thing under its constraints. Tony Soprano, say, is quite literally bound by mobster dispositions, but at least some of the time tries to choose the least bad option — and (in some admittedly limiting sense) we feel that he does the right thing, if not the best thing, when he does.

Finally, the idea that such an agent acts rightly may seem counterintuitive because in general there is an intuitive problem with bad outcomes being right ones. This is why some people are moved by the idea of strong moral dilemmas: if all of the actions are in some way disastrous, it seems little compensation to say that an action was the right one because it was least disastrous. And in general there is a temptation to ignore the fact that an option was the best available because we are (often rightly) wary of excuses. If

an agent tells you that they chose a bad option because it was the best available to them due to some lamentable past poor choice, one might be suspicious. And this suspicion may manifest itself not in questioning the truth of the claim that it was the only option available (particularly since the truth of this is hard to establish) but of doubting that the action was therefore right. But all intuition really is doing is signalling that *something* is fishy; that it is the local action which is strictly speaking not right is just the guess one may have as to the location of the smell. The exact location is up for grabs by the best theory. Importantly, on my account, in these cases rightness of the action itself is not the most important point of intervention. The action may be right, but that is not sufficient for praise or approval if it does not spring from a good disposition. And of course we may have a generalized sense that we should somehow intervene in the situation: but it may be subjectively hard to determine whether that is a need to intervene in the agent's dispositions, or to complain about the act. This sense may be the source of our ire and disapproval in these cases, even when it turns out that we can be brought to believe that the right action was locally out of the agent's power.

A final sweetening of the pill, if it is needed, is that much badness is precisely not of the kind where dispositions for the bad have been inculcated. It is rather exactly of the kind where dispositions of a moral kind have *not* been inculcated. Instead the agent takes on the perspective of choice on a case-by-case basis. They choose between the selfish and the virtuous act, and choose the selfish. Thus their act does indeed count as wrong. It's an empirical guess, but I think that most wrong action arises from allowing oneself to unboundedly choose, rather than lashing oneself to the mast of the bad.

5.1 Temporally extended acts

The individuation of acts is a very controversial topic, but one which is very relevant to the issues here. For depending on how we individuate acts, it may turn out that when someone inculcates a disposition, they are performing some other act as well—some temporally extended act. In this section I will deal with some difficulties for my view which such an idea creates, but also use it to propose another way to make sense of evil folk who choose the best available option.

Suppose someone presses a button, knowing that it will result in plague germs being dropped on some village or other in twelve months (but not which village). They could plausibly be held to have performed an act of biological warfare against the village. This act, we can assume, would not be a right act. They chose from various options the one of pressing the button and thus having the village destroyed.

I take it that the above is uncontroversial. But what should we say if instead of being a button that causally impacts the external world, it's a (metaphorical) button that causally constrains the agent's future psychology? Imagine that the button is a kind of pill which the agent takes, knowing that it will eliminate as options for him all acts which fail to result in his delivering the plague germs in person in twelve months. Perhaps he knows that he might be talked out of the act in the future if he doesn't constrain himself.

We might not think that the act the agent performs *at the time of delivering the germs* is one that is free with respect to not delivering the germs. However, with respect to the actions performed at the time of taking the pill, it is hard to see how there are relevant differences between the actions performed when the pill is taken, and actions performed when the button is pressed. In each case the agent knowingly chooses certain outcomes, and causally impacts on the future to ensure that outcome. In one case the part of the world used as means is some device connected to the button, in the other it's the agent's own body and psychology affected by the pill, but I do not see an argument for the relevance of that difference. If this is so, we must assume that the agent who takes the plague pill has performed a temporally extended act of biological terror at the point of taking the pill.

This situation raises problems for the kind of direct consequentialism I advocate here. For on my analysis, when someone adopts a disposition which is maximizing, and then does some individual act under its sway which is not maximizing, they still can be counted as having acted rightly insofar as they have chosen the best act from a limited range of alternatives. However, what if it turns out that such a person can be accused of having performed a temporally extended act which is wrong, since he earlier acted in a way that he knew would have the consequence of sometimes producing bad outcomes? At the time of choice the alternatives were not yet removed, and at that point he performed a wrong act which culminates in the non-maximizing

outcome (of course at the same time he performed many right acts which culminate in all the maximising outcomes).

