

Doing the Best We Can't

Evaluative Conflict and the 'Ought' Implies 'Can' Principle

Daniel Parmeggiani Gitlesen



Thesis presented for the degree of

MASTER OF PHILOSOPHY

Supervised by Ole Martin Moen

Department of Philosophy, Classics, History of Art and Ideas

Faculty of Humanities

University of Oslo

Spring 2016

One must perform the lower act which one can manage and sustain: not the higher act which one bungles. . . . Self-knowledge will lead us to avoid occasions of temptation rather than rely on naked strength to overcome them. We must not arrogate to ourselves actions which belong to those whose spiritual vision is higher or other than ours.

—Iris Murdoch, *The Bell*, 193, 195.

For if the moral law commands that we *ought* to be better human beings now, it inescapably follows that we must be *capable* of being better human beings.

—Immanuel Kant, *Religion Within the Boundaries of Mere Reason*, 94.

© Daniel Parmeggiani Gitlesen

2016

Doing the Best We Can't: Evaluative Conflict and the 'Ought' Implies 'Can' Principle

<http://www.duo.uio.no/>

Trykk: Allkopi

IV

ABSTRACT

The aim of this thesis is to defend *global consequentialism* from its main objection, specifically the objection that it allows *evaluative conflict*.

Global consequentialism differs from traditional forms of consequentialism in that it does not only focus on one type of thing, like acts or rules. Act consequentialism focuses on the right acts directly, and evaluates rules indirectly according to whether they lead to the right acts or not. Rule consequentialism focuses on what the right rules are, and evaluates acts indirectly by appealing to whether they conform to the right rules. Global consequentialism will rather evaluate *any* evaluand directly in terms of its consequences, whether it is an act, a set of rules, a law, a character trait, etc.

But what should we say if having the best motives implies that you *cannot* do the best act. That is, if your motives are so strong they make it causally impossible for you to do the otherwise best act? This would amount to evaluative conflict, and it can supposedly happen between any set of evaluands. To put it colloquially, evaluative conflict means you are “damned if you do, and damned if you don’t”.

I argue that evaluative conflict cannot occur since the principle ‘ought’ implies ‘can’ includes evaluands that are pragmatically impossible in conjunction with each other. If you cannot have the best motives *and* do the best action, then it will not be the case that you ought to have the best motives *and* do the best action.

Derek Parfit argues that this defense would render wrongness obsolete. If determinism is true then all acts are determined by motives, and so no acts would be wrong. I argue that his solution also fails via a distinction between objective and subjective rightness. For subjective rightness his argument results in the counter-intuitive claim that we *ought* to do something we *know* that we cannot do. For objective rightness his argument leads to wrongness becoming obsolete for other theories as well, or it merely shifts wrongness to the level of motives as opposed to acts.

I think that the best way of escaping these problems is to reject objective rightness. This in no way means we have to give up on a robust a notion of right and wrong, since subjective and objective rightness relates to our epistemic situation in relation to an objectively true morality, not a rejection of objectively true morality itself.

ACKNOWLEDGEMENTS

Several people have helped me, either by answering questions about their work, or by giving me invaluable feedback. A large gratitude is owed to Ole Martin Moen for being my excellent supervisor. Many of the ideas that I explore in this thesis were developed during my time at Rutgers University, especially during conversations with Holly Smith, but also during my classes with Derek Parfit. Both of them have also been so gracious as to respond to my subsequent queries about their work via e-mail, as has Elinor Mason, and Shelly Kagan. A great gratitude is also owed to Ainar Petersen Miyata, Conrad Bakka, and Mariona Eiren Bohlin Sturm for giving me feedback on the finished thesis. And, thank you to all of the great philosophers who have written so insightfully about the subjects that I discuss.

TABLE OF CONTENTS

1 THE PROBLEM	1
1.1 Preliminaries.....	1
1.2 Consequentialism in Context	3
1.2.1 The Three Leading Traditions.....	3
1.2.2 Subjective and Objective Rightness.....	7
1.2.3 Evaluative Focal Points.....	10
1.3 The Motivation Behind Global Consequentialism.....	12
1.3.1 Foundational Consequentialism.....	12
1.3.2 The Worship Objection	13
1.4 Global Consequentialism	19
1.5 Evaluative Conflict.....	24
1.6 Recap and Roadmap.....	28
2 IS RIGHTNESS TRANSFERABLE?.....	29
2.1 Parfit and Louise Versus Mason	29
2.2 How Transferring of Rightness Works	33
2.3 A ‘Transfer’ of Rightness?.....	37
2.4 Intermediate Conclusion	38
3 ANEWCRITERIONOFROLE-BASEDOBLIGATION.....	39
3.1 The Criterion	39
3.2 Is This Still a Form of Consequentialism?.....	43
3.3 Is This Still a Form of <i>Foundational</i> Consequentialism?	44
4 I MUST, BUT I OUGHT NOT	47
4.1 Parfit and OIC	47
4.2 The First Reason why Claim (1) is Not Claim (3).....	50
4.3 Is Parfit an Ethical Possibilist?.....	53
4.4 OIC and Subjective Rightness.....	55
4.5 The Second Reason why Claim (1) is Not Claim (3).....	61
4.6 Indeterminism I: General Psychological Indeterminism.....	69
4.7 Indeterminism II: Local Determinism.....	74
4.8 Global Determinism: the Hard Case	77
4.8.1 The Rejection of Objective Rightness.....	80
4.8.2 OIC Does Not Apply to Objective Rightness	84
4.9 Psychological Determinism <i>Without</i> Physical Determinism	85
4.10 Rejecting ‘Ought’ Implies ‘Can’ Outright	89
4.11 Recap.....	93
5 UNIVERSALIZING MASON’S DEFENSE.....	95
5.1 Decision Procedures and Acts.....	95
5.2 Rules and Acts.....	96
6 CONCLUSION	99
BIBLIOGRAPHY	103

1 THE PROBLEM

1.1 Preliminaries

The aim of this thesis is to defend global consequentialism from the objection that it allows for evaluative conflict. In order to do so, I should explain what evaluative conflict is and why it can be seen as a problem for global consequentialism. But first I shall provide some context, as well as explain some terminology that will be important in order to understand what evaluative conflict is.

In chapter one, I will start by outlining the more traditional variants of consequentialism. Then I will define two concepts and explain one distinction. The distinction I wish to explain is the one between *objective rightness*, and *subjective rightness*. The two concepts are *evaluative focal points* and *foundational consequentialism*. Once these concepts are clear I will show that the traditional forms of consequentialism are open to a strong counter-argument I call the *worship argument*. This section also serves to show how global consequentialism avoids the objection.¹ In section 1.4 I give an outline of what global consequentialism is and how it works. Then, finally, in section 1.5 I will present the evaluative conflict objection, which has been seen as particularly problematic for global consequentialism.

The core aim of chapter one is to arrive at what, exactly, evaluative conflict is, but it also tries to show what is at stake if global consequentialism fails. Very roughly, evaluative conflict can be described as a specific kind of case where one ought to x , and one ought to y , but one cannot both x and y , for example if one ought to have the best motives and do the best act, but one cannot do the best act with the best motives. If global consequentialism allows for this, it seems to demand more than what is possible of the agent. One way to describe the core intuition behind consequentialism is to say that it boils down to one single obligation; that we

¹ The argument is not of my making though. Philip Pettit and Michael Smith, "Global Consequentialism," in *Morality, Rules, and Consequences: A Critical Reader*, ed. Brad Hooker, Elinor Mason, and Dale E. Miller (Edinburgh: Edinburgh University Press, 2000), 124–130; and Shelly Kagan, "Evaluative Focal Points," in *Morality, Rules, and Consequences*, ed. Brad Hooker, Elinor Mason, and Dale E. Miller (Edinburgh: Edinburgh University Press, 2000), 135.

always ought to do the best we can.² But if global consequentialism allows for evaluative conflict, it appears like it demands that we ought to be doing *better* than we can. In other words, that we ought to be *doing the best we can't*.

In chapter two I will sketch the outline of a global consequentialist reply, inspired by a paper written by Elinor Mason. In the crudest terms possible, the reply is that global consequentialism is just as entitled as any other theory to appeal to the 'ought' implies 'can' (OIC) principle, which states it is never the case that you ought to do something you cannot do, and that this principle typically includes impossible conjunctions of acts and other evaluands. Also, if global consequentialism is interpreted so as *not* to demand the impossible, we should remember that consequentialism generally places very high demands on a moral agent. So, as long as the demands are merely the highest possible demands, and not impossible demands, consequentialists should remain untroubled by the demands of global consequentialism.

In chapter three I simply refine the characterization of global consequentialism that I gave in the first chapter, in order to make sure that it coheres with the global consequentialist reply to evaluative conflict.

Chapter four contains the most important parts of the thesis. Derek Parfit has made an argument which appears to effectively block the reply I provide in chapter two. His argument is roughly that the interpretation of OIC I need in order to mount my defense leads to absurd consequences. First of all he argues that the defense presupposes psychological determinism, and if so it follows that wrongness becomes obsolete because there will always be a single possible act. If you ought to do the best possible act, and there is never more than one option, then there will never be wrong acts.

This is a seemingly devastating objection, but he has a solution to this problem. However, this solution leads to evaluative conflict for global consequentialists. I show that his argument has a larger scope than global consequentialism, so that it effectively makes wrongness obsolete no matter what theory one ascribes to, and that it may lead to even more counterintuitive claims than wrongness being obsolete, depending on what kind of rightness we are talking about. If so it seems his argument has some rather troubling and quite general implications. I then attempt to provide a solution as best as I can. Even though I do not find

² Fred Feldman, introduction to *Doing the Best We Can: An Essay in Informal Deontic Logic* (Dordrecht: D. Reidel Publishing Company, 1986), ix–xi.

one that is *entirely* satisfying, I hope I succeed in showing that there are some possible routes out of the quagmire that Parfit's argument wades us into. Roughly, the answer I give is that we can retain one type of rightness, namely *subjective* rightness, but may have to give up on *objective* rightness. This in no way commits us to a relativistic view of morality, since the term 'subjective rightness' has to do with our epistemic relation to an ethical theory; it is not a claim about the truth of ethical theories as such.

Even though global consequentialism was introduced partly in order to make it possible for consequentialists to directly evaluate the rightness of motives, character traits, laws, rules, etc. I mostly focus on acts and motives, so in the fifth and final chapter I attempt to show that the answers I provide in the earlier chapters also apply to other evaluable entities.

There are many practical implications of this view, but perhaps the most interesting one is that it claims that, sometimes, not being a good enough person or not having good enough motives can mean that you do the right thing, even though a better person might have a duty to do something better. One example could be a psychopath, it has been claimed that they are unable to be motivated by moral considerations, and if so my view has the implication that they not only have an excuse for their behavior, but that they never actually do anything morally wrong.³ It is the view of a so called "bleeding heart liberal", and allows a certain degree of redemption for those of us who are morally imperfect. On the other hand, I do not claim to have an answer to how many cases there are where it is truly *impossible* to do some act because of the motive set you have. It is entirely possible that this bar is set quite high. Still, as we shall see, it seems the possibility of evaluative conflict rests on the assumption that motives *can* determine acts.

1.2 Consequentialism in Context

1.2.1 The Three Leading Traditions

Although there are many types of ethical theories, the three most widespread traditions of normative ethics are probably; *consequentialism*, *deontology* and *virtue ethics*, and there is one key distinction between consequentialism and the other two. Consequentialists believe that all that ultimately matters when attributing moral properties are consequences.⁴ That is,

³ Jonny Anomaly, "An Argument against External Reasons," *Sorites* 18 (2007): 56–59.

⁴ Rightness, wrongness, permissibility, impermissibility and supererogation are common moral properties. If an act is supererogatory, it means it is above and beyond the call of duty. Consequentialists generally don't believe this moral property exists. If the best action is obligatory, then there are no supererogatory actions. If there are

they reject the notion that any further constraints apply when attributing a moral property to (for example) an act. Any theory that fulfills this criterion counts as a consequentialist theory, so the mere fact that a theory falls under this rubric does not say very much. It does not say anything about—among other things—what is valuable, which consequences matter, or even whether value is to be maximized.

Deontologists and virtue ethicists do believe that we have to accept additional constraints when attributing moral properties to something (an act, a motive, a character trait, etc.). Specifically, they often accept constraints like acting from the right motives or fulfilling specific generalized duties.

The point is that almost all ethical theories acknowledge that consequences matter, but consequentialism stands out by claiming that the consequences are *all* that matters for the attribution of moral properties. I am going to assume that some form of consequentialism is true for the remainder of this thesis.

There are more consequentialist theories than you can shake a stick at, and they differ on whose consequences are to be taken into account (everybody's, only the agent's, etc.), what has value, what we should count the consequences of (acts, motives, rules, everything, etc.), whether it is actual consequences or expected consequences that determine rightness, whether good consequences should be maximized, and much more. Still, the core elements of a consequentialist theory usually consist of two parts. The first part tells you (roughly) that x is right if and only if x is *best*, and the second part tells you what the good is (along with the best, worst, worse, better, etc.). That is, the first part tells you something about how to assign moral properties, which in turn tells you something about your moral *obligations*. Consequentialists usually say that if there is a single best option it is obligatory, and if there are several equally optimal options they are all permissible, whilst performing *one of* the best options is obligatory. The second part is a *value* claim; it tells you something about what has intrinsic value.

several equally good choices they are all permissible, yet no one act is *beyond* the call of duty, it is simply obligatory to perform one of the acts. One notable exception is *satisficing consequentialism* which rejects the notion that maximization is necessary for rightness. That is, they think that rightness relies only on the consequences, but less than maximal production of good might qualify as *good enough*, and so as right. Also, 'consequences' is to be defined loosely. If you are evaluating an act, for example, the act itself may count as one of the consequences. The act of wallowing in misery for example could in itself include suffering, and that suffering is just as relevant as any subsequent misery the wallowing results in.

As an example, a good start for a characterization of one kind of act utilitarianism, which is a sub-category of consequentialism, could be:

Normative claim: “An act is right *if and only if* it will lead to at least as much good as any alternative act.”⁵

Value claim: Pleasure is the only thing that has intrinsic positive value. Pain is the only thing that has negative value.⁶

This is just a first approximation. While this characterization allows for ties between acts that will lead to equal amounts of good, it is not clear about what is meant by ‘will lead to’, amongst other things. Does this mean ‘will *actually* lead to’ or ‘will *probably* lead to’? But, my purpose is just to provide a rough template of the underlying logical structure of consequentialist theories. If one were to provide a full definition of a consequentialist view this two-part characterization would have to be filled out and adjusted.

The two most prominent kinds of consequentialism are *utilitarianism* and *ethical egoism*, and they differ primarily in *whose* good is to be taken into account. Utilitarians usually take the consequences for all beings capable of having positive and negative experiences into account, whilst ethical egoists only count the consequences that are in the self-interest of the agent.⁷ In addition to this, there are related disputes over whether future generations are to be taken into account, and if so how.⁸ I will not be concerned with ethical egoism in this thesis, and I will remain agnostic when it comes to future generations.

Another important dispute is over what has intrinsic value. Consequentialists usually accept some sort of distinction between intrinsic and extrinsic good, but they disagree on what is intrinsically good. Intrinsic good is good *in itself* or *as such*, extrinsic good is good that is not intrinsic.⁹ Here are three examples of common conceptions of intrinsic good:

⁵ Toby Ord, *Beyond Action*, (Ph.D. diss., University of Oxford, 2009), 1.

⁶ Pleasure and pain may be defined in a range of ways; I will not take a stand on this issue.

⁷ In theory you could make any division you wish. You could be a consequentialist *ethical altruist*, someone who only takes into account the consequences for everyone *but* the agent herself, and a consequentialist Nazi might (woe be the day) only take Aryans into account, but I am not aware of any theories like these.

⁸ Tim Mulgan, *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations* (Oxford: Oxford University Press, 2006), 16–21.

⁹ Michael J. Zimmerman, "Intrinsic vs. Extrinsic Value," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, (Spring 2015 Edition,) §2, §6, <http://plato.stanford.edu/archives/spr2015/entries/value-intrinsic-extrinsic/>.

Hedonism: Pleasure is the only thing that has intrinsic positive value. Pain is the only thing that has negative value.¹⁰

Desire satisfaction: The satisfaction of rational and informed desires is the only good.¹¹

Objective list theories: There is some objective list of things that are good. Likely entries on such a list are: happiness, desire satisfaction, knowledge, love, etc.¹²

In spite of listing its conception of good in terms of both positive (pleasure) and negative (pain), hedonism is often considered *value monist*, they believe that the value of any state of affairs has to land somewhere on a single spectrum going from very bad to very good. Desire satisfaction theories are also mostly considered value monist since they claim that only one type of thing has value. Objective list theories on the other hand are value pluralist; they take several different things to have intrinsic value.

A closely related, yet separate question revolves around whether the happiness that comes from reading *Hamlet* counts more than a quantitatively equal amount of the happiness that comes from reading *Fifty Shades of Grey*. That is, we can ask whether there are *qualitative* differences as well as *quantitative* differences. Mill famously held that there are such qualitative differences, while Bentham made no such distinction.¹³

There are further distinctions and controversies, but I will not get into these problems. I will assume that the good can be summed up *somehow* though. That is, I will assume that arithmetic applies to the good just as much as it does to atoms, ounces or apples.¹⁴ This means that if we take both knowledge and pleasure to have value, there will be some definitive amount of knowledge, specifically described, that is twice as good as a specific amount of pleasure.

Utilitarianism is the most common form of consequentialism. Because of its initial intuitive appeal, many non-consequentialists have thought of utilitarianism as a foil for other

¹⁰ Roger Crisp, "Well-Being," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, (Forthcoming), §4, <http://plato.stanford.edu/archives/sum2016/entries/well-being/>.

¹¹ Ibid.

¹² Ibid.

¹³ John Stuart Mill, *Utilitarianism*, ed. Roger Crisp (New York: Oxford University Press, 1998), 55–56.

¹⁴ Henry R. West, "Mill's 'Proof' of the Principle of Utility," in *An Introduction to Mill's Utilitarian Ethics* (Cambridge: Cambridge University Press, 2004), 141.

theories. Philippa Foot writes that “Utilitarianism tends to haunt even those of us who will not believe in it”, and Will Kymlicka writes that “in our society utilitarianism operates as a kind of tacit background assumption against which other theories must assert and defend themselves”.¹⁵ T. M. Scanlon characterizes act-utilitarianism as “the view toward which [ethicists] find themselves pressed when they try to give a theoretical account of their moral beliefs. Within moral philosophy it represents a position one must struggle against if one wishes to avoid it”.¹⁶ This does not necessarily mean that it is true; none of these three philosophers are utilitarians. But it serves to show that the underlying framework of consequentialism (in this case utilitarianism) has an appeal that reaches further than the theories themselves have.

1.2.2 Subjective and Objective Rightness.

Up until this point I have mainly been giving some basic background information and made it clear that some of the more famous distinctions within consequentialism are not going to play a large part in what follows. A distinction that I will be returning to, however, is the distinction between *subjective rightness*, *objective rightness* (as well as wrongness, obligatoriness, permissibility, etc.). This distinction can be most readily explained by an example borrowed from Holly Smith:

Strong Medicine: Patient Ron consults his physician, Sue, about a moderately serious ailment. Sue can treat Ron with either of two drugs. She believes that giving him no treatment would render his ailment permanent; that drug *X* would cure Ron partially; and that there is an 80 percent chance that drug *Y* would cure Ron completely, but a 20 percent chance that *Y* will kill him.¹⁷

The obvious choice here is to give the Ron drug *x*, since it is only a moderately serious illness which drug *x* can partially cure, while drug *y* runs a significant risk of killing him. That is, the *expected value* of administering drug *x* is higher than any of the other alternatives. To explain expected value, let’s assume a set of values for each outcome:

The value of Ron continuing to be ill = -500

The value of Ron being partially cured (by drug *x*) = 100

The value of Ron being fully cured (by drug *y*) = 1000

¹⁵ Eggleston, Ben and Dale E. Miller, Introduction to *The Cambridge Companion to Utilitarianism*, ed. Ben Eggleston and Dale E. Miller (New York: Cambridge University Press, 2014), 2; and Philippa Foot, “Utilitarianism and the Virtues,” *Proceedings and Addresses of the American Philosophical Association* 57, no. 2 (November 1983): 273.

¹⁶ Eggleston and Miller, *The Cambridge Companion to Utilitarianism*, 2.

¹⁷ Holly M. Smith, “Subjective Rightness,” *Social Philosophy and Policy* 27, no. 2 (2010), 77.

The value of Ron dying (from drug y) = -25000

Let us also stipulate that these are the only numbers that matter in this case. Doing nothing, then, has only one possible outcome, which means that the expected value just is -500, the same goes for giving Ron drug x which will partially cure him yielding an expected value of 100. Giving Ron drug y on the other hand has two possible outcomes, so you have to multiply each by its probability and then sum the two possibilities together.

Ron has a .8 chance to be fully cured by drug x, which has a value of 1000.

$$1000 \times .8 = 800$$

Ron has a .2 chance of being killed by drug x, which has a value of -25000.

$$-25000 \times .2 = -5000$$

The expected value of giving Ron drug x will be the sum of the value of these two possible outcomes.

$$-5000 + 800 = -4200$$

In other words, it seems you should you should administer Ron drug x, since it has the highest expected value of the three possibilities of doing nothing (-500), administering drug y (-4200) and administering drug x (100). (This is only meant to illustrate, most consequentialist will not actually reason quite so crudely, but this is the basic idea.)

But imagine if, a few years later, a test is discovered which tells you what would have happened to Ron if he had taken drug y, and it turns out that Ron is one of the 80 percent that would have been completely cured. Then there seems to be a clear sense in which giving Ron drug y would have been the right thing for Sue to do.

To dissolve this problem philosophers have claimed that there are (at least) two separate senses of rightness.¹⁸ Somewhat loosely we can say that an action is subjectively right if it is “the action that is morally more appropriate in light of the agent’s beliefs about those circumstances, even if the beliefs turn out to be wrong” and objectively right if it is “the morally best action in the actual circumstances”.¹⁹ It would in fact have been best to give Ron drug y, as it would actually maximize value, so in an objective sense that would have been the

¹⁸ Holly M. Smith, “Subjective Rightness,” 77.

¹⁹ *Ibid.*, 65.

right thing to do. This holds even though you could argue that Sue would be strongly *blameworthy* for doing the objectively right thing at the time, as it is also true that the expected value of giving drug *x* was much lower than drug *y*.²⁰ Sue was subjectively right in giving drug *y*, but objectively wrong.

In fact though, there are two versions of subjective rightness. First there is *pure subjectivism*, which generates obligations by appealing to the agent's opinion of what the moral facts are, as well as the non-moral facts.²¹ This roughly equates to the view that what is right is what the agent herself thinks is right, no matter how crazy. If *S* thinks that she should do whatever the Illuminati tell her to do, then she is obligated to do so. I take this version of subjectivism to be severely flawed, so whenever I mention subjective rightness from now on I will rather be referring to a kind of *theory-relative subjectivism* about rightness. This defines subjective rightness through some objective moral theory, but also the agent's beliefs about what would fulfill the criterion that theory has for rightness. Sometimes this may be very easy, but other times it may be much harder. Sue, for example, is unable to do what would have the best consequences because she does not know that drug *x* would fully cure Ron. One could complicate this further, but this sketch should suffice for the purposes of my thesis except to say this; there is a further kind of rightness that has been proposed in the literature, namely *prospective rightness*. It occupies something like the middle ground between subjective and objective rightness. But I think that adding this sense of rightness would mainly complicate matters without contributing any greater clarity, especially since there seems to be some significant differences in the literature concerning how to define prospective rightness.²² I will be assuming that one of the arguments I shall level at OIC according to either subjectivism or objectivism about rightness can be levelled at the prospective sense as well, and I will take consequentialism as the theory that subjective rightness is supposed to relate to for the remainder of the thesis.

This distinction is not exclusive to consequentialism. In fact, it was initially noted by philosophers from several camps, most notably Bertrand Russell, C.D. Broad, H.A. Prichard,

²⁰ Holly Smith, "Culpable Ignorance," *The Philosophical Review* 92, no. 4 (October 1983): 543. Roughly, you are blameworthy if you deserve to be blamed. It is possible that you do something objectively wrong for which you have an excuse, for example, in which case it may still be wrong, but you are not blameworthy.

²¹ Elinor Mason, "Objectivism, Subjectivism, and Prospectivism," in *The Cambridge Companion to Utilitarianism*, ed. Ben Eggleston and Dale E. Miller (New York: Cambridge University Press, 2014), 179–180.

²² *Ibid.*, 179–180; and Holly M. Smith, "The "Prospective" View of Obligation," *Journal of Ethical and Social Philosophy*, Discussion Note, (February 2011): 1.

and W.D. Ross.²³ It springs from ignorance or uncertainty about what is right according to a theory, either because one does not know what would fulfill the criterion the theory proposes, in *Strong Medicine* the uncertainty about what the consequences of your actions will be, but it could be ignorance about what you have promised, or even ignorance about what the criterion of a theory really is.

Let's say you are a deontologist, and that you have promised your wife to get some milk from the store, but when you get there you cannot remember what you were supposed to get. If you think there's about an 80% chance that it's orange juice, and a 20% percent chance that it's milk, then *subjectively* you should bring orange juice, but *objectively* you should bring milk.

Accepting these senses of rightness dissolves the paradoxical nature of the *strong medicine* example, in which the same act is apparently both right and wrong, by showing that while it is both right and wrong, it is so in two different senses. I will be employing this distinction quite a bit in my main arguments.

1.2.3 Evaluative Focal Points

One important disagreement amongst consequentialists concerns what, exactly, should be evaluated morally. *Classical consequentialism*, which can be loosely demarcated as the consequentialism of Bentham, Mill, and Sidgwick, was primarily concerned with evaluating acts, though they also made some utilitarian claims that pertained to character and motives.²⁴ But since its heyday we have seen a proliferation of different kinds of consequentialisms that either focus on something besides acts, or focus on something in addition to acts. *Rule utilitarianism* is the most familiar such theory, but *motive utilitarianism*, *virtue utilitarianism*, and *world utilitarianism* have also attracted a good deal of attention from leading philosophers.²⁵

²³ Bertrand Russell, "The Elements of Ethics," in *Readings in Ethical Theory*, ed. Wilfrid Sellars and John Hospers, 2nd ed. (New York: Appleton-Century-Crofts, 1970), 10–15, Originally published in *Philosophical Essays* (London: Longmans, Green & Co., 1910); C.D. Broad, *Ethics*, ed. C. Lewy (Dordrecht: Martinus Nijhoff, 1985), 124–194; H.A. Prichard, "Duty and Ignorance of Fact," in *Moral Obligation and Duty and Interest* (Oxford: Oxford University Press, 1968), 18–39; and W.D. Ross, *Foundations of Ethics* (Oxford: Clarendon Press, 1939), chapter 7.

²⁴ Toby Ord, "How to Be a Consequentialist about Everything," lecture presented at ISUS X (2008): 1–3, <http://www.amirrorclear.net/academic/papers/everything.pdf>.

²⁵ Brad Hooker, "Rule Consequentialism," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, (winter 2015 edition,) § 1–§ 5, <http://plato.stanford.edu/archives/win2015/entries/consequentialism-rule/>; Fred Feldman, "World Utilitarianism," in *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*

As we shall see, the reasons that led philosophers to adopt these new forms of consequentialism are parallel to the reasons given for going one step further to global consequentialism, which evaluates *all* evaluands directly in terms of the good. That is it focuses on acts *and* motives *and* rules *and* everything else. In order to explain this more fully we need to examine a concept taken from Shelly Kagan, the *evaluative focal point* (or simply focal point).²⁶

The gist of it is that we can evaluate (either in terms of rightness and wrongness or in terms of value) many different things: acts, motives, sets of motives, intentions, norms, character traits, decision procedures, institutions, dispositions, lives, states of affairs, etc.²⁷ These are all examples of evaluative focal points. Just as you can ask what act you ought to do in some specific situation, you can ask which motives you ought to have, or how you ought to decide what to do.

One way of explaining evaluative focal points is to say that they are whatever is evaluated from the point of view of some ethical theory. For act consequentialism, acts are the primary evaluative focal point. All other focal points, like motives or rules, are evaluated indirectly in terms of whether they lead to the right acts. For rule utilitarians, rules are the primary evaluative focal point. All other evaluative focal points are evaluated indirectly in terms of whether they relate to the right rules in the right way.²⁸ It is natural to think of evaluative focal points as *types* of things such as the *category* of acts or the *category* of motives, but this is not strictly necessary. The idea is that whatever receives *direct* evaluation by a theory is its primary evaluative focal point.

Another way of explaining evaluative focal points is by looking at the difference between what has been labeled *direct* and *indirect* types of consequentialist theories. Act-utilitarianism has often been labeled as *direct consequentialism*, since it focuses directly on acts, and evaluates them directly in terms of the good. Rule utilitarianism on the other hand

(Cambridge: Cambridge University Press, 1997), 17–36; R.M. Adams, “Motive Utilitarianism,” *Journal of Philosophy* 73, no. 14 (1976): 467–481; and Roger Crisp, “Utilitarianism and the Life of Virtue,” *The Philosophical Quarterly* 42, no. 167 (April 1992): 139–160.

²⁶ Shelly Kagan, “Evaluative Focal Points,” 134–155.

²⁷ Derek Parfit, *Reasons and Persons*, 24; Jennie Louise, “Right Motive, Wrong Action: Direct Consequentialism and Evaluative Conflict,” *Ethical Theory and Moral Practice* 9, no. 1 (2006): 65; and Shelly Kagan, “Evaluative Focal Points,” 134.

²⁸ Shelly Kagan, “Evaluative Focal Points,” 135; and Shelly Kagan, e-mail message to author, Jan 24, 2015.

has been characterized as *indirect utilitarianism* since it evaluates acts only indirectly by appealing to how they relate to the right set of rules.²⁹

As Kagan points out, however, act consequentialism is only direct with regard to acts; it is indirect with regard to rules and other evaluative focal points. For act consequentialists, the right rules to have in a society are the ones that lead to the right acts. Rule utilitarianism evaluates rules directly in terms of their consequences, so it is direct when it comes to rules.³⁰ (But, since rules are not concrete entities, they cannot have consequences unless they are followed, accepted, flouted or otherwise embedded in a society).

In other words, calling act consequentialism direct consequentialism is somewhat misleading; it can only be seen as legitimate as long as we accept that acts are somehow a privileged evaluative focal point. Rule consequentialists deny that acts are privileged; global consequentialists also deny this, and they also reject the notion that rules or any other evaluative focal point is privileged. From the point of view of global consequentialism, the notion of a primary evaluative focal point is lost, since global consequentialism does not privilege any particular focal point; it rather evaluates all focal points directly. Since it is very awkward to call all evaluative focal points primary, global consequentialists dispense with the notion of primary evaluative focal points.

1.3 The Motivation Behind Global Consequentialism

1.3.1 Foundational Consequentialism

Consequentialism can be *foundational* or *non-foundational*, and an example of a non-foundational consequentialist theory is *satisficing consequentialism*. It admits that *only* consequences matter for the attribution of moral properties, making it a consequentialist theory, yet it also claims that maximization is *not* necessary for optimization. That is, you can do the right thing without doing the best thing. Foundational consequentialists think that a *maximally* good outcome is what matters for the attribution of moral properties. According to Parfit it is the claim that “there is one ultimate moral aim: that outcomes be as good as possible.”³¹ Or, in the words of Douglas Portmore:

Foundational consequentialists hold that the ranking of outcomes (or possible worlds) in terms of their goodness is at the foundation of all moral assessment. They hold that moral

²⁹ Shelly Kagan, “Evaluative Focal Points,” 135.

³⁰ Ibid.

³¹ Derek Parfit, *Reasons and Persons*, 24–25.

assessments of right and wrong acts, good and bad rules, virtuous and vicious character traits, etc. are all ultimately a function of how outcomes rank.³²

Act consequentialism is (at least seemingly) foundational, then, since it claims that you ought to do the best thing you can do, but as we shall see, it is far from clear that even act consequentialism counts as a foundationally consequentialist theory, since it is open to the objection that it is guilty of *act worship*.

1.3.2 The Worship Objection

What, then, is the worship objection? At heart it is the accusation that some theory fails to be foundationally consequentialist. That is, it in some way privileges some evaluative focal point at the expense of the maximization of value. For non-consequentialists this might not seem particularly strange, but for consequentialists it is a problem, at least when it is not separately argued for. The basic idea is that a theory is guilty of worshipping an act, a motive, a rule, a decision procedure, etc., if it tells you that it is right because it is the best act, rule, motive, etc., but also that *not* doing the act, having the motive, or following the decision procedure would have better consequences. But how could this happen?

Rule utilitarianism, for example, has been accused of *rule-worship*, as it tells us to obey a rule even if it will not maximize value. J.J.C. Smart was the first philosopher to develop the objection.

Suppose that there is a rule *R* and that in 99% of cases the best possible results are obtained by acting in accordance with *R*. Then clearly *R* is a useful rule of thumb; if we have not time or are not impartial enough to assess the consequences of an action it is an extremely good bet that the thing to do is to act in accordance with *R*. But is it not monstrous to suppose that if we *have* worked out the consequences and if we have perfect faith in the impartiality of our calculations, and if we *know* that in this instance to break *R* will have better results than to keep it, we should nevertheless obey the rule? Is it not to erect *R* into a sort of idol if we keep it when breaking it will prevent, say, some avoidable misery? Is not this a form of superstitious rule-worship?³³

What's so special about this rule? If it is better not to follow the rule, then why should a consequentialist still follow it? The point is not that we cannot generally follow rules in order

³² Douglas W. Portmore, "Foundational Consequentialism and its Primary Evaluative Focal Point," (working paper, accessed June 28, 2014), 1, <https://dl.dropboxusercontent.com/u/14740340/Foundational%20Consequentialism.pdf> .

³³ J. J. C. Smart, "Extreme and Restricted Utilitarianism," *The Philosophical Quarterly* 6, no. 25 (October 1956): 348–9.

to decide what to do, or that following a rule cannot sometimes be better overall than not doing so. The point is that *if* there is an *exception* to what is generally true—which is that following the rule will be better—*then* it seems strange for a consequentialist to stubbornly continue to follow the rule. But rule utilitarianism is not the only form of consequentialism that is open to this objection.

Motive utilitarianism was proposed by R. M. Adams explicitly because motives can affect value independently of the value that comes from the actions themselves. That is, because he thinks that act utilitarianism sometimes fails to maximize value because it worships acts. In order to make his point he presented the following example:

Chartres: “Jack is a lover of art who is visiting the cathedral at Chartres. . . . He is so excited that he is spending much more time at Chartres than he had planned . . . In fact, he is spending too much time there, from a utilitarian point of view . . . and this is going to cause him considerable inconvenience and unpleasantness. He will . . . have to do several hours of night driving, which he hates. . . . On the whole, he will count the day well spent, but some of the time spent in the cathedral will not produce as much utility as would have been produced by departing that much earlier. . . . Jack is studying the sixteenth to eighteenth century sculpture on the stone choir screen. . . . It is not completely unrewarding, but he would have more happiness on balance if he passed by these carvings and saved the time for an earlier departure. Jack knows all this . . . [but he] goes on looking at the choir screen because he is more strongly interested in seeing, as nearly as possible, everything in the cathedral than in maximizing utility. This action of his is therefore wrong by act-utilitarian standards, and in some measure intentionally so.”³⁴

So, from an act consequentialist perspective, Jack is acting wrongly in staying as long as he does, and he knows this. But his motives are such that he wants to see everything more than he wants to maximize happiness.

Jack would not have omitted these things unless he had been less interested in seeing everything in the cathedral than in maximizing utility. And it is plausible to suppose that if his motivation had been different in that respect, he would have enjoyed the cathedral much less. It may very well be that his caring more about seeing the cathedral than about maximizing utility has augmented utility, through enhancing his enjoyment, by more than it has diminished utility through leading him to spend too much time at Chartres.³⁵

³⁴ R.M. Adams, “Motive Utilitarianism,” 470–471.

³⁵ *Ibid.*, 471.

The point is that even though Jack is acting wrongly according to act utilitarianism, since the act of staying so long does not maximize happiness, it is better that Jack is as motivated to see Chartres as he is, since this set of motives will create more utility in sum than the set of motives that would make Jack do the right thing. This is a version of the worship objection and it points to the fact that act consequentialism prescribes doing the best act, *even though it will not maximize utility* in this case, since it follows from doing the best act that you have inferior motives, and less value will be created with inferior motives than by doing an inferior act. That is, it seems that act consequentialism is not foundationally consequentialist after all as it permits rightness to be dictated by considerations apart from the aim that outcomes should be as good as possible.

According to Pettit and Smith, motive consequentialism encounters a parallel problem. It too is a form of local consequentialism, that is, it does *not* evaluate everything directly in terms of its consequences, which leads it to privilege motives at the expense of value maximization.³⁶

To see why motive consequentialism encounters a parallel problem, we can employ Pettit and Smith's arguments against it. According to motive consequentialism, the right set of motives to have is the one that will produce the greatest amount of utility of all the possible motive sets for that agent. The rightness of acts, for example, can be determined in relation to this set of motives in several ways. According to Pettit and Smith, motive consequentialism can define the right acts in one of the following ways, i), ii), or iii):

[i)] Those which are caused by the right motives; or . . . [ii)] to be those which would have been caused by possession of the right motives; or . . . [iii)] to be those which would have been caused by the motives that it would be best for someone to try to have; or it might define right acts in some further way.³⁷

But all of these, like *Chartres*, seem to fall prey to a form of worship objection. Consider what happens if motive consequentialists define the right acts in the first way mentioned in this quote, as i) those that are caused by the right motives. Then if someone who does *not* have the right motives but still does the value maximizing act of the ones available to him will

³⁶ Pettit and Smith, "Global Consequentialism," 124.

³⁷ Ibid.

be doing the wrong act, which seems absurd from a consequentialist perspective.³⁸ Suppose that *S* is asked to provide a benefit either to her children or to a stranger, and that giving the benefit to the stranger will produce the best possible consequences. Suppose also that if the best overall motive set to have includes loving your children very much, but since *S* does not love her children as much as she should she gives the benefit to the stranger. According to i) she *does* the wrong thing, even though the act itself maximizes value.

It also seems absurd, from a consequentialist perspective, to say that *any* act that is caused by the right set of motives is the right act. That would mean that even when there is an alternative act available to the agent which will produce more utility, the act that is not value maximizing is right. This would mean that if *S* does love her children as much as she ought to, for example, and she gives the benefit to her children, then that act is right even though giving the benefit to the stranger would be much better.³⁹

So act consequentialists end up worshipping acts because they don't evaluate motives directly, but motive utilitarians who define the right act in terms of i) are open to the objection that they worship motives, because they don't evaluate the *act* directly in terms of the good, but rather indirectly through the right motives.

It appears like the right acts cannot simply be the ones that are caused by the right motives. But what should we say about the second version mentioned by Pettit and Smith, which defines the right act as ii) the one that *would* be caused by possession of the right motives? This version says is that your act is right if and only if it is the act that the right motives would produce, but it does not require that the act is *actually* caused by these motives. Because of this, definition ii) allows that someone may do the right thing even though they have the wrong motives. The upshot is that it avoids the result that someone with the wrong motive set can never do a right act. However, it does not avoid the problem that any act that would be caused by the right motives is right. It is no less puzzling why it should be seen as right, by consequentialist standards, for someone to perform an act that is not value maximizing.

³⁸ One example of this is if what someone *wants* to do happens to maximize value, but her motives are not themselves value maximizing. Let's say she *could* have wanted it even more for example, which might have further enhanced the value of doing it.

³⁹ Although, if *S* really *cannot* love her child this much, and also give the benefit to the stranger, then it might be permissible for *S* to give the benefit to the stranger. More on this later.

Chartres II: Imagine Eartha, who *ought to be* as thoroughly motivated to see all of Chartres as Jack is. That is, the right motives for her to have would include the motives that Jack has. Let us also assume that if she were to have these motives, she would also spend too much time at Chartres.

Would it be right for her to spend too much time at Chartres, even though she is *not* motivated to do so? From a consequentialist perspective that clearly cannot be the case, in fact, since she is not so motivated, this seems doubly wrong. She has the wrong motives and because of this gets very little out of her visit to *Chartres*.

The final version that Pettit and Smith mentions says that the right act is iii) the one that would have been caused by the motives it would be best for someone to *try* to have. But this too fails the test of value maximization. To show why, they provide the following example:

Consider a possible world in which there is a mad scientist who will make millions of people miserable if certain individuals don't have malignant motives in the future, but who couldn't care less which acts they perform now. Let it be agreed that it would therefore be best for them to try to have malignant motives. Yet it is surely quite implausible to suppose that these people act rightly, here and now, if they act as if they had these motives already.⁴⁰

It seems absurd, from a consequentialist perspective, to say that just because you should be *trying* to attain a set of motives, you should be acting on those motives already. What you should be doing now is whatever maximizes utility. It seems reasonable in this case to try to maximize utility by inculcating malignant motives in yourself, but in that case you would be doing so because you have *non-malignant* motives, namely to maximize utility. In fact, if you were to act as if you already had these malignant motives, you would probably be doing everything you could *to lose* your malignant motives, since this would have such dire consequences.

We have considered three motive consequentialist criteria for the rightness of acts, and they have all been found to be guilty of motive worship. So has act utilitarianism and rule utilitarianism. This suggests that the best strategy might be to say that the right act is the one

⁴⁰ Pettit and Smith, "Global Consequentialism," 130.

that maximizes value, the right motive is the one that maximizes value, and the right rule is the one that maximizes value. This is what global consequentialism proposes.⁴¹

The problem with worship arguments is that there are so many possible versions of consequentialism that it would be a herculean task to write one for every single theory. What Pettit and Smith and Kagan do is provide worship arguments for the most common forms of consequentialism. Between them they have found worship arguments for act consequentialism, rule consequentialism, and motive consequentialism, which are what you may call the usual suspects.⁴² They admit that their argument is inductive and that there may therefore be exceptions.⁴³ Clearly though, their view is that this is a remote possibility. There are probably no exceptions, and all forms of non-global consequentialism will fall prey to some version of the worship argument. I cannot go through them all given the limited length of this thesis. From this point on I will therefore assume that all versions of consequentialism are vulnerable to a version of the worship objection, or at least that all versions of consequentialism except global consequentialism are vulnerable to it. Global consequentialism is designed specifically to avoid this objection.

We are now in a position to state more clearly what the worship objection is. It is an accusation that some theory claims that x is right (where x can be any evaluative focal point), even though x would not have the best consequences, or that x *does* have the best consequences, yet x is *not* right. This conflicts with foundational consequentialism, and it seems very strange for a consequentialist to say that act x would produce the best outcomes, and is therefore best, but it is not *right*. Or the other way around, that x is not the best, but x is still right.⁴⁴

The worship objection motivated philosophers to create new versions of consequentialism that focus on different focal points, but in doing so they have fallen prey to the same objection. That is because, as we have seen, their new proposals were local, just like the versions they were criticizing. That is, the new theories all had some *primary* evaluative

⁴¹ Of course, if you cannot have the best motives and do an act that goes against those motives then that is another matter. But as I shall be considering this point in very close detail in the next chapter, I choose not to delve into it here as well.

⁴² Pettit and Smith, "Global Consequentialism," 125–132; and Shelly Kagan, "Evaluative Focal Points," 137–150.

⁴³ Pettit and Smith, "Global Consequentialism," 130–132; and Shelly Kagan, "Evaluative Focal Points," 152.

⁴⁴ Jennie Louise, "Right Motive, Wrong Action," 70.

focal point(s) that other focal points were evaluated in terms of. Global consequentialism is not open to this objection since all focal points are evaluated directly.

Beyond the worship argument, there is a positive reason to be a global consequentialist. It makes it possible for consequentialists to incorporate motives, characters, laws, etc. into their theory directly. The weight of this argument should not be overstated however, since global consequentialists have to admit that some sets of motives which *prima facie* appear to be horrendously bad sets of motives, will actually be the right ones according to global consequentialism. Consider the malignant motives acquired in Pettit and Smith's mad scientist example, while they are clearly the right motives to acquire from a consequentialist perspective, they are clearly not what common-sense morality would accept as right or good motives.⁴⁵ From a consequentialist perspective, selfish or malignant motives may be right and good, as long as they have the best consequences. In other words, global consequentialism cannot accommodate the idea that motives are *intrinsically* good.

1.4 Global Consequentialism

Until now, I have not said much about what global consequentialism is besides that it evaluates all focal points directly in terms of value. Giving a clearer explanation of exactly what it is will be the aim of this section. As a starting point for a definition, we can use Pettit and Smith's formalization.

Global consequentialism identifies the right x , for any x in the category of evaluands [evaluative focal points] — be the evaluands acts, motives, rules, or whatever — as the best x , where the best x , in turn, is that which maximizes value.⁴⁶

So the right climate is the climate that will have the best consequences, the right motives are the motives with the best consequences, etc.

There are several issues that present themselves. My aim is not to defend global consequentialism from all possible objections that can be aimed at it, yet having an understanding of what these issues are will be helpful to the reader in what follows. For this reason I will present a few relevant considerations before I expand on Pettit and Smith's version.

⁴⁵ Elinor Mason, "Against Blameless Wrongdoing," 294–295.

⁴⁶ Pettit and Smith, "Global Consequentialism," 121.

First of all, you might wonder how a bottle of shampoo (for example) can be described as the *morally* right bottle of shampoo. Some global consequentialist theories have assessed ordinary objects and events as merely ‘better’ and ‘worse’ and remained silent on their rightness, while others, like Pettit and Smith, couch their assessments in terms of ‘rightness’ and ‘wrongness’, even for shampoo.⁴⁷ (Remember that ‘rightness’ and ‘wrongness’ do not automatically translate to *obligation*, especially when applied to that which we have no control over, say, death or taxes.) I will employ ‘right’ and ‘wrong’ for all focal points in this thesis for stylistic reasons, but I am certainly open to the possibility that this is not strictly correct when it comes to shampoo, for example. If you find this counter intuitive, just replace ‘right’ with ‘best’.

Second, travelling at the speed of light could (perhaps) be described as better than travelling at half the speed of light in certain cases, but because that is impossible, global consequentialists cannot say we *ought* to travel at twice the speed of light. In other words, there is a question of how we find the set of speeds that are available, in the appropriate sense. Put slightly differently, we should ask ourselves how global consequentialists should interpret the ‘ought’ implies ‘can’ principle (OIC). OIC is an oft-used ethical principle which tells us that it cannot be the case that you ought to do something if you are unable to do it.⁴⁸ For example, it will probably not be the case that you ought to rescue someone from a burning building if you are completely paralyzed. Though OIC is ubiquitous in ethical literature, it is also difficult to parse how it should be interpreted. The most relevant term in this doctrine—for our purposes—is ‘can’. We need to employ a workable possibility constraint in order to generate an ‘ought’ statement, since the right *x* is the best *possible x* within the category of the same evaluative focal point. Parfit employs causal possibility, but there are other possible options.⁴⁹ Global consequentialists have to employ *some* version of OIC, but it is not necessary to commit to one version at this point.

Third, evaluating a shampoo bottle normatively makes little sense without evaluating it in some context. If all we know about *x* is that *x* is a token shampoo, we cannot say much about what consequences it will have. We need to know who has the shampoo, what it will be used for, *etc.*, before we can assess whether it is the right shampoo. Of course, any actual token shampoo bottle will have one actual set of effects, which may or may not be causally

⁴⁷ Pettit and Smith, “Global Consequentialism,” 121; Derek Parfit, *Reasons and Persons*, 24–26.

⁴⁸ Elinor Mason, “Objectivism, Subjectivism, and Prospectivism,” 188.

⁴⁹ Derek Parfit, *Reasons and Persons*, 14.

determined, but if all we know is that x is a bottle of shampoo, the goodness or rightness of that bottle cannot be determined. In other words, there is no answer to the question ‘what consequences shampoo x have?’, any more than there is an answer to ‘what consequences will doing laundry have?’. That is, unless you are speaking about some specific person doing some specific laundry at some specific time. As Ord puts it:

In the case of climate, Parfit writes that ‘the best possible climate is the one that would make outcomes best’, But in what role are we to evaluate this climate? Are we to consider it as the climate for Oxford? for England? for the entire world? It is not clear.⁵⁰

There is no answer to the question because the question is incomplete; it might have bad consequences if had by Oxford and good consequences if had by, say, Equatorial Guinea. So we need to evaluate it through what Ord calls its ‘*role*’:

We could assess a climate in the role of being had by Oxford in the 18th Century, or being had by Madagascar throughout its existence. This conception of roles is philosophically simple. A role is just a single-place predicate. When given a token from the class under consideration, it forms a proposition. For example, the first role above is that of *being a climate had by Oxford in the 18th Century*. When a climate, such as ‘tropical’ (or something much more specific) is provided, this forms the proposition: *a tropical climate was had by Oxford in the 18th Century*. We can then assess the outcome that would result from this being true.⁵¹

When we have this extra information about the evaluand, an answer to what consequences it had or will have becomes possible.⁵² Even though there is no answer to what consequences a tropical climate will have, there is an answer to what consequences a tropical climate had by Oxford in the 18th century had, and in principle there is no greater difficulty in figuring out what the expected consequences of some climate in a specific role are, than what the expected consequences of an act are. (Though figuring out either is immensely difficult in practice of course.)