The response I want to make to this case is that although the bad effect was foreseeable (in type if not in token) it was the result of a disposition taken on precisely to minimize effects of that kind. That being so, it seems implausible to claim that in this case there was a temporally extended act of that sort. Indeed if we allowed it to be the case that there was a temporally extended act of that kind, then at the time of taking on the disposition, countless acts—right and wrong—would have been performed. The principle I invoke here is not the strong doctrine of double effect that all foreseen but unintended consequences should not count towards act individuation, just the weaker one that foreseen and unintended consequences which are expected given the disposition chosen, *insofar as the disposition minimizes consequences of that type*, should not count. This is in contrast to the plague case, where bad outcomes (indeed in the particular example the token biological terror) are the intended consequences. This is no doubt a controversial principle but then act individuation is a very shaky branch of philosophy. I think intuitions of moral responsibility, and the rightness and wrongness of behaviour in situations, are much more robust than theories of act individuation. Thus if someone were to charge me with manufacturing a theory of act individuation to match my ethical theory, that might not count as a complaint: more an observation that best practice is at work.

It remains to note that this account of act individuation hands us another nice consequence for the purposes of the previous section. If an agent takes on an evil disposition for the purpose of ensuring bad outcomes, then it is open on this account to impute to her a temporally extended wrong action for each consequence which is on the whole bad. For the extended acts are ruled out only in cases where the acts later chosen fail to promote the values that justified the disposition. Thus we have another way to undermine the worry of the previous section that no wrong act is performed when the evil agent acts (in the best available way – perhaps trivially because the only way) under an evil disposition. There may be no wrong act wholly located at that time; but there may be a temporally extended wrong action. And because the bad outcome was exactly the kind of outcome that the evil agent was trying to

bring about by taking on the disposition, the bad consequences can count as the result of a temporally extended action.

5.2 Agents and time slices

So far I have talked as though agents should be thought of as local strings of time slices, just long enough to deliberate. So, for example, one way to think of the comparison in the previous section between the earlier and later stages of the good and the evil agent is this:

Call the earlier, disposition-changing time slices the 'bosses'.

Call the later, acting sets of time slices, the 'footsoldiers'.

1. Evil Boss chooses (wrongly) to constrain the choices of his footsoldier Elvis.
2. Good Boss chooses (rightly) to constrain the choices of her footsoldier Gladys.
3. Elvis chooses (rightly) to do the best he can in choosing the least evil alternative (which is not the maximising one unavailable to him).
4. Gladys chooses (rightly) to do the best she can in choosing the best available alternative (which is not the maximizing one unavailable to her).
5. If we allow telegraphic acts, then Evil Boss performs a telegraphic act that is wrong, for there is an act which is available to him which is better.
6. If we allow telegraphic acts, the Good Boss does not perform a wrong telegraphic act, for there is no better alternative available to her.

This may be enough to satisfy many, but some may think that all this talk of time slices misses the point: the real issue is whether the person acts rightly in performing an act. In particular, does the person act rightly in performing the unmaximizing act under the control of a disposition that is maximizing?

Given a metaphysics of temporal stages, then the right thing to say is that it is often a contextual matter which temporal stage's behaviour we concentrate on when we make a judgement of rightness. This explains the mixed feelings that we might have when making judgements about the rightness of acts where the acts are not maximizing, or where they are maximizing but we disapprove somehow. The point is that in all of these cases there are two temporal sequences to consider; and one acts rightly and

the other does not. Which one is relevant depends on what purposes we might have: correcting overall dispositions, or improving the strength of will at moments of deliberative choice. But where we have no particular purposes in mind, the question 'did she act rightly' becomes ambiguous. We are happy to answer yes when all stages that are relevant act rightly (as when the boss and the footsoldier act rightly) and we are happy to answer no when all act wrongly (as when the evil boss acts wrongly, and the footsoldier does not choose the least bad alternative) but our moral evaluations may be more confused or nuanced when the evaluations of different relevant time slices are out of synchrony. I take it that is a strength of my view that explains why these cases are disturbing.

6. Conclusion

So the judgements in cases that plausible kinds of indirect consequentialism may deliver can be reconciled with an act consequentialist theoretical framework, just so long as we take seriously the causal impact that the moral dispositions that we adopt have on our future capacity to deliberate.

This picture has more far reaching benefits than simply getting the cases right and making the theory neat. It also, I think, is a far more convincing account of moral life. Moral life is complex: sometimes it really is the case that agonizing calculation and decision has to be made. These are cases where we are relatively unconstrained by moral dispositions, and deliberate on a case-by-case basis. Unusual moral situations will be among the causes of such case-by-case deliberation. And if this is a very different kind of deliberation, then I think it an advantage of my account that it makes it so theoretically.

Weak moral dispositions are also a common feature of our ethical life. We act on a moral autopilot, but feel always able to step in and intervene if we notice that the case is special. When we follow the autopilot we get none of the phenomenology of being restricted in our choices by who we are.