There is one snag that needs to be addressed. If x is the best act available to you then x is the right act and it follows that you ought to do x , or, if x and y are equal, they are both right and you ought to do one of the two. Likewise, you might think it follows that if B is the right

⁵⁰ Toby Ord, “How to Be a Consequentialist About Everything,” 6.

⁵¹ Ibid.

⁵² Though in this case it is a false statement of course, in order to be genuinely interesting, such claims about the past should be true, or at least possible. And if we ask about what climate Oxford ought to have had, or ought to have in the future, things become complex rather fast. But these are practical matters. It is true that forms of utilitarianism are open to the objection that because of lack of information we do not know what will produce the best outcome, to this they have always replied that we should do the best we can with the information we have. The point is that there *is* an answer to what consequences a (maximally specific) description of a climate in a role will or would have, at least in probabilistic terms.

person for *A* to kill then *A* ought to kill *B*, but this does not make sense. It clearly cannot be the case that *A* ought to kill *B*, just because *B* is the right person for *A* to kill. Even though *B* is the right person for *A* to kill, it remains very much of an open question whether *A* ought to kill anyone whatsoever, most likely he ought *not to*. Ord specifies that “*x* is the right *y* to be in role *r*” does not imply that “It is right for *x* to be in role *r*”.⁵³ Rather it is:

right for *x* to be in role *r* iff *x* is the right *y* to be in role *r* and it isn't better to have no *y* in *r* than to have *x* in *r*.⁵⁴

For example,

it is right . . . for Susan to be in the role ‘being killed by Mark’ iff Susan is the right person to ‘be killed by Mark’ and it isn't better to have no person ‘be killed by Mark’ than to have Susan ‘be killed by Mark’.⁵⁵

We would not need the last qualification for acts since not doing anything is an act. That means that you would never get results like these by claiming that bestness means rightness when it comes to acts, they are always the obligatory ones, barring ties.

Pettit and Smith's formalization does not account for this last peculiarity of global consequentialism, so by introducing the concept of roles, Ord provides a more precise formalization of global consequentialism than Pettit and Smith does. I will now give a fully fleshed-out formalization of Ord's role-based global consequentialism. It has three parts, first it defines *best in a role*, then it defines *right in a role*, and then finally it defines *role-based obligations*.

Global consequentialist criterion of best within a role

x is the best *e* in the role *r* iff the outcome that would occur were *x* in role *r*, is better than the outcomes that would occur were any other *e* in role *r*.⁵⁶

Example of the global consequentialist criterion of best within a role

Wella is the best *shampoo* to be in the role ‘being used regularly by Tim’ iff the outcome that would occur were *Wella* in the role of ‘being used regularly by Tim’, is better than the outcomes that would occur were any other *shampoo* in this role.

⁵³ Toby Ord, “How to Be a Consequentialist about Everything,” 11.

⁵⁴ Ibid.

⁵⁵ Ibid.

⁵⁶ Ibid., 16.

Notice that strictly speaking this is too non-specific to work. In order for this to work in a real life case you would need to refer to some specific bottle of Wella shampoo, at some more specific time than ‘regularly’, just like you would have to specify who is referred to by ‘Tim’, since there is more than one Tim. But my aim is to give an overview of how role-based global consequentialism works, and I think this gives a decent template for how ‘bestness within a role’ would be defined. All you need to do is exchange the somewhat vague descriptions used here with *maximally specific* descriptions.

Global consequentialist criterion of rightness within a role

x is the right *e* in role *r* iff *x* can be in role *r*[,] and for each other *e* that can be in role *r*, *x* is better than it in *r*.⁵⁷

Example of the global consequentialist criterion of rightness within a role

Wella is the right *shampoo* in the role ‘*being used regularly by Tim*’ iff *Wella* can be in the role ‘*being used regularly by Tim*’[,] and for each other *shampoo* that can be in the role ‘*being used regularly by Tim*’, *Wella* is better than it in this role.

And finally,

Global consequentialist criterion of role-based obligations

s ought that *x* be in role *r* iff *x* can be in role *r*; for any other thing that can be in role *r*, *x* is better than it in role *r*; and it isn’t better to have nothing in role *r*.⁵⁸

Example of the global consequentialist criterion of role-based obligations

Tim ought that *Wella* be in the role ‘*being used regularly by Tim*’ iff *Wella* can be in the role ‘*being used regularly by Tim*’; for any other thing that can be in the role ‘*being used regularly by Tim*’, *Wella* is better than it in this role; and it isn’t better to have no *shampoo* be in the role ‘*being used regularly by Tim*’.

One oddity about this formalization is that it says that ‘*S* ought that *x* ...’, this is clearly ungrammatical, and may seem strange. In order to address this, Ord appeals to a set of examples discussed by John Broome.⁵⁹ As I have already mentioned, there are ‘ought’ statements that do not refer to acts. Broome adds that there are ‘ought’ statements where it is not clear who is responsible, his example revolves around Julie, of whom it is true that she ought to have more work to do. Who does this ought-statement belong to? That is, who is

⁵⁷ Toby Ord, “How to Be a Consequentialist about Everything,” 16.

⁵⁸ Ibid.

⁵⁹ Ibid., 19. According to Ord, Broome discusses it in an unpublished manuscript.

responsible for this ought statement, Julie or her boss? Broome argues that you should be able to make ought-statements like ‘*S* ought that *x*’ even though they are ungrammatical. In fact, Broome sees this as a “grammatical innovation” which “extends the grammar of ‘ought’, and overcomes a merely quirk of the English language”.⁶⁰ If you are able to make statements like these, then you can attribute statements like ‘Julie ought to have more work to do’ to someone by saying that ‘Julie’s boss ought that Julie ought to have more work to do’.

You may not find this convincing, and I have not gone through all the objections to a role-based global consequentialist framework. My goal is to defend global consequentialism from the objection that it allows evaluative conflict, not to provide a full-fledged defense of the theory. I will be using this as a rough framework because it is the most complete framework I know of, and also the most straight-forward and intuitive one.

I have now explained what evaluative focal points are, and what foundational consequentialism is. I have also made a distinction between objective, subjective and prospective rightness. Additionally, I have given a framework for global consequentialism and shown why it is needed, which is that it evades the worship objection. This should suffice for us to finally take a closer look at the notion of *evaluative conflict*, which will be the main focus of this thesis.

1.5 Evaluative Conflict

Once we make the leap to evaluate more than one focal point directly, the possibility of evaluative conflict arises. Since different focal points can affect each other causally, it appears that different focal points can also conflict. The most common type of example is when the right set of motives (the value maximizing set) would lead you to do the wrong act (an act that is *not* value maximizing). If so, should you have the right motives or should you do the right act?

The possibility of evaluative conflict has been seen as troublesome by many, often sufficiently so that they reject theories that allow for it. According to Brad Hooker it is paradoxical.⁶¹ It has also gone by several names; there is quite a bit of literature on what Parfit calls *blameless wrongdoing*, blameless wrongdoing is not itself a name for evaluative conflict, but rather one possible reply to it. Ord calls it ‘the inconsistency objection’.

⁶⁰ Toby Ord, “How to Be a Consequentialist about Everything,” 12–14.

⁶¹ Brad Hooker, “Rule Consequentialism,” § 5.

There are slight differences between the meanings of the inconsistency objection, evaluative conflict, and the cases Parfit describe, which lead him to espouse blameless wrongdoing. Perhaps the easiest place to start is with an example I call *Hooker's paradox*. In this case, the conflict occurs between a decision procedure and an act.

Suppose, on the whole and in the long run, the best decision procedure for you to accept is one that leads you to do act x now. But suppose also that in fact the act with the best consequences in this situation is not x but y . So global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. That seems paradoxical.⁶²

Ord holds that the objection the example is indicative of, can be made explicit by saying that cases where something morally right necessarily leads to something that is morally wrong, are inconsistent with an intuitive principle. Specifically he thinks the objection is that cases like *Hooker's paradox* clash with the principle "if you ought to X , and you cannot X without Y -ing, then it is permissible to Y ".⁶³ But, as a global consequentialist, Ord argues that we should reject this principle, no matter how intuitive it may be. I shall return to this in section three.

Julia Driver says that the problem with *Hooker's paradox* is that it provides conflicting guidance.

One problem raised for every form of global consequentialism is that it provides conflicting guidance. . . . Brad Hooker views this problem as raising a genuine paradox for global consequentialism.⁶⁴

According to Driver the problem is that the example prescribes a set of pragmatically (jointly) impossible actions that you ought to do. The problem with this is that the prescribing theory ceases to be a useful guide to action. Sets like these may also conflict with OIC if they tell you that you ought to do something that you cannot do.

So, there are different takes on what the problem is supposed to be. Ord takes the objection to be a problem with the logic behind obligation; Driver takes it to be a practical one. Hooker's example is ambiguous between the two. Shelly Kagan also sees it as mainly a practical problem,

⁶² Brad Hooker, "Rule Consequentialism," § 5.

⁶³ Toby Ord, *Beyond Action*, 92.

⁶⁴ Julia Driver, "Global Utilitarianism" in *The Cambridge Companion to Utilitarianism*, ed. Ben Eggleston and Dale E. Miller (New York: Cambridge University Press, 2014), 175.

What are we to say in cases of conflict? If the best rules direct us to different acts from the best motives, which in turn direct us to different acts from the best norms, and so on, what are we to do?⁶⁵

Kagan refers to prescription; unlike Ord he focuses on what we should do and not the logical form as such. Jennie Louise relies on Adams' example *Chartres*, which I have already presented, and Parfit's example *Clare*.

Clare: Imagine that Clare has the best possible set of motives she can have, which includes a strong love for her child, and that: "Clare could either give her child some benefit, or give much greater benefits to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger."⁶⁶

Since Clare is not doing what will maximize value she does something wrong, and she does something wrong *because* of her right motives. This is what makes it a case of evaluative conflict. Clare can either do the right thing or have the right motives, but she cannot both have the right motives and do the right act. If this is what global consequentialism prescribes then it appears not to provide an answer to what Clare *really* ought.

Hooker's paradox, *Chartres* and *Clare* arguably make up the three most common examples of evaluative conflict.⁶⁷ *Chartres* and *Clare* indicate a conflict between motives and acts, whereas *Hooker's paradox* points to a conflict between a decision procedure and an act. In essence, evaluative conflict arises when a theory tells you that you ought to *x* and you ought to *y*, but you cannot both *x* and *y*.

There are some ambiguities that need to be cleared up. Louise describes the problem as one of causally related evaluative focal points. Since our motives affect our actions, a right set of motives can lead to the wrong action.⁶⁸ It is important to note that she uses the phrase 'lead to', as does Hooker, and that this is not sufficiently precise. Ord argues that a motive set merely leading to an act does not pose a serious problem. He asks us to consider a case:

Pride After a Fall: "Penny is crossing a bridge when she hears a sudden splash and a frantic cry for help. She sees that a young boy has fallen into the river and that he cannot swim. Since no-one else is around, she wades in and rescues the boy, thereby saving his life. Later that day

⁶⁵ Shelly Kagan, "Evaluative Focal Points," 152.

⁶⁶ Derek Parfit, *Reasons and Persons*, 32.

⁶⁷ Jennie Louise, "Right Motive, Wrong Action," 68–69.

⁶⁸ *Ibid.*, 66.

she is so proud of having saved someone's life that she decides to take the rest of the afternoon off, breaking a promise to help a friend move house."⁶⁹

In this case a right act leads to a wrong act, but there is nothing intuitively problematic about this. This kind of "evaluative conflict" happens frequently within the framework of classical utilitarianism, many deontological theories, and common-sense morality, so a theory allowing right simply *leading* to wrong does not seem to be a problem. This is why Ord uses the term 'cannot' in the principle that he thinks cases of evaluative conflict clashes with, that is, "if you ought to X, and you cannot X without Y-ing, then it is permissible to Y".⁷⁰

This is presumably also why Parfit regards motives as *necessarily* leading to actions in his *Clare* example.⁷¹ It at least appears like we should say the same thing about *Chartres* and *Hooker's paradox*; that the connection has to be a necessary one for there to be an obvious problem.

On the other hand, concerning *Chartres*, Adams says that Jack's staying at the choir screen later than is optimal, and his previous attainment of the desire to watch choir screens, are "separable in a way that striking the ground after jumping from a burning building is not".⁷² He thinks that, in some way, Jack could choose not to stay all day even with his motive set perhaps. I will return to this disagreement at length in chapter three, which contains the lion's share of my defense.

There is one last difference that should be mentioned, Mason and Louise keep referencing evaluative conflict and characterizing it as a "*best* motive leading to a *worse* act" and other similar turns of phrase.⁷³ They speak of it in value terms as opposed to in deontic terms.⁷⁴ A best motive unavoidably leading to a non-best act does not necessarily pose a problem though, unless we assume that a non-best act and motive also has to be wrong. This is a common assumption among consequentialists, but as I shall argue below, it might not be a good one in this particular case. Or, at least the sense in which this assumption is true might not be the one that leads to evaluative conflict.

⁶⁹ Toby Ord, *Beyond Action*, 92.

⁷⁰ *Ibid.*

⁷¹ Derek Parfit, *Reasons and Persons*, 14.

⁷² R.M. Adams, "Motive Utilitarianism," 473.

⁷³ Elinor Mason, "Against Blameless Wrongdoing," *Ethical Theory and Moral Practice* 5, no. 3 (2002): 294.

⁷⁴ 'Deontic' means that it pertains to moral *obligation*.

1.6 Recap and Roadmap

I have defined what consequentialism is and how it relates to other types of theories, as well as delineated some key concepts, such as evaluative focal points, foundational consequentialism, and the difference between objective and subjective rightness. We have also seen how local consequentialist theories give rise to the worship objection, and that global consequentialism is immune to it. Finally, we have seen that solving the worship objection may come at the cost of allowing *evaluative conflict*. We should now have enough of an overview, so that that we can start looking into possible solutions to the problem.

In the following section I will take a closer look at one solution to the problem; that any act that necessarily has to follow from a right set of motives inherits the rightness of the motives. This might seem reasonable, since it would mean that you don't end up in situations where you are "damned if you do, and damned if you don't". I will conclude that this strategy appears to work, but as we shall see in section four, Parfit springs a seemingly devastating objection on this strategy. From section 4.2 and onward, I will attempt to show that Parfit's objection has a series of unattractive consequences. Since these unattractive consequences also seem to apply to other theories than global consequentialism, I will attempt to thread the needle just right, so that evaluative conflict, and these other unattractive consequences, is avoided.

2 IS RIGHTNESS TRANSFERABLE?

2.1 Parfit and Louise Versus Mason

Can the rightness of one focal point, such as motives, be inherited by another focal point, like the acts they cause?⁷⁵ Take Clare, could the rightness of her motives be transferred to the act of benefiting her child? If it could be shown that this is not only the case sometimes, but it is *always* true of purported cases of evaluative conflict, it would show that these cases were not really cases of evaluative conflict after all, because the right motive would never lead to a wrong action, but rather to a right action.

This might seem like a plausible option since it would allow that if you tie yourself to the mast, so to speak, by acquiring the right motives, you will not later be doomed to do the wrong thing because of those motives. Parfit and Louise argue against transferability, whilst Mason argues for it.⁷⁶ I agree with Mason.

According to Parfit, if the following principle G4 is true, it would seem that rightness is, in some sense, transferable:

G4: “if someone acts on a motive that he ought to cause himself to have, and that it would be wrong for him to cause himself to lose, he cannot be acting wrongly”⁷⁷

Parfit proceeds to argue that G4 is false by providing a counterexample, but in the footnote to the example he admits that it is defective.⁷⁸ His counterexample might be salvageable, but since Louise provides a clearer one, I will use that instead. It is parallel to Parfit’s case, only tidier.

My Malevolence: “The universe has been taken over by a perverse demon who wishes to be worshipped as a great and kind god, but who (for some reason) wishes me to be reviled as a monster. To that end, he offers me a drug that will cause me to become malevolent. If I do not take the drug, he will make everyone in the universe miserable for all eternity. If I *do* take the drug, he will create a state of affairs in which the only unhappiness that will ever exist from then on is the unhappiness which I create through my actions: That is, with the exception of the unhappiness I choose to cause, there is universal happiness. Only the act of taking (or refusing) the drug is relevant: If I take the drug, the demon will make everyone happy regardless of what my later actions are (i.e., even if I do not later act malevolently). But because taking the drug

⁷⁵ I am using ‘transference’ and ‘inheritance’ as synonymous.

⁷⁶ Derek Parfit, *Reasons and Persons*, 38; Elinor Mason, “Against Blameless Wrongdoing,” 287–294; and Jennie Louise, “Right Motive, Wrong Action,” 70–73.

⁷⁷ Derek Parfit, *Reasons and Persons*, 35.

⁷⁸ *Ibid.*, 509.

will make me malevolent, I will for the rest of my (mortal) life go around randomly torturing people.”⁷⁹

Let’s call the agent in this example Molly. According to Louise and Parfit, cases like this one show *G4* to be false, since the malevolence was right for Molly to cause herself to have, it would be wrong for her to cause herself to lose it, and yet, according to them, it is clearly not *right* to go around randomly torturing people after taking the pill. Therefore, rightness is not transferable, and the upshot is evaluative conflict.

Mason disagrees; she responds that though *G4* is false, it is false in a trivial way. If we formalize *G4* we get “For an act ϕ , it is right to ϕ if one is acting on a motive it is right to cause oneself to have, and wrong to cause oneself to lose.”⁸⁰ She argues that one way of filling out this formalization is by replacing ϕ with ‘a wrong act’, if so the principle becomes self-contradictory, as it can never be right to perform a wrong act.

Consequently she agrees that the principle is false, but Mason thinks examples like *My Malevolence* don’t accomplish what Louise and Parfit want them to, it proves the principle is false but does not further the case that rightness is non-transferable. Her point seems to be that while *My Malevolence* shows that the *principle* *G4* is false, it would need to show more than this in order to show that rightness is not transferable. The principle says that an act that results from a motive you ought to have *cannot* be wrong, but all Mason needs is that there are *some* cases where rightness or permissibility is transferred, as long as they are all of the ones that would otherwise result in evaluative conflict. Mason is not claiming that right motives result in right acts and *only* right acts; she is merely claiming that the resulting acts are right if you *cannot* do anything better, given the right motives.⁸¹

Specifically, she thinks that *My Malevolence*—and examples like it—*beg the question* against those who believe in transferability of rightness. They show that *G4* is false, but Mason is not committed to it anyway, and they may also show that rightness *has* to be transferable. Since it is right for Molly to cause herself to become malevolent, it cannot be the

⁷⁹ Jennie Louise, “Right Motive, Wrong Action,” 72.

⁸⁰ *Ibid.*, 71.

⁸¹ In fact, the claim is even weaker than this, since the necessary connection could go the other way. That is, that you truly *cannot* have the right motives if you do the right act. This might happen if you know you have to do some heinous act that will in fact save a lot of people, but also result in you becoming callous and disaffected, which in turn might happen if you will not *know* that you saved these people afterward for example. As long as the value of that single act outweighs all of the disvalue of the subsequent bad motives, you still ought to do it, and your subsequent bad motives will not be wrong if rightness is transferable.

case that she acts wrongly when she later acts according to those motives. That is, Parfit and Louise think that since the acts of torture cannot be right, rightness is not transferable, but Mason thinks that since Molly ought to take the pill, and that will lead to torturing people, the subsequent acts of torture have to be right.

“But this is precisely why someone might claim that . . . rightness must be *transferable*! Parfit has shown that a notion of transference is necessary, as the acts performed after tying oneself to the mast are not ‘right’ . . . by virtue of their own qualities. Parfit has certainly not shown that a motive it is best to have can lead to an act it is worse to do”.⁸²

Mason’s phrasing in this passage may not have been the most fortuitous. Louise responds that the latter is *exactly* what has been shown. Mason accepts that in *My Malevolence* the subsequent acts of torture are not right by virtue of their own qualities, and then, Louise argues, she moves to the *stronger* claim “that it is not true in such cases that a *best* motive can lead to a *non-best* act”.⁸³ But the example *does* show that this can happen according to Louise, because torturing people is worse than not torturing people, and in *My Malevolence*, Molly tortures people because she has the right motives.⁸⁴

It might look like Louise is straightforwardly correct, but concluding in her favor would be premature, as we shall see. But first, what we have here is a standoff between two views. Parfit, Mason and Louise all agree that taking the drug, becoming malevolent, and then acting malevolently is *better* (produces more value) than not taking the drug and then acting non-malevolently. They also agree that taking the drug is the right thing. Presumably, they also agree that the *best* thing would be to take the drug and then not torture people as well. But on Mason’s view, transference is needed *because* without it we get the unsatisfactory answer that Molly is obligated to acquire the motives, and then she necessarily has to do the wrong thing because of those motives. Louise thinks that *even though* acquiring the motive was obligatory and therefore right, the necessary subsequent acts are clearly wrong. On Louise’s view “we have not yet been given any reason—other than discomfort with evaluative conflict—to suppose that the worse act in question can be described as *right*”, and she thinks that this is “more than a mere oddity”.⁸⁵ This is a standoff, but Mason moves to a

⁸² Elinor Mason, “Against Blameless Wrongdoing,” 294.

⁸³ Jennie Louise, “Right Motive, Wrong Action,” 72.

⁸⁴ *Ibid.*

⁸⁵ *Ibid.*

much stronger claim in the passage above.⁸⁶ Why does she do this? It is time to clear up some confusions.

First, an obvious one, ‘not right by virtue of their own qualities’ is ambiguous between “it is not the case that the acts of torture are right, this is because of their own qualities” and “it is not the case that the acts of torture are right by virtue of their own qualities, but they may be right by virtue of something else”. Mason clearly means the latter, since she is arguing that the acts can be right because they inherit the rightness of the motives.

But there is a more subtle ambiguity involved. Louise is *certainly* right that a motive set it is *best* to have can lead to an act that it is *worse* to do, if by worse you mean worse than some act that is possible with *other* motives. There is little doubt that having the consequentially speaking best possible set of motives can lead you to do something that is worse than what you would do with a lesser set of motives. This is true simply because the best possible set of motives results in more than one act, and it is the sum value of the resulting acts—infused with the value of the motives themselves—that determine the overall value of the motive set. But Mason is not really arguing against this obvious claim.

After the quote where she does use ‘better’ and ‘worse’, Mason immediately switches back to writing as if what she has argued is that *rightness* is transferable. I think it is clear that Mason does not take ‘worse’ to mean worse than some other act that is possible with *other* motives, she means an act which is worse than some other act that is possible with the *same* motives. In other words, I think the use of the terms ‘better’ and ‘worse’ is simply an inaccuracy on Mason’s part, which Louise—of course—takes at face value. What Mason really intends, I think, is to show that the *right* set of motives cannot (necessarily) lead to the *wrong* act. Louise is successful in showing that *My Malevolence* shows that a best motive can lead to a non-best act compared between different motives, but if I am right this is irrelevant to the real argument that Mason is making. There are three reasons why I think this is the case.

First of all, rightness and wrongness is *what is at issue*. They disagree whether rightness can be transferred, not whether bestness can be transferred. Specifically they disagree whether the right motive can *necessarily* lead to the *wrong* act, as we remember from Ord’s *Pride after a fall*.

⁸⁶ Jennie Louise, “Right Motive, Wrong Action,” 73.

Second, as Louise correctly observes, by using ‘best’ instead of ‘right’, Mason’s claim becomes much stronger. In fact, it becomes unnecessarily strong, since there is no benefit to Mason’s argument in making this strong claim. Her point is that the actions that would have been possible for Molly to perform if she had non-malignant motives are irrelevant, since she does not have those motives, and Molly can only do as good acts as she can with her actual motives.⁸⁷

And third, in a series of personal correspondence e-mails Mason agreed with me that what she had in mind was the same argument with the terms ‘right’ and ‘wrong’ instead of ‘better’ and ‘worse’, and she directed me to a paper she has written where she warns against the dangers of conflating the two.⁸⁸

Still, based on *My Malevolence* and what has been said thus far, it might seem like Louise has the stronger intuition on her side, since it undeniably sounds strange to say that randomly torturing people is the right thing to do. It is time to look more closely at Mason’s reasons for thinking that it is incorrect to call actions done on the basis of rightly acquired malignant motives *wrong*.

2.2 How Transferring of Rightness Works

Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off—then, I account it high time to get to sea as soon as I can.⁸⁹

—Ishmael, *Moby Dick*, by Herman Melville

There are cases where it is completely natural, from a consequentialist point of view, to say that torture is right. For example, if your only two choices are torturing someone or letting millions of people die. In these cases, the *explanation* for why it is right to torture, is that out of the acts available to the agent, torture is the best act. Similarly, if you can give an explanation of how or why rightness transfers in cases like *My Malevolence*, then that will serve to significantly lessen the impact of the intuition Louise depends upon, that torture

⁸⁷ Elinor Mason, “Against Blameless Wrongdoing,” 294.

⁸⁸ Elinor Mason, “Objectivism and Prospectivism About Rightness,” *Journal of Ethics and Social philosophy* 7, no. 2 (March 2013): 17; and personal e-mail to the author 27.3.2015.

⁸⁹ Herman Melville, *Moby Dick or the White Whale* (Boston: The St. Botolph Society, 1892), 7.

cannot be right. And it *is* in a parallel sense that torture would be seen as right in *My Malevolence*; torture is the best option available.

Mason's point is that when Molly has taken the drug and chooses to torture someone, it *is* the best act causally available to her. *My Malevolence* says that taking the drug "will make you for the rest of your (mortal) life, go around randomly torturing people".⁹⁰ You *will*. Mason takes this to mean that Molly will be *unable* to do anything else after taking the motive-altering pill. If we accept that 'ought' implies 'can' and that torturing people is the best act she can do, then how could torture be wrong?⁹¹ The idea is that not torturing people is no longer a part of the set of available options for Molly after taking the pill, just like magically making everyone in the world happy is not a part of the options available to me now. This seems to imply that that *My Malevolence* is just like the case where you have a choice between torturing someone or letting millions of people die, the right choice cannot be to do neither, since that is *stipulated* to be impossible.

But it is not clear that this answer is entirely satisfactory. Is it reasonable to juxtapose magical utile distribution with a very strong motivation?⁹² Ishmael is able to stop himself from methodically knocking people's hats off with the application of strong moral principle. The fact is that we don't usually accept that having a set of motives excuses you from having done something wrong. In fact, whether someone really *wants* (is motivated) to torture people is seemingly irrelevant to whether she ought to do so. In other words, even though *My Malevolence* is phrased in a way that stipulates that Molly cannot refrain from torturing given her motives, it may not be true that such motives actually exist.⁹³

To solve this puzzle we need to understand what a consequentialist would say if it *were* still possible for Molly to refrain from torturing people after taking the pill, which—as luck would have it—has a simple answer. If she *can* refrain from torturing people after taking the pill, then she *ought* to refrain, plain and simple. No matter how difficult it might be to refrain from torturing people, consequentialists will say that Molly ought to refrain from it since this will produce more value than torturing people.⁹⁴ If we assume that taking the pill

⁹⁰ Jennie Louise, "Right motive Wrong action," 72.

⁹¹ I shall return to this in chapter four.

⁹² A utile is the unit of measurement for utilitarian value. There is no definite answer to how much value one utile has, except that it is half the value of two utiles, one third of three, etc.

⁹³ Of course, Molly could presumably be put into an artificial coma immediately after taking the pill, and thus not torture anyone, but we have to eschew possibilities like these.

⁹⁴ They might say that she is very weakly to blame in some cases though.

makes it *almost* impossible to refrain from torturing people, this is certainly very demanding of Molly. But demandingness is not particular to global consequentialism. Classical utilitarianism will demand that Molly refrains from torturing after taking the pill as well, and the demandingness objection to consequentialism has several familiar consequentialist replies.⁹⁵

If it is plain wrong for Molly to go around torturing people, then no evaluative conflict arises. Even though her right motives in a weak sense leads her to perform the wrong act (through making her want to torture), Molly *chooses* to do the wrong thing.⁹⁶ Global consequentialism prescribes her to take the pill and then not torture people, if she fails to do the right thing there is little more to say about it than “she did the wrong thing”. It is *only* if Molly really *cannot* refrain from torturing after taking the pill that a global consequentialist may want to say that torturing people can be right, and only if she does the best possible version of torturing people available to her, whatever that may be. If she cannot, and ‘ought’ implies ‘can’, then Molly ought not to refrain from torturing people.

To make this point even clearer, I will reiterate *Pride after a fall*:

Penny is crossing a bridge when she hears a sudden splash and a frantic cry for help. She sees that a young boy has fallen into the river and that he cannot swim. Since no-one else is around, she wades in and rescues the boy, thereby saving his life. Later that day she is so proud of having saved someone’s life that she decides to take the rest of the afternoon off, breaking a promise to help a friend move house.⁹⁷

It is Penny that is at fault in this example, not that we are considering multiple evaluative focal points directly. It is true that her motives led her (in a weak sense) to do the wrong thing, but all we need to say about this is that she should have done better. The supposed problem with evaluative conflict is that you ought to *x* and you ought to *y*, when you *cannot y* given that you *x*. If Penny and Molly *are* able to refrain from torture then that would be a case where you *can y*, which is unproblematic.

In other words *My Malevolence* requires some impossibility clause in order for evaluative conflict to arise. It truly has to be the case that Molly *cannot* do anything but torture people, otherwise global consequentialism will tell her to refrain from doing so. On the

⁹⁵ David Sobel, “The Impotence of the Demandingness Objection,” *Philosophers’ Imprint* 7, no. 8 (2007): 1–17.

⁹⁶ There is an ambiguity in the phrase ‘lead to’, I accept that your motives might lead you to do something in the sense that loving your child makes it difficult not to favor him unjustly, but I do not accept that this difficulty makes it any less true that you should not favor him unjustly. It is probably possible to do what you do not wish *sometimes*.

⁹⁷ Toby Ord, *Beyond Action*, 92–93.

other hand, if such an impossibility clause is included, it turns out that Molly cannot refrain from torturing people, and since ‘ought’ implies ‘can’, it cannot be the case that she ought to do anything else. If so, then it will never be the case that global consequentialism results in evaluative conflict, because rightness does transfer. This is in accord with a commonsensical consequentialist slogan, that our moral obligations boil down to one single obligation, the obligation of ‘doing the best we can’, one implication being that we should not be doing *better* than we can.⁹⁸

To be clear, it might have been the other way around. That Molly choosing not to torture people somehow had better consequences than taking the motive-altering pill (which saves so many people from horror), and she *could not* take the pill if she refrained from torturing. In *that* case she ought to refrain from torturing people. This is exactly like the case where you have to choose between torturing someone and letting millions of people die. There exists some number of people you would have to torture high enough so that it would be worse than letting millions of people die, and if so a consequentialist will tell you to let them die instead. These are extreme cases of course, but the point is that as long as the conjunction of two acts is impossible—or an act and a motive, an act and a rule, a motive set and a decision procedure, or any other set of focal points—then the conjunction is not obligatory.

I take it to be an open question whether there are any sets of motives that make it truly impossible to do something that would otherwise be possible, though Parfit certainly thinks such motive sets exist, but the point is that if they don’t exist, then it amounts to the perfect outcome for the global consequentialist. It would truly be the final nail in the coffin of evaluative conflict, because it would mean that Molly simply ought to take the pill and not torture people.⁹⁹

It remains an open question whether such act-restricting motives exist, but I am nonetheless inclined to believe that they do. I am of the opinion, that while we think of the external world as containing both liberty and the lack thereof, we tacitly assume that our minds are either winged or shackled. However, why should not what is true about the world be true of our minds. If being paralyzed restricts us from certain actions, why could not some

⁹⁸ Fred Feldman, *Doing the Best We Can*, ix–xi.

⁹⁹ This might mean that the clause in *My Malevolence*, which states that she *will* go around torturing people, is wrongheaded since there are no motives that can guarantee this.

psychological inabilities exist? The mind resides, after all, in a physical thing. There is a tendency to fear the prospect of an unfree mind so much that we run to the opposite extreme. But if Jenny is unfree in the sense that she cannot bring herself to slaughter her true love, it in no way *follows* that she cannot bring herself to, say, give up chocolate for lent. Consequently, there may or may not be unseen barriers and bulwarks in our psyche.

This reply assumes that OIC *applies* in cases like these, that is, it assumes that Molly's pill involves the sort of impossibility that is relevant for OIC. Parfit does not think so, as we shall see in section four, but before we tackle his argument there is a final point that should be made clear, concerning whether this reply truly amounts to rightness being 'transferable'.

2.3 A 'Transfer' of Rightness?

Even if we assume that there are sets of motives which will make Molly unable to refrain from torturing people, and that the pill in *My Malevolence* gives Molly one of these sets, and that Molly ought to take the pill, *and* we make the assumption that this makes torturing right because of the ought 'implies' can 'rule', then it still seems rather odd that this would be called a 'transfer' or 'inheritance' of rightness.

What happens in *My Malevolence*, according to Mason's reply, is that a change in motivation alters Molly's set of available acts and thus alters what she ought to do.¹⁰⁰ This is the only sense in which rightness is transferred. She ought to do something she did not before taking the pill, but only because her set of available acts changes. But I think this shows it to be a bit of a misnomer to call it a 'transfer' or an 'inheritance'. What happens is that her *circumstances* change, and when circumstances change, what you ought to do often changes too. The implication is not that torture somehow *acquired* some of the rightness that belonged to the motives. This can be shown with an example based on *My Malevolence*.

Margie's Button Pushing: Let's say Margie is a brilliant doctor who could and would easily cure cancer if she only had a few years to work on it, and assume that this is the best thing she could possibly do with her life (except for accepting a demon's deal). And let's say the universe has been taken over by a jealous demon who wishes to be worshipped as a great and kind god, but who (for some reason) wishes her to be ridiculed by her colleagues. To that end, the demon offers her a deal in which Margie either (1) presses a button which will remove all of her memories from medical school and medical practice for the rest of her (mortal) life, as well as make it impossible for her ever to learn these skills again. If she presses the button, the demon will create a state of affairs in which everyone will be happy (though Margie can still make people marginally happier even after pushing the button). Or (2) if she does not push the button, the demon will make everyone in the universe miserable for all eternity. Only the acts

¹⁰⁰ Elinor Mason, "Against Blameless Wrongdoing," 294.

of pushing or not pushing are relevant: If Margie pushes, the demon will make everyone happy, regardless of what her later actions are (i.e., even if she pointlessly tries to cure cancer, or if she does the best possible things available to her).

Pushing the button is obviously the right thing to do, and Margie's subsequent actions may also be right if they are value maximizing. If she *does* do some act that is value maximizing after pushing the button, then that act is right by virtue of being the best available act at the time, *even though it is worse than curing cancer*. This act has not acquired its rightness from the act of pushing the button. Rather, it is right in virtue of being the best act available to Margie at the time, which is the normal way for acts to become right. It is unclear to me why anyone would wish to say anything different about *My Malevolence*, so I will use the term 'transfer' of rightness in what is, strictly speaking, simply shorthand for this process of circumstances changing. That is, as something entirely straightforward, commonplace and non-mysterious.

2.4 Intermediate Conclusion

Given these considerations, I think we are in a position to say that merely brandishing *My Malevolence* as a counter-example to the transferability of rightness is neither sufficient nor particularly convincing. There is a commonsensical explanation for how torture becomes the right thing to do. Since it can be explained in a straightforward manner, it is no longer enough to say that it seems counterintuitive in a set of cases; you have to give some reason why the explanation does not hold.¹⁰¹ In this case you have to show why it would be wrong, *from a consequentialist perspective* even though it is the *best* thing you can do. Parfit has an argument to this effect that I shall turn to in chapter four. But first I should clear up how this defense squares with global consequentialism.

¹⁰¹ Of course, one common argument against consequentialism is that it does not fit with our intuitions. I am not claiming that this argument has no merit, merely that if you are a consequentialist it should not be any more convincing than other cases where consequentialism leads to what a deontologist, for example, would see as wrong.

3 A NEW CRITERION OF ROLE-BASED OBLIGATION

3.1 The Criterion

This defense, for which I have built using Mason's argument as a template, tells us that jointly impossible evaluative focal points cannot *both* be obligatory.¹⁰² But given Ord's formalization of role-based global consequentialism, we *do* end up with jointly impossible prescriptions. His version of global consequentialism does not allow for this defense, and this needs to be addressed.

Julia Driver thought the main problem with evaluative conflict is that it tells you that you ought to *x*, and you ought to *y*, but you cannot do both, so if global consequentialism allows for evaluative conflict is not *action guiding*.¹⁰³ But another way of seeing the problem is that, according to global consequentialism, you are unable to do all that you ought, which breaks with OIC. You will either be wrong in not following the best decision procedure or by doing an act that is not optimal compared to other acts available to you.

Mason's defense solves the second problem, but as a result, Ord's formalization does not properly become action guiding. In order to make this clear, consider what he would say about *My Malevolence*.¹⁰⁴ Ord thinks that the problem with evaluative conflict is that it breaks with the principle:

Rejected Principle: "If *s* ought to *X*, and cannot *X* without also *Y*-ing, then it is permissible for *s* to *Y*."¹⁰⁵

He does not think that the principle is sound, even though it appears reasonable. He gives a reason for this that relies on a specific version of *actualism* that Jackson and Pargetter advocate, but I will give a much simpler reason not to accept it.¹⁰⁶ As it stands, the principle tells you that if *S* ought to *x*, it is *always* permissible for *S* to *y*, if *y* is necessary for *x*. This is much too permissive. To see this, consider that it entails the following:

¹⁰² It is mainly her defense; all I have done is show that her defense is isomorphic to how classical consequentialism works, which I think increases its plausibility. I shall call it Mason's defense from here on.

¹⁰³ Julia Driver, "Global Utilitarianism," 175.

¹⁰⁴ He mentions *Hooker's paradox*, but the rejected principle is equally problematic for *My Malevolence*.

¹⁰⁵ Toby Ord, *Beyond Action*, 92.

¹⁰⁶ Frank Jackson and Robert Pargetter, "Oughts, Options, and Actualism," in *The Philosophical Review* 95, no. 2 (1986): 235. I shall return to what actualism and possibilism is in section 4.3, so let's stick a pin in that for now.

Principle Entailment: If S ought to x , and cannot x without also y -ing, then it is permissible for S to (y and $\neg x$).

In other words, if S ought to save a village, and in order to save the village she has to blow up a guard post with some guards in it, then it is permissible to blow up the guard post and *not save the village*. But if she just blows up the guard post and then watches the village burn, then blowing up the guard post was surely impermissible.

If we wish to make the principle robust we need to add some clause, for example that S actually *does* x .

Possible Principle Extension: If S ought to x , and cannot x without also y -ing, then if S does x it is permissible for S to y .¹⁰⁷

This attempt at an extension of the principle might seem strange; it demands that you actually succeed. If you merely try your very hardest to save the village after blowing up the guard post, yet fail, it is not necessarily permissible to y . This may or may not be plausible, but it should make it clear that we need some way to adjudicate when it is permissible to accept a subpar decision procedure (e.g.) in order to do the right thing (e.g.).

I think there is a good way of doing this which does not lead to any additional problems. I said that given Ord's formalization we do end up with evaluative conflict. That is because he gives us a criterion of obligation which is incompatible with Mason's reply. But if we alter his formalization a little, we get a formalization that *is* compatible. First let us revisit Ord's original formalization of role-based obligation.

Original global consequentialist criterion of role-based obligations

s ought that x be in role r iff x can be in role r ; for any other thing that can be in role r , x is better than it in role r ; and it isn't better to have nothing in role r .¹⁰⁸

This is not in itself sufficient to settle whether Molly ought to take the motive-altering pill or refrain from torturing, therefore it is not action guiding. Having the pill-induced motives can be in role r , for any other set of motives that can be in that role, it is better than that set of motives; and it clearly isn't better to have *no* motives, so this criterion tells us that S ought to take the pill and have the subsequent motives. *But the same goes for not torturing people*. It is

¹⁰⁷ There might be other exceptions here, for example if S tries to save the village but fails, then killing the civilians may also be permissible. But surely if S actually does x then x is permissible in this case. So this criterion is sufficient for permissibility, but I am not claiming it is necessary.

¹⁰⁸ Toby Ord, "How to Be a Consequentialist About Everything," 16.

also possible, and better than any other act that she could do. (Although, it is impossible not to do *anything* so the last clause does not apply to it.) In other words, this formalization tells us that *S* ought to take the pill and not torture people, and it is silent on what to say if she cannot do both.¹⁰⁹ It is open to a much clearer accusation of evaluative conflict than if we remove this possibility.

Just to be clear, it is (perhaps) possible that what Ord has in mind when he writes this is for the ‘can’ clause to be interpreted in a way that does not allow for vagueness, but if so I would say that it has been presented somewhat misleadingly, given the strong focus of bestness *within* a role. If this is what he has in mind, it still cannot hurt to hone the criterion somewhat, so that it resonates more clearly with Mason’s defense. It might also be that he is unperturbed by evaluative conflict, given the fact that he thinks the rejected principle breaks with actualism.

Furthermore, it would be possible to squabble about what Molly ‘can’ at t_1 , before she takes the pill, and what she ‘can’ at t_2 , after taking the pill. So you might want to say that at t_1 , Molly ought not to torture people at t_2 , but when t_2 actually arrives, Molly ought to torture people (if we assume it is the best she can do). This might depend on what it means that ‘you will’ go around torturing people, and much else. There are several fancy ways of carving out this *dated* notion of rightness, and this certainly ought to be considered in relation to global consequentialism, but not by me. Evaluative conflict is what I am addressing, and it is (supposedly) problematic because it tells you that you ought to not torture people even *at* t_2 , simply because you could have been able to torture people if you had not taken the pill. I shall focus on this aspect. Holly Smith provides an excellent account of how these questions relate to one another, and how such dated rightness might be incorporated into both consequentialism and deontology.¹¹⁰

I think there is a reasonable alteration to Ord’s formalization that could make it work with Mason’s reply. First we *remove* the clause that x needs to be better than any other thing that can be in role r . Then, we add the clause that if x is in role r , then outcomes are better than if x is not in role r . This yields the criterion:

New global consequentialist sufficiency criterion of role-based obligations

¹⁰⁹ In section 5.1 I will show how Mason’s reply can work for *Hooker’s Paradox*.

¹¹⁰ Holly S. Goldman, “Dated Rightness and Moral Imperfection,” *The Philosophical Review* 85, no. 4 (October 1976): 449–487.

S ought that *x* be in role *r* iff *x* can be in role *r*; it isn't better to have nothing in role *r*; and if *x* is in role *r*, outcomes are better than if *x* is not in role *r*.¹¹¹

This does not allow for ties, so the criterion is merely sufficient, not necessary for obligation. So we have to add a separate clause that says that if there are ties then all are permissible, and one, no matter which one, is obligatory. But I will not give one here, since it is not really pertinent to my overall case.

We have to make a similar change to the global consequentialist criterion of rightness within a role. First we take away the clause that for each other *e* that can be in role *r*, *x* is better than it in *r*. Then we add the clause that if *x* is in role *e*, outcomes are not worse than if *x* is not in role *e*. This is needed because the term 'right' is ambiguous between meaning *obligatory* and *optional*.¹¹² With the changes to both criteria, any token within an evaluative focal point will be merely right if it is equally good as another, but obligatory if it is better than all others. This yields the new criterion:

New global consequentialist criterion of rightness within a role

x is the right *e* in role *r* iff *x* can be in role *r*; and if *x* is in role *e*, outcomes are not worse than if *x* is not in role *e*.¹¹³

With these changes we can assess the rightness of disparate focal points and also incorporate Mason's reply to evaluative conflict. If we follow Ord's wording we will get conflicting prescriptions since both accepting/following the decision procedure and doing *y* are obligatory, as there are no other acts or decision procedure that will be better *in that role*, yet they are jointly impossible. With my wording, however, we are free to say that if you cannot do act *y* as long as you follow the best decision procedure, then you may have no obligation to do act *y*. Whether you should follow the decision procedure or not depends on whether following the decision procedure and *not* doing act *y* has better consequences than following a worse decision procedure but also doing act *y*. That is, the new clause *includes* all the causal restraints that one focal point can put on others. If you cannot follow the decision procedure

¹¹¹ This way of spelling out the criterion does not say anything about the subjective/objective rightness distinction, since it simply states that "if outcomes are better" and the other clauses are fulfilled, an obligation is generated. It does not say what it means by 'outcomes are better'. I am not going to delve into this distinction, except to say that we just have to make separate criteria for each sense of rightness that specify whether it is expected utility or actual utility that is the basis for the obligation.

¹¹² Mark Timmons, Introduction to *Conduct and Character: Readings in Moral Theory*, 6th ed. (Boston: Wadsworth Cengage Learning, 2006), 3.

¹¹³ Toby Ord, "How to Be a Consequentialist About Everything," 16.

and do act y then it might be that following the decision procedure without doing y will have better consequences than doing y and not following the decision procedure, and *vice versa*.¹¹⁴

In other words, if following the best decision procedure precludes you from doing the best act, and if doing the best act has better consequences all things considered than following the best decision procedure, then you ought to *not follow that decision procedure since outcomes will be worse*. This is just like act consequentialism, which does not expect that if you have a choice between doing act x and act y , you should do both.

3.2 Is This Still a Form of Consequentialism?

Making the changes mentioned in the previous section means that global consequentialists give up the idea that the bestness of a token within a category of focal points always means that this token is right. That is, sometimes an act you can do that is better than all other acts you can do, is still not obligatory. The explanation for this is that doing the act will have *worse* consequences than not doing it, as long as you have the motives (for example) that you ought to have, since these motives and another set of acts will have better consequences in sum. This will be the case for Molly before she takes the motive-altering pill for example. It is possible for her at that time to not torture people in the future, but this will have worse consequences than if she does torture people because of taking the motive-altering pill. Still, giving up the notion that bestness means rightness might be seen as contrary to the spirit of consequentialism, or even to be something else entirely. Louise considers this in a footnote.

A reviewer has suggested that the claim that a non-best alternative is (nevertheless) right may be compatible with the view that what ultimately matters is neutral value maximization, if it is also held that logically possible states of affairs ought to be ranked according to the amount of neutral value they realize. However, I do not see how this would count as a consequentialist view. Such a view would retain part of the characterization of consequentialism (namely, the definition of bestness in terms of neutral value maximization); however, the other part (the definition of rightness in terms of bestness) is lost. If rightness is not determined by bestness, then it seems odd to claim that bestness is what *really matters* morally.¹¹⁵

But there are good reasons to discard this objection. First of all, it is far from obvious that this criterion of role-based obligation has given up the definition of rightness (or obligation) in terms of bestness. It is true that the best act to do at t_2 , compared within the category of acts that *were* possible at t_1 , is not necessarily the right act. But it is still the case that the best

¹¹⁴ In this case, it seems that you should always follow the best decision procedure. Since the best decision procedure necessarily has equal or better consequences as/than doing the best act and following some other decision procedure. After all, if some act alone has better consequences than an entire decision procedure, then it seems to follow that you should follow some decision procedure that allows for that act to be done.

¹¹⁵ Jennie Louise, "Right Motive, Wrong Action," 73.

token compared across *any* evaluative focal point is right. If doing the best act you can do is only possible if you accept a suboptimal decision procedure, then that act may not be obligatory. But it will always be the case that the best of either an act, decision procedure, set of motives, etc. that is available to you, is obligatory, since this will maximize value *across* all focal points. I would say this plainly counts as defining rightness in terms of bestness, and if so bestness is what really matters morally. I would even go so far as to say that this counts as defining rightness in terms of bestness even *more* clearly than other forms of consequentialism, they are open to the worship objection.

Second, this means that even with my changes, global consequentialism is *foundationally* consequentialist. It is foundationally consequentialist because this conforms to the requirement that what really matters morally is that outcomes are as good as possible.¹¹⁶ If it is foundationally consequentialist it is presumably also a form of consequentialism.

Third, does it *matter* if it is a form of consequentialism as long as it is *foundationally* consequentialist? I cannot see how. If it is not then here is a view that resembles consequentialism immensely, yet is not consequentialism. Call it ‘schmonsequentialism’ if you will, the question would be whether the view is consistent and whether it solves or creates any new problems. All I can see is that it solves problems, namely the problem of evaluative conflict.