But there is a widespread class of acts where we may get to choose, but from a constrained set of options. Some options are excluded because, it seems, of who we are, or because of our deep moral dispositions. Even if we see that the act we are about to perform doesn't have as good consequences as another physically in our power, we feel driven to it. This is satisfactory if the deep moral disposition remains one that we judge best. On my account we

still act rightly. That we are not free to choose outside this range is testified to by the fact that if we do decide that our disposition is not a maximizing one, we can't just immediately choose differently. It takes work on the self to reform, to change one's moral dispositions.

Acting rightly is doing the best we can do. That is the key insight of consequentialism. What I hope has been added here is the observation that it is our own past choices that influence what the best we can do is. We act freely within the bounds imposed by those choices. What once we may have been able to do we sometimes cannot, and for good reason. And what we now cannot do, we can sometimes come to be able to do, if that would be best.

References

- Adams, R. (1976). "Motive Utilitarianism." *The Journal of Philosophy* 73: 476-81.
- Braddon-Mitchell, D. (2003) 'Qualia and Analytical Conditionals' *The Journal of Philosophy* 100(3): 111-136.
- Driver, J. (2001). *Uneasy Virtue* Cambridge University Press, New York.
- Elster, J. (1979). *Ulysses and the Sirens: studies in rationality and irrationality*. Cambridge, Cambridge University Press.
- Frankfurt, H. (1971) 'Freedom of Will and the Concept of a Person' *The Journal of Philosophy* 68: 5-20, Dworkin, G. 1970 'Acting Freely' *Nous* 4:367-383.
- Gauthier, D. (1986). *Morals by Agreement*. New York, OUP.
- Jackson, F. (1991) 'Decision Theoretic Consequentialism and the Nearest and Dearest Objection' *Ethics* 101 (3): 461-482.
- Lyons, D. (1965). *The Forms and Limits of Utilitarianism*. Oxford, Clarendon Press.
- Mintoff, J. (1997). "Rational Cooperation, Intention and Reconsideration", *Ethics* 107: 612-643.
- Mintoff J. (2000). "Is Rational and Voluntary Constraint Possible?" *Dialogue* 39: 339-364.
- Pettit, P. and M. Smith (2000). Global Consequentialism. *Morality, Rules and Consequences*. B. Hooker, E. Mason and D. Miller. Edinburgh, Edinburgh University Press: 121-133.
- Pettit, P. and G. Brennan (1986). "Restrictive Consequentialism." *Australasian Journal of Philosophy* 64: 438-456

Sidgwick, H. (1930) *The Methods of Ethics* (reprinted 1981) Hackett Indianapolis.

Smart, J. J. C. (1973). An Outline of a System of Utilitarian Ethics. *Utilitarianism: For and Against*. J. Smart and B. Williams. Cambridge, CUP.

Stamp, D. and Gobson, M. (1992) 'Of One's Own Free Will' *Philosophy and Phenomenological Research* 52:529-556.

Williams, B. (1973). A Critique of Utilitarianism. *Utilitarianism: For and Against*. J. Smart and B. Williams. Cambridge, CUP

* Thanks to Justine Kingsbury, Michael McDermott, Jonathon McKuen-Green, Richard Joyce, Kristie Miller, Taha O'Leary, Philip Pettit and Caroline West for helpful discussion of this paper.

ⁱ Or satisfice. Throughout the paper I talk of maximizing, but there is a version of everything I say that takes satisficing as the required relation between value and evaluation or choice.

ⁱⁱ This way of explaining the distinction is due to Pettit, P. and M. Smith (2000). Global Consequentialism. *Morality, Rules and Consequences*. B. Hooker, E. Mason and D. Miller. Edinburgh, Edinburgh University Press: 121-133. They have called the usual versions of direct and indirect consequentialism *local* consequentialisms: local because each prioritizes one kind of entity – acts, or dispositions – and evaluates these consequentially, and then evaluates other kinds of entities according to whether they produce the maximizing version of the locally preferred kind of entity. Thus local act consequentialism evaluates acts consequentially, and judges dispositions rules or whatever by whether they produce the best acts. Local rule, disposition or character consequentialism evaluates the rule (disposition or character) consequentially, and evaluates the act according to whether it is produced by the right rule. In this paper I defend a version of act consequentialism, since the consequences of dispositions or rules are salient only in the context of acts that may affect the dispositions or rules that one instantiates.

ⁱⁱⁱ Williams, B. (1973). A Critique of Utilitarianism. *Utilitarianism: For and Against*. J. Smart and B. Williams. Cambridge, CUP.

^{iv} Smart, J. J. C. (1973). An Outline of a System of Utilitarian Ethics. In Smart and Williams *op cit* p10

^v Lyons, D. (1965). *The Forms and Limits of Utilitarianism*. Oxford, Clarendon Press. In fact I am unmoved by this objection. The benefit of the rule and character may be manifested in ways other than the acts it produces. If this is so, the best rule or character may well not be the one which produces the maximizing acts, if there is no way to get the good consequences of the rule or character without the non-maximizing acts.