3.3 Is This Still a Form of *Foundational* Consequentialism?

I would give similar answers if someone were to object that even though this amounts to consequentialism, it does not amount to *global* consequentialism. Given that global consequentialism has typically held that the best act is the right act *and* the best decision procedure is the right decision procedure, etc. this objection might hold more sway though. Since my way of formalizing global consequentialism is ‘global’ in a different way than other types of global consequentialism, it is not entirely clear cut that this is a form of global consequentialism. But, just like it does not really matter whether you put it under the rubric of consequentialism, it does not really matter whether you put it under the rubric of global consequentialism. If it is not then we can call it ‘galactic consequentialism’, but for the remainder of this thesis I am going to assume that it is a form of global consequentialism, and whenever I refer to global consequentialism, I will be referring to global consequentialism including my alterations.

¹¹⁶ Derek Parfit, *Reasons and Persons*, 25–26.

In the following section, however, we will see that Parfit has an argument that makes both Mason's reply and this way of formalizing global consequentialism significantly less attractive. If it becomes necessary to abandon Mason's defense, we may also have to either simply *accept* evaluative conflict, or give up on global consequentialism entirely.

4 I MUST, BUT I OUGHT NOT

4.1 Parfit and OIC

Parfit argues that OIC does not *apply* in cases like *My Malevolence*. According to him, Molly's inability to refrain from torturing is not the right kind of inability. Hers is not the same *sense* of 'cannot' as that which is involved in the principle.¹¹⁷

On this view, Molly has the right motives and this makes it impossible for her not to torture people, yet she still does the wrong thing in torturing people since it is irrelevant that she is unable to refrain from doing so. Alternatively, if she refuses to take the pill, she may still *do* right things (whatever she does instead of torturing people), but she has the wrong motives. In other words this would allow evaluative conflict.

Parfit focuses on his example *My Moral Corruption*, but as I mentioned earlier he admits that this example is flawed, so I will continue to utilize *My Malevolence* which was constructed by Louise to be parallel, yet negate the faults of the former.¹¹⁸ I will assume that Parfit would argue identically in the two cases.

According to Parfit, the following defense fails,

Defense: Molly could have refrained from torturing people, but she could only refrain if she did not take the pill, and this would be much worse overall. So, she cannot both have the best motives and do the best act (compared across motives). She has to choose, and she ought to choose the pill.

Parfit argues that it fails because

This reply assumes one view about voluntary acts: *Psychological determinism*. On this view, our acts are always caused by our desires, beliefs, and other dispositions. Given our actual dispositions, it is not causally possible that we act differently.¹¹⁹

He thinks that Clare's defense presupposes that, given our actual motives and dispositions, there is only one possible resulting act (in a particular situation). Crucially, he goes on to say that if we also interpret OIC in a way that lets Clare appeal to it with this line of reasoning, it results in wrongness becoming obsolete; that nobody can ever do anything wrong.

¹¹⁷ I am assuming he would say the same thing about *My Malevolence* as he would about *Clare*. They are slightly different, most importantly because Molly takes a pill that alters her motives, whilst Clare does not. Still, his argument applies to both cases, and I will assume that Parfit would not find this to be a significant difference.

¹¹⁸ Derek Parfit, *Reasons and Persons*, 509.

¹¹⁹ *Ibid.*, 14–15.

On the [OIC] doctrine, I ought to have acted in this other way only if I could have done so. If I could *not* have acted in this other way, it cannot be claimed that that this is what I ought to have done. The claim (1) that I could not have acted in this other way is not the claim (2) that acting in this way would have been impossible given my actual desires and dispositions. The claim is rather (3) that acting in this way would have been impossible, even if my desires and dispositions had been different. If claim (1) was claim (2), determinists would have to conclude that it is not possible for anyone ever to act wrongly or irrationally. They can justifiably reject this conclusion. They can insist that claim (1) is claim (3).¹²⁰

In his later book *On What Matters* he calls claim (3) the *hypothetical* or *iffy* sense of ‘can’ or ‘cannot’, and he calls claim (2) the *categorical* sense of ‘can’.¹²¹ Claim (3) is hypothetical because it states that the relevant sense of ‘can’ is that you could do something if you had different motives, that is, in a hypothetical situation, (2) is categorical because it tells you that the relevant sense of ‘can’ is that you ‘can’ do something in your actual circumstances, so the categorical sense just means what is *actually* possible. Both types of claims can be true or false, of course, but the categorical sense of ‘can’ is much stricter. I shall mostly be using this terminology because it is more mnemonically helpful.¹²²

In embracing this hypothetical sense of ‘can’, he does not commit himself to the view that free will and determinism are compatible, nor does he commit himself to the view that they are not. As he himself puts it, “the sense of ‘can’ required for free will may be different from the sense of ‘can’ in the doctrine that ought implies can”.¹²³ So, he thinks determinism might be true without OIC rendering wrongness obsolete. His point here is that if you assume determinism *and* the categorical sense of ‘can’, then wrongness becomes obsolete, so we should reject the categorical sense.

Also, Parfit focuses on *psychological* determinism, not physical determinism. But the most credible reason for being a psychological determinist is that you believe in physical determinism.¹²⁴ I will assume that Parfit mentions psychological determinism because of

¹²⁰ Derek Parfit, *Reasons and Persons*, 15.

¹²¹ Derek Parfit, *On What Matters vol. 1*, (Oxford: Oxford University Press, 2011): 260.

¹²² I shall also be writing as if OIC is consequentialist, so I will assume that OIC demands that we do *the best we can* for example, even though a requirement of bestness is not really integral to OIC as such. Since both the philosophers I refer to and I are focusing on consequentialism, this makes things much easier to state.

¹²³ Derek Parfit, *Reasons and Persons*, 15.

¹²⁴ That is, determinism on the macro level (including, of course, our brains). You could still believe that randomness exists on the quantum scale but that it has no effects on how our brains function.

physical determinism until section 4.9. He seems to suggest that this is his reason for mentioning it in *On What Matters*, where he is discussing physical determinism.¹²⁵

Parfit's interpretation of OIC blocks Mason's reply to evaluative conflict, which argues that you either *can* do better, in which case you ought to simply do so, or you cannot, in which case you will not be doing anything wrong in doing whatever the best you can is. Parfit blocks her reply because it relies on the categorical sense of 'can'; the sense in which Molly 'cannot' refrain from torturing after taking the motive-altering pill is categorical and not hypothetical. That is, if she had other motives or dispositions she *could* refrain from torturing, and according to Parfit's argument that is insufficient for appealing to OIC.

Mason considers this. Essentially she replies by saying that even if acts are causally determined by motives, the choice of motives need not be determined, and so we can accept the categorical sense of 'can' without wrongness becoming obsolete.

Parfit is under-motivated to his account of available alternatives. A determinist can still claim that an act is available (possible) only if it is possible given the actual dispositions an agent has, without making wrongness obsolete. The wrongness is just shifted to the level of choosing a disposition. The determinist can say that an agent acts wrongly when she chooses to cultivate the wrong disposition.¹²⁶

This is an insufficient reply, as Louise argues,

If the agent's *acts* are involuntary because caused by motives (or dispositions, as in Mason's account), then it seems that the agent's decision to cultivate a disposition will likewise be involuntary, since this decision will be caused by motives or dispositions the agent already has. (Indeed, Mason specifically describes such choices as acts, which brings them obviously under the scope of her argument).¹²⁷

Mason's reply falls prey to Parfit's analysis of OIC because she categorizes choosing to cultivate the wrong disposition *as* an act, and she gives no arguments to the effect that *acts* do not fall prey to Parfit's argument.

But, even though choosing to cultivate certain motives were somehow *not* an act, it is difficult to see how Mason's line of thought is supposed to help. If the cultivation of a set of dispositions or motives is not an act, then they do not fall directly under the scope of Parfit's argument, but at what price? You are left with a strange situation in which acts are determined by motives, but motives are not determined by anything. On what plausible view do you get this result? It appears rather unmotivated and *ad hoc*. As mentioned above, the most credible

¹²⁵ Derek Parfit, *On what Matters*, 258–261.

¹²⁶ Elinor Mason, "Against Blameless Wrongdoing," 291.

¹²⁷ Jennie Louise, "Right Motive, Wrong Action," 78.

reason for believing that acts are psychologically determined, is that the physical world is determined (with a possible exception for quantum mechanics). I shall consider the alternative in section 4.9, that is, what we should think if psychological determinism is not a result of physical determinism.

I will not address this issue further at this point; if Mason's line of thought succeeds, there is no problem of evaluative conflict to be found in global consequentialism. Basing OIC on the categorical sense of 'can' does not amount to wrongness being obsolete, and so we can say that Mason's reply to evaluative conflict is reasonable. If Louise is right, on the other hand, then the following sections will show that it is still far from obvious that Parfit's argument succeeds, even though it looks rather devastating. This will become clearer once I go through my arguments for why Parfit's analysis cannot be correct.

4.2 The First Reason why Claim (1) is Not Claim (3)

I will give some reasons why we might think that the categorical sense of 'can' has to be the correct one below, but first I should point out a rather strange upshot of Parfit's claim hypothetical sense, that is, a reason for thinking that claim (1) is *not* claim (3). If this argument succeeds, what Parfit has shown is not that we should accept the hypothetical sense of 'can', but rather that there is a dilemma, with two seemingly bad options, the hypothetical interpretation and the categorical one. If there is such a dilemma, it might serve as a good reason to search for some third option.

If we base OIC on the hypothetical sense of 'can' then the rightness of Molly's acts depends on what might have been, as opposed to what is actually the case. This is not always a problem, but in this case it should give us some pause. That is, sometimes it could make sense to say that rightness depends on what might have been; act consequentialists do this when they compare one act in the past to the other acts that were then possible, and designate an act as right if it was the best of the ones the agent could have done.

Parfit's claim is different. He is suggesting that we break with a fairly fundamental aspect of consequentialism, which is that we ought to base our decisions about what to do on what is possible given the actual state of the world, or at least what we think is the actual state of the world, as opposed to basing them on pure fantasy. In Molly's case, basing a decision on whether to torture people on what would be possible if she had a *different* set of motives than the one she actually has. There is a big difference between basing rightness on what *was*

possible, and basing it on what would have been possible if things were not as they were. Or, in the present, on what is possible as opposed to what might have been possible. From a consequentialist perspective it is downright weird to say you should do the best thing possible in some other possible world than the one we actually inhabit. I consider this the basic intuition behind many of the the arguments I shall be presenting from here on, that is, some of the force comes from how strange this result actually is.

If a standard consequentialist is considering her options, for example how much money she should give to Oxfam, and if we assume that she ought to give some sum of money to Oxfam, that the more she gives the better the consequences, and that she has exactly 10000 dollars, then she ought to give 10000 dollars. It is not true that she ought to figure out how much money she *could have had*, and give that amount, say 100 trillion dollars, because she cannot give that amount. This shows just how unattractive Parfit's solution is. Parfit is telling us that we ought to do something that is literally impossible.

And I believe this (rather large) quirk of Parfit's view leads to an even bigger problem.

Ellie's Brain Scan. The year is 2525 AD, and Ellie has her brain scanned. The doctor tells her that based on her current brain-state, which is the basis for her motives, she will not be psychologically (causally) able to give her money to Oxfam, though if she could do so it would be better than any of the actions that she is able to do.

If we base OIC on the hypothetical sense of 'can', then it follows that Ellie ought to give her money to Oxfam. This should *really* give us some pause. It tells us that she ought to do something she *knows* she cannot do. This seems much weirder than merely telling us she ought to do something that she in fact cannot do, which is already pretty strange. Perhaps the most problematic feature of this conclusion is that it fails to provide adequate guidance. For example, if you know you ought to give money to Oxfam, and you know you cannot give money to Oxfam, what should you do? You could try to give money to Oxfam anyway (perhaps), yet it seems very strange to say that you ought to try to do something you know you cannot do.¹²⁸ You would, for instance, be walking down the street knowing that at some point along the way to the bank, you are going to turn around and do something else. You could try to do whatever the best thing you are causally able to do is, but according to the hypothetical sense of 'can' (the hypothetical interpretation of OIC from here on) this action

¹²⁸ Especially if the reason you ought to do it is that, if you succeeded, it would be the best thing you could do.

will be *predictably wrong* since it is worse than giving the money to Oxfam, and any theory that tells you that you ought to try to do the wrong thing is *seriously* defective.¹²⁹

This becomes even *more* untenable if we assume that by *attempting* to give the money to Oxfam; she will forego an opportunity to do the second best thing, so that by her futile attempt at doing the impossible (yet somehow right) thing, she is actually foregoing the possibility of doing the best *possible* thing. Let us assume that what she *would* have done if she were not to attempt to give it to Oxfam would be to give it to UNICEF, which would be the second best outcome. In this case it actually follows that she ought to give the money to Oxfam, but she ought *not* to try to do so, which would be very strange indeed.

The hypothetical interpretation of OIC *does* pick out an act as right (in tandem with a normative theory). According to it and consequentialism, the right act is giving the money to Oxfam, but that right act is clearly not an *option* in any meaningful sense of the word, and may lead us to miss out on doing the best we can. I believe that this is not a tenable view. However, more complications are going to surface as we proceed, so it might be that there are no good ways to interpret OIC if determinism is true.

Now, Parfit might argue that this example is unrealistic since this kind of brain scans do not exist and likely never will, but remember that the rationale for the hypothetical sense of ‘can’ is that given the categorical sense in conjunction with psychological *determinism*, wrongness becomes obsolete. If psychological determinism is true, the impossibility of these brain scans is merely contingent. It just happens to be the case that we cannot make them, and it is untenable to hold that the merely contingent fact that we cannot scan brains means that the problem goes away.¹³⁰

The assumption of determinism appears to have weird results on both the categorical and hypothetical interpretation of OIC. If both options are bad we should either find a *third* option, which might or might not be damaging to Mason’s reply to evaluative conflict, or accept that the categorical sense having some weird results does not amount to a refutation of it. I will give some reasons for thinking there is a third option below.

¹²⁹ Barring ties, that is. But of course, not all cases are ties, so in all non-tie cases this problem holds. I do not think anyone wishes to defend a theory that only works when there are two equally good options.

¹³⁰ You could deny this by saying that even though psychological determinism is true, perfect knowledge of your brain states do not amount to knowledge of your motives or how they combine to create actions. If this makes sense to you just replace the brain scanner with a motive scanner and add a scientist that knows how these motives combine to form actions, this is equally unrealistic, in an equally contingent manner.

Eschewing third options for a moment, wrongness being obsolete is certainly the more *spectacular* consequence, but that does not mean it is more absurd or less plausible. In fact, there are several strands of philosophy that actively argue for the conclusion moral nihilism; error theory and expressivism are the most common.¹³¹ As far as I can tell, no one has ever argued that something like this could be true, that is, argue that you should rather try to do the best thing you know you cannot do, than the best option you think you *can* do. Still, both of these options are highly unattractive.

4.3 Is Parfit an Ethical Possibilist?

I will get to the second reason why Parfit's solution is not attractive in the section 4.5, but first I should make some clarifications and present some defensive routes available to Parfit. One possible way of looking at his analysis of OIC is to say that what Parfit is espousing is a version of *ethical possibilism*. Consider the following case:

Professor Procrastinate: Procrastinate receives an invitation to review a book. He is the best person to do the review, has the time, and so on. The best thing that can happen is that he says yes, and then writes the review when the book arrives. However, suppose it is further the case that were Procrastinate to say yes, he would not in fact get around to writing the review. Not because of incapacity or outside interference or anything like that, but because he would keep on putting the task off. (This has been known to happen.) Thus, although the best that can happen is for Procrastinate to say yes and then write, and he can do exactly this, what would in fact happen were he to say yes is that he would not write the review. Moreover, we may suppose, this latter is the worst that can happen. It would lead to the book not being reviewed at all, or at least to a review being seriously delayed.¹³²

An *actualist* will say that Procrastinate ought to say no, since if he were to say yes he would not write. *Possibilists*, on the other hand, will say that since Procrastinate *could* say yes and then write, he ought to say yes.¹³³ Possibilism seems at least somewhat similar to holding that we should accept the hypothetical interpretation of OIC. The possibilists are saying exactly that Procrastinate ought to do something, even though that *will* lead to him not writing the review, just like they might agree with Parfit and say that Ellie ought to give the money to Oxfam since she can do so. ('Can' in the sense that she according to the hypothetical sense of 'can' could and would have done so *if* her motives had been different.)

¹³¹ Walter Sinnott-Armstrong, "Are Any Moral Beliefs True?," in *Moral Skepticisms* (Oxford: Oxford University Press, 2006), 32–37.

¹³² Jackson and Pargetter, "Oughts, Options, and Actualism," 235.

¹³³ *Ibid.*

There are some differences between *Ellie's Brain Scan* and *Professor Procrastinate*. There is at least one sense in which Ellie not only *will* not, but *cannot* give money to Oxfam. This is different from actualism/possibilism cases. When talking about *Professor Procrastinate*, for example, Jackson and Pargetter write that “we must not let the mere fact that someone would not do something show that they could not”.¹³⁴ So Jackson and Pargetter seem to think that ‘will not’ is a weaker term than ‘cannot’, but if determinism is true it is causally impossible that it is going to happen, meaning that in at least one sense, she cannot give money to Oxfam. If so it is unclear how much traction Parfit could get out of this line. And, as we see in the case itself, Jackson and Pargetter specify that it is not the case that Procrastinate will not do it *because of incapacity*.

Another difference is that the *Professor Procrastinate* case is diachronic, whilst Ellie has to decide what to do right now. This might be thought to be a relevant difference. Unfortunately for me there are non-diachronic cases that display the actualism/possibilism distinction. The following case was devised by Holly Smith:

Jones is driving through a tunnel behind a slow-moving truck. It is illegal to change lanes in the tunnel, and Jones's doing so would disrupt the traffic. Nevertheless, she is going to change lanes—perhaps she doesn't realize it is illegal, or perhaps she is simply in a hurry. If she changes lanes without accelerating, traffic will be disrupted more severely than if she accelerates. If she accelerates without changing lanes, her car will collide with the back of a truck.¹³⁵

In this case, an actualist would say that Jones ought to accelerate, since she is going to change lanes no matter what, and it is better to change lanes and accelerate than to change lanes and not accelerate. A possibilist would say that Jones *can* stay in lane and not accelerate, and this is even better than changing lanes and accelerating, so she ought to do this. Since this entails not accelerating, she ought not to accelerate.

There are proponents of both sides of this dilemma. And the mere fact that *Ellie's Brain Scan* is non-diachronic does not help my case much. However, there is another difference between the cases that I think shows that *Ellie's Brain Scan* is more problematic than mere possibilism.

Actualism/possibilism cases are usually about what the agent objectively ought to do. E.g. Jackson and Pargetter explicitly state that they are talking about objective rightness in the

¹³⁴ Jackson and Pargetter, “Oughts, Options, and Actualism,” 240.

¹³⁵ Holly S. Goldman, “Doing the Best One Can” in *Values and Morals*, ed. Alvin I. Goldman and Jaegwon Kim (Dordrecht: D. Reidel Publishing Company, 1978), 186.

Professor Procrastinate case.¹³⁶ Parfit, on the other hand, seems to be talking about subjective rightness. He mentions early in *Reasons and Persons* that “in most of what follows, I shall use *right*, *ought*, *good* and *bad* in the objective sense. But *wrong* will usually mean *subjectively wrong*.”¹³⁷ In the cases that are relevant to us, he is talking about wrongness. Also, he often adds he is talking about what the agent *believes* to be wrong, which indicated that he is talking about subjective rightness.¹³⁸

Ellie's Brain Scan is most problematic when we are talking about subjective rightness. The fact that she *knows* that she will not give the money to Oxfam makes it seem straightforwardly irrational to attempt to do it. It would mean that seen from Ellie's perspective at the time she is making the choice, she subjectively ought to give her money to Oxfam, while knowing she will not do so. If this is supposed to tell her something about how to act it eludes me. Proposing possibilism does little to mitigate this since actualism/possibilism cases are really about the objective sense of rightness.

To make this clearer we can imagine what a subjective version of possibilism would look like. All we need to do is take *Professor Procrastinate* and add the fact that Procrastinate *knows* he will not write the review if he says yes. If what we are after is what he subjectively ought to do, then the answer is clearly not to say yes, since there is a 0% chance of it being the better option, and the expected utility then drops below the alternative (to say yes).

4.4 OIC and Subjective Rightness

It seems like we have to say that given determinism and that the relevant sense of rightness is subjective rightness, Parfit's hypothetical sense of 'can' will not work. I shall return to objective rightness shortly, but first we need some way of telling what is subjectively right in a way that incorporates something like OIC. Luckily there are some ways of doing this that have been proposed in the literature which don't have the same problems that Parfit's solution has.

One possibility—proposed by H.A. Prichard—is that we 'can' do something in the sense relevant for OIC, only if we think that setting ourselves to bring about something would

¹³⁶ Jackson and Pargetter, “Oughts, Options, and Actualism,” 236.

¹³⁷ Derek Parfit, *Reasons and Persons*, 25.

¹³⁸ *Ibid.*, 25, 35–40.

make it likely that it actually happens.¹³⁹ This would have two interesting consequences for a global consequentialist. First of all, it would mean that *Ellie's Brain Scan* has no traction. Ellie believes it impossible that giving the money to Oxfam would happen if she set herself to do so, and so she clearly does not think it likely. Therefore, subjective global consequentialism does not prescribe that she set about doing so according to Prichard's version of OIC.

Second, it would mean that Mason's reply is available to the global consequentialist. If this is the sense of the 'can' in OIC, then wrongness is *not* obsolete even if determinism is true, since it is not what is actually causally possible that is relevant given this sense of 'can', but rather Ellie's *beliefs* about what is possible. Assume (once again) that she is predestined not to give money to Oxfam. Provided she believes that setting herself to give money to Oxfam would likely result in her giving money to Oxfam, and that this is the best thing she could do, then doing so will be right. When she does *not* do this, she will be doing the wrong thing because she could have given the money to Oxfam (in the relevant sense of 'could').

But there are also times where she *can* appeal to OIC. It is, after all, not always the case that you believe that setting yourself to do something will bring it about. This means that a global consequentialist can appeal to OIC if she believes that trying to *x* will not result in *x*-ing, if so it follows from OIC that it is *not* subjectively right. (This also holds if she simply does not believe that trying to *x* will result in *x*-ing.) On this interpretation, Ellie subjectively *cannot*, in the relevant sense, give money to Oxfam and so she *can* appeal to OIC.

Saying that she 'can' *x* is somewhat weird in this case of course, since it is causally impossible. But since we are considering subjective rightness, it cannot be the case that the meaning of 'can' is 'it is predetermined that it will happen', even though the predetermined is the only causal possibility if determinism is true. This is because we are sometimes unaware of what is possible and what is not, and it makes no sense to say that *S* subjectively ought to *x* even though she believes it is impossible. Subjective rightness is a concept whose primary motivation is prescriptive, and the *mere fact* that something is not actually going to happen in no way helps her to know what her options are. Only what she believes and knows about what is going to happen is important for subjective rightness. In other words, if *S* believes that *x* is

¹³⁹ H.A. Prichard, "Duty and Ignorance of Fact," in *Moral Writings*, ed. J. MacAdam (Oxford: Clarendon Press, 2002), 95–98.

best and that it is possible, then *S* seems to be rational in *x*-ing. But if *S* believes that *x* is best but *impossible*, she appears *irrational* if she attempts to *x*.

Mason's reply to evaluative conflict is still viable on such a view, since you could say that what you cannot do (in Prichard's sense) is not a relevant option, and so only what you actually believe you can do counts. If we assume you have the right motives, you subjectively ought to do whatever you believe is best and possible given those motives, which in normal cases (that do not include knowledge of what is possible with your motives) will mean all the acts that seem physically possible. In other words, whenever you have information that tells you that you will not *x* given your motives, it will not be the case that you ought to *x*. That is, it will never be the case that you subjectively ought to *x* but you cannot (in the relevant sense) *x*.

Another way of seeing the same point is that Prichard's version of OIC tells Molly, from *My Malevolence*, that it is right for her to torture people after taking the motive-altering pill, as long as we assume that she *knows* that she cannot refrain from doing so after taking the motive-altering pill, which is a reasonable assumption since she *chooses* to acquire the motives and the demon tells her that the pill will make her torture people. But if she believes that she can choose not to torture people, but still does it, then she does the wrong thing subjectively, because she believes she could have done something else.

However, there is a special case that opens up. What if Molly's memories are also erased by taking the motive-altering pill? Then it seems like it will still be the case that she subjectively ought to acquire the torture-inducing motives, and because her memories are erased she may believe she can refrain from torturing people afterward. If so she subjectively ought to refrain, even though she causally cannot do so, and this sounds like evaluative conflict. She ought to *x* and she ought to not *y*, but she cannot do both. Still, I don't think this objection makes much sense at this point. Yes, it is true that this leads to a form of subjective evaluative conflict, but this is a result of the fact that we have given up the notion that what is causally impossible also has to be subjectively neutral *generally*, which means that there are several causally impossible things that you ought to do, yet only subjectively. In other words, this type of evaluative conflict is just a particular case of causally impossible yet subjectively obligatory actions. We have to remember that we are assuming determinism, and some weird results are to be expected. The alternative is that wrongness becomes obsolete, since any

theory that sets causal possibility as the relevant sense of can will have only one possible outcome, which will obviously always happen.

As with Parfit's claim that we ought to think of 'can' in his hypothetical sense, Prichard gives an answer that will sometimes prescribe an action that is causally impossible given determinism. That is, sometimes we *literally* cannot do what Prichard's principle tells us to. This is what I professed earlier that I did not want, but this, it seems, is going to be true of any solution. If determinism is true there will always be one single causally possible act, so any theory that ever prescribes anything else than that one act will necessarily prescribe something literally impossible. I would wish for a solution that avoided this, but it seems I am out of luck in that regard.

Still, I don't think this is as bad a solution as Parfit's. Both solutions tell us to do something we literally cannot do, but we are considering *subjective* rightness, and the fact that you cannot *causally* do something might not be all that damaging, as long as you *believe* that you can. If you yourself think that you can do otherwise, and you yourself think that what you are going to do is wrong, perhaps that is all that is needed for subjective wrongness. Parfit, on the other hand, tells you that even if you *know* you cannot do something, you still subjectively ought to do it, and that is what I find so hard to believe. The literal impossibility of doing anything but what you are going to do is an inescapable part of hard determinism, this second problem is not. If so, then *Ellie's Brain Scan* is just a specific instance where, for once, you are able to know in advance that you are not going to do something.

Another possibility is proposed by Frances Howard-Snyder. She believes that in order for someone to be able to do something they have to know *how* to do it. For example, I may be physically able to beat Magnus Carlsen at chess, all it would take is moving the pieces around the board in the correct order, but I don't know *how* to beat Magnus Carlsen at chess.¹⁴⁰ Subjectively then, according to Howard-Snyder, I would not have an obligation to beat him, even if that were somehow the best possible action available to me, because I cannot do so in the relevant sense of 'can'.

But consider Ellie, she presumably cannot know how to give money to Oxfam if she cannot causally give money to Oxfam. She could know how-one-would-go-about giving money to Oxfam, of course, but since she knows that she is not causally able to do so, she

¹⁴⁰ Frances Howard-Snyder, "The Rejection of Objective Consequentialism," *Utilitas* 9, no. 2 (1997): 245–246.

must also know that this general approach would not work in her case, and presumably what is needed for her to count as subjectively able (according to OIC) to give to Oxfam is for her to know how *she* would go about doing so, not how someone else would go about doing it.

If this is how we should understand Howard-Snyder's claim that you are unable to do something in the sense relevant to OIC if you do not know how to do it, then we may have to conclude that wrongness is obsolete yet again. It may be that the *only* thing you could know how to do in this latter sense would be what you are causally able to do, and there is, according to determinism, only one thing that you are causally able to do. If so we seem to have to conclude that wrongness is obsolete after all.

Now, even though we would always be causally able to do what we know how to do, it does not immediately *follow* that we always know how to do what we are causally able to do. The case can be made that we don't always know how to do what we are causally able to do. If I am asked to throw a quarter into a cup ten feet away it may be determined that I will succeed. But it is at least not obvious that I know *how* to do so. I may be able to loosely describe how to throw the penny in a way that will send it in the direction of the cup, but I am not able to give a complete description of how (the way every muscle moves, etc.). In other words, it would seem that I have knowledge-how to do it in the sense that I will be able to do it this once, but I don't necessarily have knowledge-how in the sense that I would be able to repeat it at will.

Perhaps this is an indication that knowing how to do something in this context should be interpreted differently; perhaps I can know how to do something even though it is not causally possible for me to do so. This would be more reminiscent of knowing how-one-would-go-about doing something than knowing how you yourself would go about doing something. If so, a global consequentialist could appeal to OIC if she did not know how-one-would-go-about doing something that she would otherwise have an obligation to do.

This means that (presto!) wrongness is not obsolete. If you have the right motives, then the relevant options for you are the ones that you know how-one-would-go-about doing, and the best of those would be right. Not doing this will be wrong and so there are wrong acts. This also means that Mason's reply works, since you will never be in a situation where you cannot do (in the relevant sense) the right thing given your motives. Although, it seems to me

to have the air of a magic trick, as opposed to that of a proper explanation of OIC in light of determinism. I therefore prefer Prichard's solution.

What happens if we apply this to Molly and the case with the motive-altering pill, *My Malevolence*? If the expected sum value of having another set of motives, and that means you can do some better act which has a value that is *greater* than the sum value of your current motive set, then you ought to have that motive set instead. If, on the other hand, the expected sum value of another set that allows you to do some better act is *not* greater than your current set, then you should have your current set. More generally, if the expected sum value of *any* focal point (set of motives, acts, rules) including all the constraints it puts on other focal points, and compared across focal points, has the greatest expected sum value, then that focal point is right. So, Molly should take the pill to acquire malignant motives, and if she knows how-one-would-go-about not torturing people she should do so, if she does not know how-one-would-go-about not torturing people after taking the pill then she can appeal to OIC. All of this is predicated on the assumption that taking the pill means that the malignant motives are the right motives though, if we remove the evil demon from the equation then these motives are wrong and she should rather have another set of motives and do another set of acts that are better than one that contains going around torturing people.

But Howard-Snyder's view falls prey to *Ellie's Brain Scan* as well. Since Ellie knows how-one-would-go-about giving money to Oxfam, and this is the best thing she can do in the relevant sense, she subjectively ought to do so, yet she knows she cannot do so as a matter of fact. Once again we seem to be in the position that our theory is telling us that she *subjectively* ought to do something that she knows will not happen, which appears highly irrational. Perhaps there is some middle ground to find here, or maybe there is some other way of resolving this. I will not delve deeper into this here since my main objective has been to show that there are better alternatives than Parfit's claim that we should accept the hypothetical sense of 'can', and I think Prichard's view is such an alternative.

The major point to make clear is that it is only if we assume the hypothetical interpretation of OIC that global consequentialism has a problem above and beyond that of other theories, namely evaluative conflict. If we reject the hypothetical interpretation, wrongness stands in danger of becoming obsolete no matter *what* ethical theory you espouse, and the hypothetical interpretation of OIC is *more* counter-intuitive than wrongness being

obsolete. But *some* separate interpretation of OIC better work, unless we want to accept that wrongness is obsolete, and this separate interpretation may not lead to evaluative conflict.

Global consequentialism evaluates different focal points directly, which is what opens up the possibility of evaluative conflict if we take OIC to mean that one ought to do the best one could have done with *different* motives. But that is not why determinism and the categorical interpretation of OIC renders wrongness obsolete. As long as there is only *one* possible outcome, and we accept that the can in OIC simply means what is *causally possible* given the *actual world we live in*, there will only be one possible outcome no matter what theory we are talking about, and if ‘can’ is then to be interpreted *directly* in terms of causal possibility it *has* to be the case that wrongness is obsolete. Parfit introduces the hypothetical sense of ‘can’ to escape this conclusion, but we have seen that when it comes to subjective rightness, his solution cannot be correct.

This is why Prichard and Howard-Snyder’s replies, which focus on what you actually believe and actually know how to do, are important. Where Parfit juxtaposes the motives you could have had with the ones you do have, they focus on what *kind* of possibility is necessary for OIC, and what they arrive at is not really dependent on causal possibility (though that might be an underlying assumption). So, there is no *special* problem for global consequentialism according to the categorical interpretation of OIC, and if we are talking about subjective rightness, then the hypothetical interpretation *cannot* be correct, as *Ellie’s Brain Scan* shows.

In response to this, Parfit might backpedal and say that it is only when it comes to *objective* rightness that I ought to do the best I could have done if my motives were different. But as we shall see in the next section, this does not work either.

4.5 The Second Reason why Claim (1) is Not Claim (3)

Ellie’s Brain Scan shows us that Ellie does not have a subjective obligation to give money to Oxfam, and I have shown that there are some possible options available to Molly in *My Malevolence*. But what if all Parfit is claiming is that it cannot be the case that Molly *objectively* ought to do the best that is possible given her actual motives? That is, she objectively ought to refrain from torturing people after taking the motive-altering pill, but not subjectively. This would be a rather large concession to make, but not an unreasonable one I think.

Ellie's Brain Scan is not effective as a counter-example in cases of objective evaluative conflict. The fact that Ellie knows that she will not give the money to Oxfam is seemingly irrelevant to whether or not giving it would be the objectively wrong thing, since objective rightness is only sensitive to what is actually possible or what actually happens, not what an agent believes about what will happen.¹⁴¹ Of course, it may still appear strange that according to Parfit, not torturing people cannot happen after Molly takes the pill, yet not torturing people is counted among the possible *alternatives* from which the objectively right act will be chosen. Bestness is a property of some option relative to some other option, and that comparison class has to be decided somehow. If it is not what you believe is possible *or* what is causally possible, then what is it? This is what the hypothetical interpretation of OIC is an answer to; it claims that you only ought to do better if you could have done so with different motives, yet the fundamental weirdness of choosing the best causally possible option among a mix of causally impossible and possible outcomes should not be lost on us even when it comes to objective rightness.

Still, when it comes to objective rightness, Parfit's solution has not yet been shown to have an equally unattractive consequence as wrongness becoming obsolete. However, there are two very unattractive consequences that pertain even to objective rightness. His view is *arbitrary*, and it is *inconsistent* with how we think about what we physically 'can' do. Let us focus on the latter first; what happens if we alter Parfit's analysis slightly?

Parfit argues that OIC cannot be interpreted so that one ought not have acted differently if acting differently would have been impossible given one's actual motives, because if it is interpreted thus and psychological determinism is true, wrongness is rendered obsolete. There would only be one possible action for any given set of motives (and initial conditions). But, you could make a parallel argument about *physical* determinism. If OIC means that you ought not have acted differently if acting this way would have been physically impossible given the actual state and laws of the universe, then it will equally follow that wrongness is obsolete, since there is only one possible outcome. So, by analogy to Parfit's argument, we should rather interpret OIC as meaning that acting this way would have been physically impossible, even if the actual state/laws of the universe had been different. Let us

¹⁴¹ Though you could argue that if she knows she will not, it is true that she will not, in which case she cannot, and by OIC that means she ought not. But given Parfit's argument, this would be begging the question.

call these two new claims (2') and (3'). OIC claims that you ought not do something if you cannot, and 'cannot' can at be interpreted in at least these four ways.

(2) That acting in this way would have been impossible given my actual desires and dispositions.

(3) That acting in this way would have been impossible, even if my desires and dispositions had been different.

(2') That acting in this way would have been impossible given the actual state/laws of the universe.

(3') That acting in this way would have been impossible, even if the actual state/laws of the universe had been different.

For now I am assuming that the reason why psychological determinism is true is that physical determinism is true; if psychological determinism is true for some other reason the following arguments would not work. But, given this assumption, and if we do say that 'ought' implies that it is physically possible given the state and laws of the universe, then wrongness is similarly obsolete. It is also rendered obsolete for *exactly* the same reason that it would be obsolete if psychological determinism is true and OIC includes the claim that you ought not do otherwise if you cannot do so with your actual motives.¹⁴² Now, if the only way to counter this were to retreat to the claim that you ought not do so if it would have been impossible even if the state and laws of the universe had been different, then OIC will be *severely* punishing. You will not be able to appeal to OIC even if the reason you are unable to save someone from a burning building is the fact that you are completely paralyzed. After all, if the state of the universe had been different, that is, if you were not paralyzed, it would have been possible for you to rescue them, and so you ought to do so even though you cannot. As mentioned, this holds as long as psychological determinism is true *because* physical determinism is true.

Besides being severely punishing, the demand that you (3') ought to do what would be best if the state or laws of the universe were different would result in evaluative conflict for other theories than global consequentialism. Let's say that if you do an act *x* now, you will not (actually) be able to do act *y*, because you will be too far away from the place where the act needs to be done. E.g. if you would have to leave now in order to do *y*, but act *x* will take about an hour. Then according to (3') you *still* ought to both *x* and *y*, even though you cannot do both. This goes for a deontologist, a virtue ethicist, etc., as much as for a consequentialist.

¹⁴² The wrongness of acts, that is.

Of course, we normally take OIC to *include* ‘if *S* does *x* then *S* will be too far away to *y*’, but (3’) effectively cuts off that possibility. This is a form of evaluative conflict since it tells you that you ought to *x* and you ought to *y*, but you cannot do both, the only difference is that in this instance both *x* and *y* are acts. It is parallel to what happens if we accept Parfit’s interpretation (3) of OIC and psychological determinism, since it would mean that you ought to have a set of motives, and with that set of motives there is something you cannot do, but which you still ought to do. This follows because neither (3) nor (3’) base rightness on what is actually possible, but rather on what could have been possible.

The main point is that we do *not* usually accept that if physical determinism is true, then OIC means that wrongness is obsolete. In some cases we take the fact that an outcome is causally determined to mean that you *can* appeal to OIC, other times we reject this and take it to mean that you *cannot*. We accept being paralyzed as a legitimate reason for not rescuing someone from a burning building, but we don’t accept the mere fact that one is physically determined to do something as sufficient for one to appeal to OIC, presumably because it would lead to wrongness becoming obsolete. But if we refuse to choose in one area of discourse, we need a reason why we have to choose in the other.

Parfit assumes that when it comes to psychological determinism we are immediately forced to choose between (2) and (3), but I don’t think he would say that we are forced to choose between (2’) and (3’). This is what I take to be *inconsistent* about his view, at least without giving some sort of reason why we treat them differently. Of course, it might be that we should simply accept that we have been wrong to delude ourselves into thinking that there is a difference between being determined and being paralyzed.

What all of this means is that wrongness is obsolete for *all* ethical theories if we take OIC to mean that you only ought to do what you ‘can’ do with your actual motives. As long as motives have causal force, which they do, it will be true of all ethical theories that there is only one possible outcome for any set of motives, and so the same abjection holds for all theories, and all theories will be treating psychological and physical determinism unequally without justification if they opt for (3). It think this is a good indicator that—as long as psychological determinism is true because physical determinism is true—there is some option beyond (3), and that option may or may not result in evaluative conflict. Perhaps there is a distinction between mere determinism and other kinds of inabilities within the mental realm

as well, or perhaps there is some further adjustment that needs to be made to the ‘ought’ part of OIC. Perhaps OIC should be rejected, or perhaps we need to accept evaluative conflict. These options will be explored below.

Now, it is important to note that the apple of discord is *not* the traditional debate on *free will* and determinism, the problem is how to interpret ‘ought’ implies ‘can’ as long as determinism is true. (What I *am* asking is whether OIC may demand that one ought to do something that is causally impossible.)¹⁴³ It is entirely possible to be a compatibilist and hold that free will exists even in an entirely deterministic world, yet still hold that OIC breaks down as a principle given determinism.¹⁴⁴ One way to reach this conclusion would be to accept the conclusion that if there is only one possible outcome, and OIC is true, it has to be the case that wrongness is obsolete because there are no reasonable ways out of this predicament. Or, you could deny that free will exists, but still accept that OIC holds. One way of reaching this conclusion would be if you believe that quantum indeterminacy could affect our actions. I.e., that certain processes on the quantum scale are irreducibly non-deterministic, and that this indeterminism can affect our actions, as some philosophers hold.¹⁴⁵ The possibility that some of our actions are affected by randomness, as opposed to completely determined, is arguably irrelevant to whether we have *free will*. As J.J.C. Smart famously argues, “Indeterminism does not confer freedom on us: I would feel that my freedom was impaired if I thought that a quantum mechanical trigger in my brain might cause me to leap into the garden and eat a slug.”¹⁴⁶ But, this randomness could make a large difference when it comes to the question of whether OIC holds. If *more* than one action is possible given a single motive set and initial conditions, it does *not follow* that wrongness is obsolete just because the categorical interpretation of OIC is correct, as more than one action would be possible given a single motive set.

Michael Stocker does not consider this (as he is not considering determinism at all), so he excludes *all* psychological disabilities from OIC without argument.

¹⁴³ In this section I focus mostly on objective rightness, as there is an independent reason not to accept Parfit’s solution on subjective rightness. But it should be noted that these arguments also work for subjective rightness.

¹⁴⁴ Compatibilists believe that determinism and free will are compatible; both can obtain. Michael McKenna and Justin D. Coates, “Compatibilism”, *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2015 Edition), § 1, <http://plato.stanford.edu/archives/sum2015/entries/compatibilism/>.

¹⁴⁵ See for instance, Robert C. Bishop, “Chaos, Indeterminism, and Free Will,” in *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford University Press, 2002), 119–120.

¹⁴⁶ J.J.C. Smart and J.J. Haldane, *Atheism and Theism*, 2nd ed. (Malden: Blackwell Publishing Ltd., 2003), 63.

It [OIC] would almost certainly be uninterestingly false if considered in light of *psychological* inability, the one present where a person simply "cannot get" himself to do the act in question—e.g., the inability of a person to stick his arm into a cesspool to retrieve his wallet. In these last two sorts of cases, the agent "really can" do the act; and for this reason it is difficult to see how the Doctrine can plausibly be said to involve such 'cans'.¹⁴⁷

Why does he put scare quotes around 'really can'? I agree, of course, that if an agent really can do an act, then she cannot appeal to OIC. The question becomes in what sense you 'really can' do something which, given the laws and the state of the universe is causally impossible? It would seem like you *really cannot* do that which is impossible in this sense. Or at least, if we do assume determinism, Stocker is treating the determination of motives and the determination of acts dissimilarly without justification. Stocker is *not* considering determinism in the text where this quote appears, so he is excused, but we are addressing Parfit's argument which *is* based on a presumption of determinism, and since determinism might be true, we need to consider the implications it has for OIC.

R.M. Adams makes some rather similar statements (though he also does so without considering determinism). He starts out by saying:

We do not say that a person rightly jumped from a burning building, saving his life, but wrongly struck the ground, breaking his leg. It is unreasonable, on this view, to separate, for moral evaluation, actions that are causally inseparable.¹⁴⁸

I agree. But he later adds another point where he is considering what Jack (the same as from the *Chartres* case) ought to do. (Adams is assuming that being motivated by something *other* than maximizing utility will maximize utility here, but that is not what is important for our purposes.)

Jack was once, let us suppose, an excessively conscientious act utilitarian. Recognizing the duty of cultivating more useful motives in himself, he took a course of capriciousness training, with the result that he now stands, careless of utility, before the choir screen. . . . It seems clear enough . . . that Jack's staying at the choir screen is separable from his earlier efforts at character reform in a way that striking the ground is not separable from jumping out of a building. Once you have jumped, it is no longer in your power to refrain from striking the ground. . . . Therefore it is inappropriate to separate the impact from the leap, as an object of moral evaluation. But even after Jack has taken capriciousness training, it is still in his power to leave the choir screen.¹⁴⁹

So he thinks that character does not follow character reform in the same way that hitting the ground follows jumping. But, as Parfit shows, once we add determinism to the mix, this

¹⁴⁷ Michael Stocker, "'Ought' and 'can'," *Australasian Journal of Philosophy* 49 (1971): 311.

¹⁴⁸ R.M. Adams, "Motive Utilitarianism," 473.

¹⁴⁹ *Ibid.*

quickly gets us into trouble, at least as far as OIC is concerned. If determinism is true it may in this particular case be that it is *not* within his power to leave the choir screen after having undergone capriciousness training. It is true that *if* he wanted to it might be within his power, but that is something else entirely than saying it is actually within his power to want to, and consequently do so. If we assume determinism (which Adams does not), Adams is simply begging the question, since he tells us that the character and character reform are separable in a way that jumping and hitting the ground is not, but on determinism this is the question that needs answering. The question is whether and in what way character and character reform could be separable, as both are determined by physical laws. One is governed mainly by chemical laws and the way his brain is built, and the other by gravity, but this does not alter my point. This is, of course, parallel to Molly and her motive-altering pill in *My Malevolence*.

We do treat being paralyzed differently than mere physical determinism, because being paralyzed somehow seems more like a hindrance than being determined to do something. We might be wrong to do so, but this shows that it is certainly not *obvious* that no psychological inabilities can count.

Still, the possibility certainly remains that they never count, though that means we have an inexplicable difference between determinism in the brain and outside it. Perhaps it's pure chance, so it *could* have gone otherwise, but in the lottery of how the world turned out it just randomly happened that all the cases OIC admit ended up outside the brain. So, Molly ought to take the motive-altering pill, acquire the torture-inducing malignant motives, and then not torture anyone, even though this is impossible. But notice that if this is the case, then it is not only according to global consequentialism that one will sometimes experience cases where one is determined to do the wrong thing; it rather goes for all ethical theories. For the sake of simplicity, imagine that I am a deontologist who believes that absolutely any lie is wrong. It will still sometimes be the case that I am going to do the wrong thing (lie) *because* I did the right thing earlier. One example:

My Egoism: There is a known killer at my door asking where my friend is, and I know he intends to kill him. Believing lying to be *absolutely* wrong, I refuse to answer, since I know he is in my kitchen. The killer becomes suspicious, enters the apartment, and kills my friend. Later, I *do* lie about my response to the killer, because I fear that others will not understand that what I did was right, and I do not have a sufficiently strong will to face their judgment.

In this case I will have done the wrong thing *because* I did the right thing. I probably would not have lied later if I did not first refuse to lie, and since I did tell the truth it was determined

that I would later lie if determinism is true. The main problem with evaluative conflict is that there are causal connections between motives and acts, so sometimes you will necessarily do the wrong thing because you have the right motives. But *all* theories have this consequence if psychological determinism is true *because of* physical determinism. Granted, I have only shown that this is true for a specific and caricaturish version of Kantianism, but I think it is clear from this case that other theories will have the same problem. It can happen simply because a right act may sometimes have the causal outcome that you do a wrong act. Global consequentialism simply has the same problem between a larger set of evaluative focal points.

Theoretically, it is possible that this never occurs for some agent. That is, it could theoretically happen that doing the right thing never has the outcome that the agent does something wrong that she would not otherwise have done. This would be the case for someone who never does anything wrong, but this would be exceedingly rare. Furthermore, the same goes for a global consequentialist and their version of evaluative conflict. It is not *impossible* to live a life where the right motive always leads to the right act, the chances of it happening are just microscopic.

But this is not only the case when one act causes another act, motives cause acts according to all ethical theories, and as long as they do, and physical determinism is the reason why psychological determinism is true, all theories have to contend with the problem Parfit describes, that they ought to do only what they can do, and they can only do one thing with the motive set that they have.

I have argued that Parfit's solution, which demands that we do the best we could have done with better motives, is inconsistent with how we think about OIC when it comes to physical impossibility, but earlier in this section I also claimed that the categorical interpretation of OIC is *arbitrary*. The reason it is arbitrary is related to the reasons why it is inconsistent. It is arbitrary because the moment we allow impossible outcomes to count as something we *can* do according to (objective) OIC, a question that immediately presents itself is *which* impossible acts ought to be included. If we allow acts that cannot occur given our motives, why not also allow acts that cannot occur because you are paralyzed? Why not allow acts like magically making everyone happy or growing wings and saving someone whose parachute has not opened? Some reason has to be given why determined motives are singled out as insufficient for OIC.

Determinism seemingly leads to some problems according to both (2) and (3), i.e. whether we think of what we ought to do as the best we can do with our actual motives or the best we could have done with better motives. Should we accept one of them anyway, or are there some ways to avoid this dilemma? In fact, there are several ways to avoid it. We could reject determinism; I consider versions of this in sections 4.6 and 4.7. We could accept determinism but make some other distinction besides the one between the categorical and the hypothetical interpretation of OIC; I consider this option in section 4.8. We could reject OIC entirely; I consider this in section 4.10. We could reject the notion that OIC applies to objective rightness, which I consider in section 4.8.2. Or, we could do what I take to be the best option, which is to reject the notion of objective rightness entirely; I consider this option in section 4.8.1.

I hope to show that there are at least *some* options open to the global consequentialist, and that there are problems related to OIC and determinism that pertain to all ethical theories alike. Therefore, even *if* global consequentialism leads to evaluative conflict, this will not show that there is anything *especially* wrong with the theory. It just happens to be the particular problem that is associated with global consequentialism specifically, whilst other theories will have other problems associated with them. And as I will show, there are ways of avoiding the problem, though they are not as satisfying as one would hope.