^{vi} Perhaps this is the rule worship point of Smart.

^{vii} Pettit and Smith

^{viii} Driver, J. (2001). *Uneasy Virtue* Cambridge University Press, New York

^{ix} Adams, R. (1976). "Motive Utilitarianism." *The Journal of Philosophy* 73: 476-81.

^x Gauthier, D. (1986). *Morals by Agreement*. New York, OUP.

^{xi} Elster, J. (1979). *Ulysses and the Sirens: studies in rationality and irrationality*. Cambridge, Cambridge University Press.

^{xii} Nuclear deterrence and evolutionary game theory both present cases with a similar structure.

^{xiii} Gauthier *op cit*.

^{xiv} Mintoff, J (1997) "Rational Cooperation, Intention and Reconsideration", *Ethics* 107: 612-643. and Mintoff J (2000) "Is Rational and Voluntary Constraint Possible?" *Dialogue* 39: 339-364.

^{xv} Of course the very best moral disposition would be the one not to kill except in very extreme circumstances. These might be circumstances where the costs of not killing are so great, that it is possible to build a psychology which is generally repulsed by killing, but may yet be able to kill when the lives of so many depend on it. This means that someone who thinks that the wrong thing is done by someone who refuses to kill when it will save a nation can still complain that an agent who fails to do so has the wrong moral sensibilities.

^{xvi} Frank, RH (1989) *Passions within reason: the strategic role of the emotions* W.H. Norton.

^{xvii} Of course backwards induction arguments show that this is much sooner than might seem obvious

^{xviii} Weak moral dispositions—ones where the dispositions are overseen by a maximizing psychological overseer—correspond to the kind of

disposition that we would internalise if we were what Pettit and Brennan (Pettit, P. and G. Brennan (1986). "Restrictive Consequentialism." *Australasian Journal of Philosophy* 64: 438-456.) call virtual consequentialists: on that view the weak moral disposition is the motive for action, allowing us to behave with non-maximizing motives while the overseer ensures that we in fact do maximize. It is not part of my view that this never happens – it clearly does – but rather that strong moral dispositions are what is required in cases where the maximizing benefit of the dispositions can only be purchased at the expense of acting on some occasions in an unmaximizing way.

^{xix} Of course it is no part of my theory that this is the *conscious* justification of adoption of such dispositions: only that they are rightly adopted because of these factors (and these factors no doubt play some indirect role in the explanation of our propensity to form such dispositions).

^{xx} Frankfurt, H. (1971) 'Freedom of Will and the Concept of a Person' *The Journal of Philosophy* 68: 5-20, Dworkin, G. 1970 'Acting Freely' *Nous* 4:367-383.

^{xxi} Stamp, D. and Gobson, M. (1992) 'Of One's Own Free Will' *Philosophy and Phenomenological Research* 52:529-556.

^{xxii} Note that this perhaps controversially distinguishes between that part of our psychology that makes decisions about actions, and that part which reasons about the desirability or utility of actions. For we might be able to reason that a certain ruled out option better promoted the very values whose promotion justified the disposition to rule out the option in general, and yet not consider the ruled out option seriously in actually deciding what to do.

^{xxiii} Braddon-Mitchell, D. (2003) 'Qualia and Analytical Conditionals' *The Journal of Philosophy* 100(3): 111-136.

^{xxiv} Sidgwick, H. (1930) *The Methods of Ethics* (reprinted 1981) Hackett Indianapolis.

^{xxv} Clearly responsibility is not univocal: the concept of Ministerial responsibility in the Westminster system, for example, is one which might be defended on broadly consequentialist grounds, even though it is at odds with the internalist conception of responsibility I am discussing here.

^{xxvi} Adams *op cit*, Driver *op cit*.

^{xxvii} Or satisficing; Driver is not in fact committed to maximizing. For expository reasons throughout I will talk of maximizing, when in fact being

neutral between the maximising and satisficing versions of consequentialism is fine for the current purposes (though on other grounds I favour a maximizing version).

^{xxviii} Though I do not favour such a view, since it would tell us that the right disposition to have would be the best one, regardless of whether there was any way to achieve it. However I do not wish to take a stand on an in-house issue amongst binding consequentialists: the choice between what might be called binding global consequentialism, and binding act consequentialism. Binding global consequentialism would evaluate all dispositions consequentially, regardless of whether any acts of inculcation or retention are performed, and separately evaluate the acts produced by them. Binding act consequentialism evaluates things only insofar as they are subjects of choice and thus there are acts to consider: and some of the acts will be acts of disposition choice. But both kinds of binding consequentialism will have the crucial feature of harmony between the evaluation of general acts, and the evaluations of the dispositions or disposition choosing acts.