4.6 Indeterminism I: General Psychological Indeterminism

This section and the next focuses on what a global consequentialist should say if she rejects determinism. This section specifically focuses on what she should say if she rejects determinism and accepts that human actions are fundamentally free. That is, if they are more or less as we think of them in our everyday lives, as motives and actions never being determined. Most of what I will say in this section also applies to subjective rightness; I place it here simply because I don't want to make the argument twice, that is, once for each sense of rightness.

Now, if determinism is simply false then Mason's reply to evaluative conflict will still be possible. It will be reasonable to require that Molly takes the motive-altering pill, gets malignant motives, and then chooses *not* act on those motives. If she does act on them, that is her responsibility, and not a problem for global consequentialism, no matter how difficult it might be. If so, wrongness is not obsolete.

This is obviously very demanding of Molly, and some philosophers have seen excessive demandingness as a problem.¹⁵⁰ It demands that she takes the pill and then does the best act she is physically able to do. Similarly, it means that even if you are tortured, it is simply your duty not to give up information as long as this is the best thing you can do. Absolutely any deviation from the very best act at any given time is straightforwardly wrong, and the same would go for motives and acts; you should have the best motives and do the best acts no matter what. As I mentioned earlier, there are good replies to the objection that consequentialism is too demanding, and it is open to the global consequentialist to accept one of these.

Parfit also considers what global consequentialists should think if psychological determinism is not true, and he disagrees with the answer I just sketched. He thinks that there will *still* be cases where you ought to do something you cannot do. If he is right, we might still end up with evaluative conflict even if psychological determinism is false, but as we shall soon see he is probably wrong about this.

To make his point, Parfit utilizes a slightly different version of the *Clare* case which I will address as *Clare II* (he calls it *case two*).

Clare II: “Clare could either save her child’s life, or save the lives of several strangers. Because she loves her child, she saves him, and the strangers all die.”¹⁵¹

This sets Clare up for a potential problem case even if psychological determinism is not true.

Pure do-gooders: “[global consequentialism] implies that we should always try [to] do whatever would make the outcome as good as possible. If we are disposed to act in this way, we are *pure do-gooders*.”¹⁵²

Being a pure do-gooder, according to this definition, means that you are always *disposed* to try to do the value maximizing act, and this carries with it certain problems.

“If I [Clare II] was a pure do-gooder, my ordinary acts would never be wrong. But I would be acting wrongly in allowing myself to remain a pure do-gooder. If instead I caused myself to have one of the best possible sets of motives, as I ought to do, I would then sometimes do what I believe to be wrong. If I do not have the disposition of a pure do-gooder, it is not causally possible that I *always* act like a pure do-gooder, never doing what I believe to be wrong. Since

¹⁵⁰ Brad Hooker, “The Demandingness Objection,” in *The Problem of Moral Demandingness: New Philosophical Essays*, ed. Tim Chappell (London: Palgrave Publishers, 2009), 148–162.

¹⁵¹ Derek Parfit, *Reasons and Persons*, 33.

¹⁵² *Ibid.*

this is not causally possible, and it would be wrong for me to cause myself to be a pure do-gooder, I cannot be morally criticized for sometimes failing to act like a pure do-gooder.”¹⁵³

There are some things to nitpick at here, like the fact that the definition of a pure do-gooder states that you are disposed to try to do the acts that will maximize value, which does not imply that you never actually do anything wrong. He also assumes that the best motives are *not* always to want to do the best thing, since being motivated to do the best thing might sometimes lead to less good than some other motive set, so in this case being a pure do-gooder means *not* having the best motives.¹⁵⁴ He thinks this means that her “ordinary acts” will not be wrong, but she will be acting wrongly in allowing herself to remain a pure do-gooder. This, of course, carries with it certain conundrums, for if she is always disposed to the best act, it seems to mean that she would *not* choose to remain a pure do-gooder, for example. But I propose to leave these little conundrums to the side. The main point is that Parfit rejects the possibility of *always* acting like a pure do-gooder without the disposition of one, because he thinks that “even those that deny Determinism cannot completely break the link between our acts and our dispositions.”¹⁵⁵

What he is arguing is that Clare II could—by an effort of will—act against her strongest desire and save several strangers rather than her own child, much like Ishmael used strong moral principle to prevent himself from stepping into the street and methodically knocking people’s hats off. And I agree that if determinism is *false* then this scenario at least seems plausible; it is how we typically think of choices. But, he continues, it is not possible that Clare II “could *always* act like a pure do-gooder *without* having a pure do-gooder’s disposition”.¹⁵⁶ That is, she could not always act against her motives.

He gives no argument for this, although it seems to require one. He says that we cannot completely break the link between our acts and our motives (dispositions), but what link is he talking about? He has just admitted that Clare II *can* act against her strongest desire *once* by an effort of will. Since he says that we cannot *always* act contrary to our motives, he has to mean something like “we cannot *always* break this link”. Indeed, this seems much harder, but is it really causally impossible? We are left with a conundrum, what is it that

¹⁵³ Derek Parfit, *Reasons and Persons*, 35.

¹⁵⁴ As we saw in section 1.3.2, simply attempting to do the value maximizing *act* will not always maximize happiness overall. Like when Jack is in Chartres looking at the choir screen, if he is as motivated to see everything as he is the overall value is greater than if he is simply trying to do the best acts.

¹⁵⁵ *Ibid.*, 36.

¹⁵⁶ *Ibid.*

makes it *possible* to break it once, but *impossible* to break it every time? This is not a mere matter of degree, breaking the link *n* times is possible, the successive number is *impossible*. Exactly how many times can you break it? Two? Three? A million?

Consider this from a different perspective, breaking the link once is hard, but is breaking it twice roughly twice as hard, and breaking it thrice roughly three times as hard? I cannot see why this would be the case. Presumably it is (roughly) equally hard to break it the second time as the first if we assume that the motivation is equally strong. And if two breakings of the link between motive and action are each very hard individually, why should breaking both be twice as hard? And furthermore, isn't there an argument to the effect that the opposite happens; that we would get better and better at acting against our motives the more we did it? Of course, I am not saying it could not be amazingly hard or extraordinarily hard to do so, it might even be the hardest thing anyone could ever do. What I am asking is whether it is *impossible*. If it is merely the hardest thing ever done, then it will simply be obligatory because one ought to have the best motives and do the best act, but if it is impossible then Parfit thinks we have a problem on our hands. These considerations should show that it is at least not *obvious* that we cannot always act against our motives given non-determinism.

The point is that if you can break it all the time then there can be no evaluative conflict. If it is possible, given non-determinism, to always act against one's motives, then there is no problem for global consequentialism and the reply will be along the lines sketched earlier in this section.

This is an empirical matter, and I am not rejecting the possibility that there is some actual number of times after which it becomes impossible for Clare II to act against her motives.¹⁵⁷ But the point is that if that is the case, then *she should act against her motives as much as she can*, and when she cannot she *will necessarily* be acting in accordance with them, and so she can appeal to OIC. She either can always break the link, in which case there is no evaluative conflict, or she cannot, in which case OIC kicks in and there is still no evaluative conflict. Any time she can and ought to break the link, but does not, what she does will be wrong, but that is *her* fault. As long as it is not necessarily the case that she has to do the

¹⁵⁷ I say 'after which' it becomes impossible. I am not suggesting that there is a successive string of occasions where she can, and then another successive string of occasions where she cannot. The occasions where she is able and the occasions where she is not able need not follow each other in any particular order.

wrong thing we should blame her for not living up to the demands of morality, not blame morality for making demands she cannot live up to.

This reply relies on OIC demanding (2), that you do the best you can with your actual motives, and not (3), that you ought to do the best you could with better motives. Parfit appears to reject my reply because he is still assuming the latter interpretation of OIC.¹⁵⁸ This is not all that strange; this is how he has argued we should interpret OIC, after all. But *why on earth* should we interpret it in such a way in this case? If determinism is *not* true then there is no rationale for basing OIC on what motives we could have had, since basing it on the motives we do have would not imply that wrongness is obsolete. Remember, we are assuming that the world is not determined, and it was the *conjunction* of determinism and OIC resting on our actual motives that would render wrongness obsolete. So, it may be true that Clare II is unable to have one of the best possible sets of motives *and* always act like a pure do-gooder, but if this *conjunction* is impossible we can say that she ought not to do both and there will still be plenty of wrong actions to go around. This impossibility in no way whatsoever makes it impossible to *ever* do anything wrong.

For the reasons I have sketched, Parfit thinks that global consequentialism tells us to always have the best motives and always do the best we can, he also thinks that this is sometimes impossible even if the world is not determined. This amounts to a form of evaluative conflict. Because of this he insists on either calling Clare II's actions or her motives wrong, even if she could not fulfill both obligations. But he says that she is not *blameworthy* for this. According to Parfit this is enough to satisfy OIC, because what he thinks is most difficult to accept about cases like these would be if you could be inescapably blameworthy.¹⁵⁹ I don't see how this satisfies OIC. I say we should rather take my approach: use Mason's reply and assume that we ought to do the best we can with our actual motives. There is little reason not to, as long as determinism is false. This view does not result in anything like *Ellies Brain Scan*, it is not arbitrary, and it does not render wrongness obsolete. In addition it works for both subjective and objective rightness and we avoid having to base

¹⁵⁸ Derek Parfit, *Reasons and Persons*, 15, 36.

¹⁵⁹ *Ibid.*

OIC on nonexistent motives. It does, however, also rely on the plausibility of the claim that our acts are not determined, which has been rejected by many philosophers.¹⁶⁰

So, what Parfit does is he implicitly assumes a form of *local* psychological determinism, because if you cannot *always* act contrary to your strongest motives then some acts must be determined by motives.¹⁶¹ This may be true, but as I have argued this does not mean we end up with evaluative conflict. It does mean, however, that some of our choices are determined. In this section I have argued that if *none* of our acts are determined by our motives, then we can never appeal to OIC because of them. I will give a somewhat fuller explanation of what I think if some acts are determined, and some are not in the next section, though the gist of the answer has been given here, since Parfit's non-determinism view *assumes* local determinism.

4.7 Indeterminism II: Local Determinism

The unsatisfactory solution Parfit gives comes from an assumption he makes when he first presents his argument the hypothetical interpretation of OIC. He assumes that for Molly to be able to defend her actions after taking the motive-altering pill by appealing to OIC, we have to assume *global* psychological determinism.¹⁶² That is, we have to assume that for *any* set of motives and initial conditions there is only one possible act that can be the upshot.¹⁶³ This is simply not true, and as we have just seen, he countermands this later when he considers non-determinism. An alternative possibility would be that *some* acts are psychologically determined by motives, while others are not. This does not have to mean that it is the *number* of times one breaks the link between motives and action that eventually makes it impossible to do so, it could rather be the *type* of motive or action that does it.

Building on the case I built in the previous section, this is clearly a plausible option if you do not think that the world is globally determined. On such a view, most cases where you are considering your options you can do any physically possible act, but other times there are

¹⁶⁰ David Hodgson, "Quantum Physics, Consciousness, and Free Will," In *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford University Press, 2002), 99–102. As I mentioned in section 4.5, quantum mechanics may be irreducibly non-deterministic, and this non-determinism may affect our decisions or it may not. There is no consensus.

¹⁶¹ Derek Parfit, *Reasons and Persons*, 36.

¹⁶² *Ibid.*, 14. By which I simply mean that *all* acts are psychologically determined, as opposed to *some*.

¹⁶³ *Ibid.*, 14–15, 32–33. (The case Parfit makes for his view is presented in the section on rationality (pp. 14–15), not ethics (pp. 32–33), but he specifically states that it is the same answer in both circumstances.)

some physically possible acts that are *psychologically* impossible. If Molly can appeal to OIC after taking the motive-altering pill, and she can do so *without* assuming global determinism, it means that wrongness is not necessarily obsolete just because it is right for her to torture people after taking the motive-altering pill. It might be that after taking the pill her brain is simply so constituted that the particular act of not torturing is impossible.

Imagine someone who gives in to torture because she really “could not take it anymore”. Let’s say she gives up the location of a group of righteous resistance members, who will be hunted down and killed by the evil torturers (of the variety that has not popped any motive-altering pills). You might want to say that she has recourse to defend herself by appealing to OIC, while still maintaining that she cannot do the same if she is caught stealing, and her only justification is that she wanted the loot. The rationale for such a view could be that it is psychologically determined that she is going to give in to torture, but it is *not* psychologically determined that she will steal, that is a free choice.

This could be true just like some physical phenomena are determined while some are not, which would be the case if global determinism is false. Even if the universe is not globally determined, there are certainly cases where it *is* determined that some particular event will happen, e.g. a certain comet will crash into a certain planet because no-one is around to stop it, and the laws of physics say they will crash.

Now, what category Molly and her torturing people falls into is obviously an open question, and depending on what category Molly is in, she either can or cannot appeal to OIC. But on such a view, Mason’s reply is still an option *without* making wrongness obsolete. If Molly really cannot refrain from torture then she can appeal to OIC, and if she can refrain from doing so, then she should refrain from it, no matter how difficult that might be, as we saw in the previous section. This, then, is the second way that a global consequentialist might reply to Parfit’s objection: reject *global* psychological determinism but embrace *local* psychological determinism. The first I presented in the previous section; to reject the claim that *any* act can be psychologically determined, so you can never appeal to OIC because of your motives.

Beyond the mere possibility of quantum indeterminacy affecting our actions, there are some *prima facie* reasons to accept that some acts are too difficult for someone to perform. One I have mentioned already is torture. Even trained professionals are probably not able to

withstand torture for as long as they want. As Fritz Allhoff puts it, “there is no evidence that anyone can resist torture-laden interrogations indefinitely; the psychological trauma and the degree of confusion are simply too severe.”¹⁶⁴ If they *cannot* withstand it indefinitely, then why should they not be able to appeal to OIC? Another example might be psychopaths. Jonny Anomaly argues that psychopaths are incapable being motivated by moral considerations, and that they should therefore be able to appeal to OIC.¹⁶⁵ Another example might actually be *My Malevolence*, in which Molly takes a pill that alters her motives in such a way that she *will* go around torturing people. I have no idea what the machinations of such a pill would be, but it does not seem impossible to imagine a pill like that, which somehow *rewires* the brain in such a way that not torturing is impossible. Or, in a similar vein you might say that a mad scientist rewires her brain instead of a pill, and that the very structure of her subsequent brain makes it impossible for her to refrain from torture. Someone who fails to do the otherwise right thing because she is “paralyzed by fear” might be another example, or someone who fails under similar circumstances because of a strong case of obsessive-compulsive disorder.

If some acts are determined and some are not, then the same defense as the one I provided in the previous section serves to save global consequentialism from evaluative conflict. We can still base OIC on what is possible with our actual motives, since this will not render wrongness obsolete. If you are being tortured, on this view, you may be able to appeal to OIC, but if you steal something just because you want the loot you probably cannot. If you are being tortured and you can choose to not reveal information then you ought to do so; if there is no other option than revealing the information then you may appeal to OIC.

In this section and the last one I have argued that there is no problem with evaluative conflict for global consequentialists if we assume that either no acts are psychologically determined by motives, or if only some acts are psychologically determined by motives. This is important because it shows that we are not forced into wrongness being rendered obsolete just because Molly can defend herself by appealing to OIC after taking the motive-altering pill and becoming a torturer.

Parfit started out arguing that if Molly can defend herself by appealing to OIC then wrongness is obsolete, because the defense presupposes global psychological determinism.

¹⁶⁴ Fritz Allhoff, “An Ethical Defense of Torture in Interrogation” In *Ethics of Spying: A Reader for the Intelligence Professional*, ed. Jan Goldman (Oxford: Scarecrow Press, inc., 2006), 136.

¹⁶⁵ Jonny Anomaly, “An Argument against External Reasons,” 56–59.

This is not true, since it could simply be that *in this particular case* she would be unable to refrain from torturing people, which is a possibility if determinism is not true. Then, when he considers the possibility that determinism is not true, he does argue based on the assumption that a form of merely *local* determinism is true. He then implausibly concludes that an inability to break the link between motives and actions *all the time* is insufficient for one to appeal to OIC. There is absolutely no rationale for this conclusion, because it would not render wrongness obsolete to base OIC on the categorical sense of ‘can’ if determinism is not true. This seems like a good turn for the global consequentialist, but of course, these arguments rely on the plausibility of the claim that some sort of non-determinism is true.

4.8 Global Determinism: the Hard Case

We have looked at how to interpret OIC given two types of non-determinism, local and global, but how should we interpret OIC if global psychological determinism is true? There are two possibilities here. Acts may be globally determined by motives, that is, all acts are entirely determined by motives and initial conditions, but motives, dispositions, etc. are not determined. Or, acts may be determined because *everything* is determined physically, that is, determined by the laws of nature and initial conditions. In this section I focus on the latter, and in section 4.9 I focus on the former. As I have mentioned, quantum mechanics may be an exception to global determinism, but it is unclear whether this can affect our actions. When I mention global determinism in *this* section I will be assuming that quantum mechanics may or may not involve genuine indeterminism, but that this potential indeterminism does not in any way affect our actions. So, the view we will examine in the present section is one where all acts are psychologically determined, and psychological determinism is true as a result of physical determinism from natural laws and initial conditions. Acts are determined by motives, which are in turn determined by the laws of nature.

Determinism is what made wrongness obsolete on Parfit’s view, but as we have seen, his way of escaping this conclusion has serious problems. We seem to have a bit of a problem on our hands.¹⁶⁶ Is there some way to avoid the problems associated with interpreting OIC in terms of either what is possible with your actual motives, or what would have been possible with other motives than those you actually have?

¹⁶⁶ We are focusing on physical determinism in this section, but physical determinism entails psychological determinism, unless we count the possibility of non-physical minds. I will assume that physical determinism entail psychological determinism.

There are two things that should be noted before we tackle this issue. The first thing to note is that if the ‘ought’ in OIC is derived from objective rightness, then OIC has some rather strange consequences even if we reject determinism. That is, even if determinism is *not* true, the notion of objective rightness seems to imply that you ought to do the best thing you are physically able to, for example you ought to *immediately* write down the cure for cancer on a napkin and hand it to the nearest drug company.¹⁶⁷ This is because you are able to do so in the regular sense of the word, you can write words and as long as cancer is actually curable, you are able to write down the words that describe how to do so. It’s just that you don’t know how to string together the words. If we assume that this is the one best thing you are able to do, it is also obligatory.¹⁶⁸ This is rather counter-intuitive. Now, if determinism is true, there is only *one* thing you are physically able to do, so this problem disappears. You are *not* causally able to write down the cure for cancer, on account of not knowing how to do so, so it is not obligatory. Unfortunately this result is reached at the price of rendering wrongness obsolete. So it seems like even if we avoid the problems associated with determinism, couching OIC in terms of objective rightness is still counter-intuitive.

The second thing I should make clear is that if we treat physical determinism and psychological determinism consistently, as we should as long as psychological determinism is a result of physical determinism, this problem is a problem for *all* ethical theories. If psychological determinism is true because of physical determinism, then there will be only one physically possible outcome in any situation no matter what theory is true, so all theories have the consequence that wrongness becomes obsolete if the ‘can’ in OIC is based on what is causally possible given your actual motives, or extremely punishing if the ‘can’ in OIC is based on what is causally possible whatever the state and laws of the universe might have been. These six options seem are the best options I can think of for how to understand OIC given determinism and assuming we are after *objective* rightness; it is not meant to be exhaustive.

(A): Consistently accept (3) and (3’) when it comes to objective rightness. That is, both that you ought to do the best you could have done with different motives, *and* what you could have done if the state and the laws of the universe were different. Wrongness is not obsolete, but we get the result that you ought to rescue someone from a burning building even though you are paralyzed.

(B): Inconsistently accept (3) and reject (3’) when it comes to objective rightness. I call this the ‘hope for a miracle’ solution. (The same goes for inconsistently accepting, (2) and not (2’), and for accepting

¹⁶⁷ Elinor Mason, “Objectivism, Subjectivism, and Prospectivism,” 194.

¹⁶⁸ *Ibid.*, 189.

(2') and not (2), and for accepting (3') and not (3), but these do not seem intuitively acceptable from any perspective I think.)

(C): Consistently accept (2) and (2') when it comes to objective rightness. That is, both that you ought to do the best you can given your actual motives *and* that you ought to do the best you can given the actual state and laws of the universe. Wrongness is obsolete.

(D): Reject objective rightness, this is what I take to be the best solution (see section 4.8.1).

(E): Reject that OIC applies to objective rightness (see section 4.8.2).

(F): Accept psychological and reject physical determinism (see section 4.9).¹⁶⁹

(G): Reject OIC outright (see section 4.10).

I take option (A) to be a non-starter. OIC appears to lose its entire *raison d'être* if we accept that you ought to rescue someone from a burning building even though you are paralyzed. The whole point of OIC is that it tells us that in exactly these kinds of cases, rescuing the stranger is not obligatory. The principle would serve absolutely no function, except perhaps pointing out that what is logically impossible is not obligatory.

Let us consider option (B), then. This may be what Parfit believes. I have already argued extensively for why I find this option unsatisfying, and I will not argue it again here. Suffice to say that this seems inconsistent as long as psychological determinism obtains as a result of physical determinism, as I am assuming for the time being. This option simply seems to say that we ought to somehow hope that there is a difference between physical and psychological determinism, in spite of the one stemming from the other, which is why I call it the 'hope for a miracle' solution.

What should we say about (C)? Most people would not want the outcome that wrongness is obsolete. Is there some way to avoid this result according to (C) and as long as we are talking about objective rightness? I don't think so. You could try and introduce the same distinction that we intuitively do for physical determinism and physical disabilities like being paralyzed. That is, accept that we base OIC on what is actually possible, but say that just like being paralyzed is mysteriously accepted as sufficient for appealing to OIC, and being determined not accepted as sufficient, there is a similar distinction to be made between psychological determinism and psychological disabilities. And I do think that if there *is* such a difference between physical determinism and physical disability, there could very well be one between psychological determinism and psychological disability. But this solution simply

¹⁶⁹ This is a form of act-specific determinism, so it is not truly a form of determinism at all.

seems like a shot in the dark, furthermore, if something like this is true, then OIC would work similarly for psychological disabilities as it would for physical ones. If so it would be wrong to kill someone if you are able to refrain from it, as a mentally able person would be, and so wrongness is not obsolete. Furthermore, there is no special problem for global consequentialism, because this would be the case for all theories. It will be wrong of you, sometimes, to do something that you could not have avoided, but this would be a general feature of OIC, not something specific to global consequentialism.

I do not think this solution works, though. It relies on the idea that even though you are supposed to base OIC on what is *actually* possible, both physically and motivationally (as one follows from the other), you somehow still have an array of different options that you can assess in some principled way. What this could be eludes me, and I cannot see any good options in the literature. In other words, this would be arbitrary.

It appears that we should rather accept the outcome that as long as we base OIC on what is actually possible, objective wrongness is obsolete, but only objective wrongness, subjective rightness does not get affected by this if we accept Prichard's version of OIC. I shall consider this option (D) in the next section, and this is what I take to be the best solution.

4.8.1 The Rejection of Objective Rightness

As I mentioned, even if determinism is false, objective rightness seems to have the intuitively strange implication that I ought to write down how to cure cancer immediately, even though I don't know how to do so. If so, I seem doomed to do the wrong thing more or less constantly. Writing down the cure for cancer is probably a better action than any act I will ever do, so it will always be the case that what I do is objectively wrong as it is worse than curing cancer. For this reason, Howard-Snyder writes that objective consequentialism might conflict with how we intuitively understand OIC, which is that it does not require me to produce the cure for cancer if I don't know how.¹⁷⁰ The objectivist seems to be caught in a dilemma where she either has to designate actions that seem to be perfectly legitimate as wrong, like not writing down the cure for cancer, which seems to be the case if determinism is false, or, if determinism is true, accept that there are no wrong actions.

¹⁷⁰ Frances Howard-Snyder, "The Rejection of Objective Consequentialism," 242.

I should provide a more detailed presentation of Howard-Snyder's argument against objective rightness according to consequentialism:

Premise one: "objective consequentialism tells us we ought to produce the best consequences."¹⁷¹

Premise two: "we cannot [always] produce the best consequences."¹⁷²

Premise three: "'ought' implies 'can'. If we cannot do something, then it is not true that we ought to do it."¹⁷³

Assume that beating Magnus Carlsen at chess would have the best consequences. She defends premise two by arguing that it is true in one sense that I can beat him, but not true in another. I can beat him in the sense that I am perfectly able to move the chess pieces across the board in the pattern necessary in order to beat him, but in another sense I cannot beat him because even if I attempt to play as well as I can, I am going to lose, and I cannot change this fact.

Her argument is an analogy, so she goes on to justify that the chess playing case is sufficiently similar to ordinary choice-situations. The idea is that any choice you make will have consequences that reach until the end of time, and some of those consequences will be morally relevant from a consequentialist point of view. This means that it will often be harder to do the objectively right thing than it is to beat Magnus Carlsen at chess, which in turn means that if I cannot beat Magnus Carlsen, I cannot do the objectively right thing.

She goes on to argue that the second sense of 'can' is how we intuitively think about the 'can' in OIC.

This accords with our practice. We say things like, 'I cannot open either of the safes: because I do not know the combination for the one and I do not have the key for the other.' This does not have the bizarre sound of, 'She came home in tears and a sedan chair.' More importantly, if someone asserted that you ought to have opened both safes, he would retract that claim on discovering that you were unable to do so, whether the inability amounted to physical inability or the inability which involves ignorance, unless your inability (of either sort) was culpable.¹⁷⁴

¹⁷¹ Frances Howard-Snyder, "Response to Carlson and Qizilbash," *Utilitas* 11, no 1 (1999): 106.

¹⁷² *Ibid.*

¹⁷³ *Ibid.*

¹⁷⁴ Frances Howard-Snyder, "The Rejection of Objective Consequentialism," 245.

It makes sense to say that inability due to non-culpable ignorance entails that you cannot do what you don't know how to do, and we generally accept the explanation that you don't know how to do something as sufficient for not having an obligation to do so *now*.¹⁷⁵ Of course, if you ought to have learned how, that is, if the inability is culpable, you have done something wrong in the past, but it is still not true that you ought to beat Magnus Carlsen now.

She assumes a genuinely *non*-deterministic universe, and I think she shows that in such a universe, objective consequentialism is unattractive. What I add is that in a determined universe we get the additional problem that I cannot beat Magnus Carlsen, even in the other sense mentioned. That is, I cannot move the pieces around the board in a pattern that would beat him. I am causally incapable of doing so as long as I am not going to do so, and since I don't know how, I am not going to do so. It says that any action you do is right, since it is the best one you can do *in either sense*.

If objective consequentialism cannot conform even vaguely to our intuitions, no matter what kind of universe we live in, it might be better to reject it outright. If the notion of objective rightness does not even hold up in the kind of universe we usually take ourselves to live in, why should we accept it? And, since even if I *were* physically able to move the pieces in the required pattern, I still would not be able to beat Magnus Carlsen, my inability is actually overdetermined. Why cling to objective rightness as long as it simply gets us into trouble?

I take this dilemma to strongly suggest that there is something wrongheaded with the idea that objective rightness and the OIC principle can be made to work together. And, as long as we retain a subjective sense of rightness, this might not be a very large concession to make. Consider the *Strong Medicine* case, where physician Sue has to choose between doing nothing or giving one of two drugs, where she believes drug *x* will partially cure Ron and she believes drug *y* may cure him completely, but may also kill him; actually it will fully cure him. What it shows is perhaps not that giving Ron drug *y* would have been objectively *right*. Perhaps all it shows is that it would have been *better* to have done so. It would also have been better for me to write down the cure for cancer, and to magically make everyone in the world

¹⁷⁵ Of course, it might be that you ought to find out how and then do it, but that is another matter. Also, some have argued that cases of *culpable* ignorance show that OIC is not always applicable. For example if you are presently unable to get to work on time, you still ought to be at work on time because you have an obligation to be on time. I shall take it up in section 4.10.

happy, but are any of these claims really that interesting from a moral point of view? Giving Ron the drug that would cure him is the best thing I can do physically (in a non-deterministic universe), curing cancer is the best thing I could do if I knew how, and making everyone happy is the best thing I could do in some other universe. All of them are better than giving Ron the drug that partially cures him, but none of them seem to be options that are relevant in the required sense. If we add the fact that rejecting objective rightness solves the problems associated with OIC and objective rightness in a deterministic world, it seems to me that what Parfit has shown is not that wrongness is rendered obsolete if we base OIC on what we can do with our actual motives, but rather that *objective rightness* becomes obsolete.

The reason we thought we needed an objective sense of rightness to begin with was to cope with cases like *Strong Medicine*, where physician Sue has to choose between different actions. She can either do nothing, which will render the condition permanent, or administer drug *x* which will partially cure him, or administer drug *y* which she thinks may cure him, but she also thinks may kill him. Actually, drug *y* will fully cure him, but she could not have known this. The idea is that administering drug *y* is objectively right because it is the option that will actually have the best consequences, but of course, if determinism is true, there is really just one option. It is no more likely that Sue will cure cancer or grow wings than give Ron drug *y* if giving drug *x* is what she is going to do. But, if we flip the question and rather ask what options she *thinks* she has, there *is* a set of available options, and this is highly relevant, because then it seems that we can say that Sue acts rationally or irrationally in choosing *x* or *y*, and we have a set of non-arbitrarily chosen options from which she can choose. She is, after all, only rational in choosing among the options she thinks are available. But the *objectively* right option cannot be determined as the best option among a set of options derived from the *subjective* criterion of what Sue *thinks* she can do. The objectively right option has to be the best among the options she actually can do, which is to give drug *x*.

In other words, *Strong Medicine* and other objective/subjective rightness cases is strongly dependent on there being some sort of epistemic deficiency in order for there to be a difference between what she subjectively ought to do and what she objectively ought to do, after all, if she knew that drug *y* would fully cure Ron, the difference between what is subjectively right and what is supposedly objectively right disappears. As long as there is an answer to what she subjectively ought to do, as I argued in section 4.4, then wrongness is not obsolete as such, you simply have not done the *morally* wrong thing if you have done what

you sincerely believed was best according to consequentialism, which is administer drug *x*, though it can still be *worse* than giving drug *y*. In at least one sense, you did the best you could. The relationship between bestness and rightness is not necessarily one to one, rightness depends on what you believe to be possible (and maybe what you ought to have believed was possible), whereas goodness and badness are objective states of affairs that simply are or are not better or worse.

I think that objective rightness falters even if determinism is not true. Howard-Snyder provides a separate argument for why this is true, but it seems even clearer that this is a reasonable way to go as long as determinism is true, as it shows how we can make sense of OIC in a deterministic world. Not everyone will be willing to accept this, though, which is why I try to provide a solution that is open to objectivists in the next section.

4.8.2 OIC Does Not Apply to Objective Rightness

A hard-nosed objectivist can still insist on there being an objective sense of rightness, but then she may have to accept one of these other unsatisfactory solutions. But, perhaps there is one final way open to the hard-nosed objectivist. She could deny that OIC *applies* to objective rightness. That, I suppose, would mean that the best *logically* possible act, motive, etc. would be the objectively right act.

On the one hand, it is difficult to imagine a more punishing solution, it would mean that nobody has ever done the objectively right thing, but this does not mean that it makes no difference whether you produce 100 utiles or 1000 utiles. First of all, some consequentialists believe it can be *more right* to do one right act than another right act, and also less wrong to do one wrong act than another wrong act.¹⁷⁶ Furthermore, no consequentialist would say that if you have a choice between doing an act that produces 1000, 100, or 10 utiles, and for some reason you choose not to produce 1000, then it is morally irrelevant whether you produce 100 or 10. In other words, you could still lead as good a life as you physically can, which would be better than many alternatives, and since you are still able to do the subjectively right thing, you would not be blameworthy for failing to make everyone magically happy (for example). This would resemble *scalar consequentialism*, which dispenses with deontic evaluations

¹⁷⁶ Martin Peterson, *The Dimensions of Consequentialism: Ethics, Equality and Risk*, (Cambridge: Cambridge University Press, 2013), 10–12.

entirely and only evaluates in terms of ‘better than’ or ‘worse than’, except that it would be scalar only for objective rightness and not rightness as such.¹⁷⁷

Being doomed to do the wrong thing *objectively* might not matter that much. The fact that you will not be able to do what would have been the objectively best thing is a sad fact about the world, but that actually seems to be the case in *Strong Medicine* as well. You could do the subjectively right thing or the objectively right thing, *but not both*, so an inability to do the right thing seems unavoidable.

As long as subjective rightness is still in place, wrongness as such is not obsolete though. Arguably, subjective rightness is the most important form of rightness, since it is what acts as a guide to how we should live our lives; it produces an answer to what ought to be done that actually matters to us when we deliberate, and that acts as the basis for blameworthiness.

I have argued that if psychological determinism is true because physical determinism is true, then all ethical theories result in evaluative conflict. This shows that evaluative conflict does not work as an argument against global consequentialism specifically, as the problem has a scope that is much larger than global consequentialism. But, the problem doesn’t disappear just because it is a general problem. We should still figure out how to solve it. I have proposed two ways of solving it. We could reject objective rightness, this has the advantage that it has been proposed by other philosophers for separate reasons, so it could hold no matter what, and if it solves the problems for objective rightness that arise if determinism is true then I think this option deserves to be taken very seriously.¹⁷⁸ That is, it seems to me that I really *cannot* beat Magnus Carlsen at chess. Or, finally, we reject that the OIC rule applies to objective rightness. On the two latter views, subjective rightness would be rightness proper.

4.9 Psychological Determinism *Without* Physical Determinism

When discussing determinism up until now, I have been assuming that the reason to believe in psychological determinism is because of physical determinism, or that non-deterministic

¹⁷⁷ Alastair Norcross, “The Scalar Approach to Utilitarianism,” in *The Blackwell Guide to Mill’s Utilitarianism*, ed. Henry West (Malden: Wiley-Blackwell, 2006), 217–232.

¹⁷⁸ Frances Howard-Snyder, “The Rejection of Objective Consequentialism,” 241–248; and Elinor Mason, “Objectivism, Subjectivism, and Prospectivism,” 184–185.

physical properties do not affect our actions whatsoever. As I have argued, if some of our acts are genuinely random, psychological determinism is not true, and it does not follow that wrongness is obsolete if we base OIC on the categorical sense of ‘can’. I have also argued that even if there is some degree of randomness to what acts we perform given a specific set of motives and initial conditions, there could be some actions that are still not possible, like withstanding torture, or refraining from torturing after taking a motive-altering pill. In the previous section I argued that even if all acts are determined by motives, so that only one act is possible, there may still be hope for the OIC principle, essentially in retreating to the view that subjective rightness is the proper sense of rightness.

But, it might be that *all* acts are psychologically determined by motives, but that motives, dispositions, character traits, etc. are not. As far as I can tell, there is no evidence to support this, but it cannot be simply dismissed. Say if motives (for example) are not determined, whilst actions (for example) are always determined by motives; does it follow that there will be no wrong acts?

I don’t think so, at least not directly. In this specific case I think something like Mason’s reply to Parfit would work. As I mentioned in section 4.1, Mason claims that even if acts are determined, motives need not be. I rejected this reply because it rests on the rather strange position that acts are determined by motives, but motives are not determined, but in this section I discuss what we should think if it is true in spite of its admitted strangeness. What her reply means is that Parfit has simply shifted wrongness to the level of motives or dispositions rather than acts.¹⁷⁹ As Adams writes:

We do not say that a person rightly jumped from a burning building, saving his life, but wrongly struck the ground, breaking his leg. It is unreasonable, on this view, to separate, for moral evaluation, actions that are causally inseparable.¹⁸⁰

If acts are determined by motives, but motives are not themselves determined then it would seemingly be unreasonable to separate the two for moral evaluation, since acts would then follow motives in much the same way that hitting the ground follows jumping from a burning building. It would be unreasonable to separate, for moral evaluation, the right or wrong motive from the act that necessarily followed from it. (Though, a motive might be right

¹⁷⁹ Elinor Mason, “Against Blameless Wrongdoing,” 291.

¹⁸⁰ R.M. Adams, “Motive Utilitarianism,” 473.

because of the acts that are going to follow it, from a consequentialist point of view. This is just how consequentialism works.) So the motives could be wrong, and the acts follow the motives.

The question becomes how to interpret this. Is it right to jump from the burning building and *right* to strike the ground too? Or is it right to jump from the burning building, and morally *neutral* to strike the ground? The former might be something akin to a genuine transfer or inheritance of rightness perhaps, as I discussed in chapter two. Perhaps it does not even make sense to call it two different acts as Adams does here. Perhaps striking the ground is simply *a part* of the act of jumping from a burning building. In our case, where we are assuming that acts necessarily follow from motives, the act of jumping would similarly be a part of having the right motives, so the resulting acts are either right or neutral as well. I don't think it would make sense to call one part of an act right and another wrong, but perhaps I am wrong about that. My two cents is that the resulting acts would be neutral if they really are separate acts.

The point is that wrongness is not rendered obsolete no matter which of these possibilities we opt for. If we assume that wrong acts necessarily follow wrong motives and right acts right motives, then there *are* wrong acts, so this option is unproblematic.¹⁸¹

If acts are an ineliminable part of a right or wrong motive, then I think it makes no sense to call one part right and one part wrong, and this explains *why* there are no wrong acts. There are no such acts because it is the *conjunction* of the motive and act that is the proper evaluative focal point, and it is either right or wrong. Or, in other words, having the right motives and whatever acts follows from that would be the best possible life you could live, since the best motives would be the ones that maximized value, and this explains why it would be strange to call the act-part wrong. Furthermore, if it *does* make sense to call one part right and one wrong, then the wrongness of acts is not obsolete. Then it could be wrong to strike the ground after jumping from a burning building even if jumping was right, and it could be wrong of Molly to torture people after taking the motive-altering pill, even though acquiring the motives was right. I don't find this plausible since it conflicts with OIC, I agree

¹⁸¹ In fact, if we do assume that motives determine acts, but motives are *not* determined, it also follows that your surroundings are not determined, since your surroundings contain people with undetermined motives which may therefore lead to different acts. In other words, your acts are not strictly speaking determined either, but I will leave this to the side, since it feels like a cheap shot to make.

with Adams that it would be more reasonable to deny that focal points that are causally inseparable can be meaningfully separated for moral evaluation. Also, why insist that they are one separate evaluative focal point in the first place, only to also insist that you have to evaluate each part separately? That appears unmotivated to me. Rejecting this latter option seems like a reasonable route for a global consequentialist to take if we assume that acts are an ineliminable part of a right or wrong motive, whilst the former is unproblematic.

If the acts that necessarily follow from wrong motives are morally neutral, then the wrongness (and rightness) of *acts* is obsolete (all acts would be neutral), but wrongness as such is not obsolete. Once again, the explanation for this is that if it is “unreasonable . . . to separate, for moral evaluation, actions that are causally inseparable”, it is also unreasonable to separate a motive and an action that are causally inseparable.¹⁸² That would break with OIC, since the act would be the best one you could possibly do provided you have the right motives. You could do a better act with worse motives, but that would necessarily mean that you did not maximize value, whilst having the best motives would maximize value. This is the same argument that was given in chapter two, which concluded that the conjunction of the best motive and the best act are not jointly obligatory if the conjunction is impossible. This last point is what Parfit’s argument was supposed to show could not work, but since wrongness is not obsolete if motives can be wrong, I don’t find his argument particularly convincing.

I am not entirely sure which of these options is the correct one, but I do think there is a reasonable option available to the global consequentialist no matter which of them is correct. Global consequentialists should see which focal points are determined by which other focal points, and then not separate, for moral evaluation, focal points that are causally inseparable. This is just the global consequentialist version of the principle Adams proposed for actions. Of course, Adams thought that acts and motives are separable in a way that jumping from a burning building and hitting the ground is not, but if for some reason motives are not determined whilst acts are determined by motives, we should treat acts and motives as inseparable. Wrongness does not become obsolete—only the wrongness of acts separately

¹⁸² R.M. Adams, “Motive Utilitarianism,” 473.

becomes obsolete—and that is explained by the fact that they are determined whilst our motives are not.¹⁸³

Now, even though I think there are some good options available to the global consequentialist if it is true that motives are not determined whilst acts are determined by motives, I do not think that this kind of act-specific determinism is very likely. There is, to the best of my knowledge, absolutely no evidence to support this hypothesis, which means that it appears *ad hoc*. I simply find it difficult to see why one would be a determinist about acts but not about motives.

4.10 Rejecting ‘Ought’ Implies ‘Can’ Outright

A final response to evaluative conflict which is open to a global consequentialist, given the problems encountered if determinism is true, would be to reject OIC outright. If ‘ought’ does not imply ‘can’ then it is not clear that cases like *Hooker’s Paradox* are troublesome. It may seem strange that you ought to follow the best possible decision procedure, and you ought to do the best possible act, and you cannot do both. Yet, if ‘ought’ does not imply ‘can’ then this just might be one of those cases where you ought to do both in spite of the fact that you cannot. Of course, you may think that this does not *solve* evaluative conflict, but rather is a way of accepting it. In this section I will be talking about both subjective and objective rightness, since rejecting OIC solves the problem for both senses of rightness, and I eschew determinism for the chapter’s duration.

Some philosophers reject the principle, largely because they think they have found counter-examples. The best way of illustrating their point is by borrowing a short dialogue from the beginning of Peter B. M. Vranas’ “*I Ought, Therefore I Can*”.

“Good morning, Professor. Unfortunately, I haven’t even started writing a paper yet. I know I promised to turn in a paper today by 9 am, but last night I didn’t feel like writing, so I went to the movies instead. What should I do?”

“You should turn in a paper by 9 am. It’s 8:57, so you’ve got three minutes left.”

“Sorry, Professor, maybe I didn’t make myself clear. I don’t have a paper to turn in.”

“You asked what you should do. Obviously, you should fulfill your obligations. You may not have a paper, but you do have an obligation to turn in a paper by 9 am.”

“But I can’t turn in a paper by 9 am, so I don’t have an obligation to do so: ‘ought’ implies ‘can’.”

¹⁸³ Of course, this means that our motives are determined entirely randomly within a set of possibilities, which is not particularly satisfying, but it is not a problem that is unique to global consequentialism. Rather, global consequentialism offers a possibility of saying that wrongness exists, i.e. *wrong motives* are wrong, even though your acts are entirely determined by motives.

“This principle would make a joke of morality: it would allow people to get rid of unwanted obligations by making themselves unable to fulfill them.”

“But with all due respect, Professor, your denial of the principle also makes a joke of morality: rather than making morality excessively lax, it makes morality absurdly demanding.”

“Was it absurd to demand that you write a paper rather than going to the movies last night?”

“No, it wasn’t absurd, but it is absurd to demand that I turn in a paper now. How can you advise me to do what you know I can’t do?”

“I’m not advising you to turn in a paper by 9 am; I’m simply remarking that you have an obligation to do so.”

“But what’s the point of morality if it’s useless for advice?”

“Morality is not useless for advice: yesterday it would have given you very useful advice indeed. As for now, I’m afraid there is nothing you can do: there is no advice to give you. So the fact that you get no advice now from a morality which denies that ‘ought’ implies ‘can’ does not show such a morality to be defective.”

“But surely, Professor, there is something I can do: I can turn in a paper tomorrow. My morality is not excessively lax: I recognize a compensatory obligation to accept a penalty for late submission.”

“Talk of compensatory obligations may be appropriate when you have to pay a penalty for late repayment of a loan, but is out of place here: you don’t have to compensate me for anything. It won’t do to turn in a paper tomorrow because I have to submit the course grades today by 10 am and I emphasized I would give no incompletes. It’s now 9 am. I’m sorry, but you just failed the course.”¹⁸⁴

Vranas believes that the Professor has the upper hand in this exchange, and unless we are presupposing consequentialism, I’m inclined to agree. However, if we do presuppose consequentialism I do not. The view on obligations proposed by consequentialists is closely tied with what you can and cannot do. It usually identifies what you ought to do as that which is both *possible* and *best*. It does *not* admit of obligations that are somehow disconnected from this, like generalized obligations to ‘not lie’ or ‘turn in your assignments on time’. The question, then, is whether letting the student turn in is possible and whether it would be best. Presumably it is neither. The student probably will not be able to write a paper in an hour, yet if the student is able to do this, and doing so would be best then the Professor ought to let the student turn it in late. It might *not* be best, since letting students turn in late will promote laziness, but *if* it is best and possible then it is right. The Professor having the upper hand seems to assume a non-consequentialist view of obligation.

Another way of saying this would be to say that even if you do take OIC out of consequentialism, you would still have to array the different possible alternatives according to what is available somehow, so in this sense OIC is a part of the very nature of consequentialism itself.

¹⁸⁴ Peter B. M. Vranas, “I Ought, Therefore I Can,” *Philosophical Studies* 136, no. 2 (2007): 167–168.

Does this make a joke of morality? Not obviously. Even though the student does not have an obligation to turn in the assignment *now*, he did yesterday if we assume that this was the best action available to him at the time. In other words he has *failed* to do the right thing; he did not get off scot-free. The way the student “got rid of an unwanted obligation” was by not fulfilling it, thus rendering himself *blameworthy*, and not only is he blameworthy, he did something *wrong*. Yesterday he had an obligation to write the essay in a timely manner, and not doing so was straightforwardly wrong.

Arguably this is always the case with time-limited obligations; if you do in fact break an obligation you are no longer bound by it. If you had an obligation to come to work on time, yet you did not arrive on time, you are not obligated now to fulfill it retroactively. If you are supposed to be at work at 9, and you call your boss at 9:10 telling him you won’t be able to make it until 9:20, the reasonable thing for your boss to answer is to say that you should get to work as fast as you *can*, not that you should get to work ten minutes ago. How can you (presently) have an obligation to be at work as fast as you can *and faster* than you can, that is, ten minutes ago?

To be clear though, there is absolutely no automaticity whereby the moment you break an obligation, you are somehow *entitled* to a new obligation toward the person you had an obligation to in the first place; that the professor should have to extend the deadline for example. In my work case, it might just as easily be that the reasonable response is to fire you on the spot, if arriving on time is very important, or if you are frequently late for example. This, I think, is where the Professor is intuitively right; he is entirely within his rights to fail the student. But, the student is right that she does not have an obligation to turn in the assignment faster than she is able to *now*.

In other words, I am challenging the assumption that the Professor makes in the dialogue, that applying OIC to these kinds of cases would make a joke morality by allowing “people to get rid of unwanted obligations by making themselves unable to fulfill them”. It is true that you can do this, but the way you do so is by doing the wrong thing, e.g. not turning in your assignment on time. And you certainly cannot simply refuse to do your assignment on time and then demand a new deadline. You may ask for one, and the professor or the university can choose whether to give you one based on whatever criteria they deem

pertinent, none of which include the fact that you cannot hand it in on time. So, what is supposed to be the problem?

Perhaps another example would be in order, consider the following case, taken from Michael Stocker's "*Ought' and 'Can*".

Jones promises to repay Smith tomorrow, and thus, to meet him at the bank between twelve and twelve-fifteen, which, as Jones knows, is Smith's only free time. At eleven the next day, fully aware that unless he leaves very soon, he will be unable to get there on time, Jones starts reading a novel; he puts it down at one.¹⁸⁵

Stocker thinks that OIC denies that Jones acted in a blameworthy fashion in not meeting and repaying Smith at the appointed time "for, between twelve and twelve-fifteen, he was in his study a good hour's drive from the bank; and no one can travel in an instant or even fifteen minutes a distance requiring sixty minutes."¹⁸⁶ But this seems like a rather peculiar claim. It seems to me that the statement 'Jones is blameworthy for not meeting Smith' is not precise enough. Put differently, Jones clearly *is* blameworthy for not meeting him. That is, he is blameworthy for not getting in his car at the necessary time in order for him to fulfill his obligation, and he is blameworthy precisely *because* it makes him unable to fulfill his obligation. But Jones is not *additionally* blameworthy for failing to insta-travel to Smith between 11:16 and 12:15.

There are a couple of reasons for this, first of all because it seems this would lead to an inflation of blameworthiness. Jones would seemingly be more and more blameworthy for each instance of not insta-travelling to Smith at all times between 11:16 and 12:15. Another reason why he is not additionally obligated is essentially the same as in *Ellie's Brain Scan*. If Jones, at 12:00 ought to repay Smith at 12:15, then he should apparently be trying to do so, yet this would clearly not be the most rational way for him to act. Should he be speeding along the motorway at breakneck speed in an attempt he *knows* will not succeed? Or should he be clicking his heels three times, saying "there's no place like the bank, there's no place like the bank, there's no place like the bank"? Saying that Jones ought to be where Smith is, at 12:15 gives him no guidance as to what he ought to be doing. And remember that what we are currently considering is rejecting OIC outright, not simply rejecting it for objective rightness.

¹⁸⁵ Michael Stocker, "'Ought' and 'can'," 315.

¹⁸⁶ *Ibid.*

If consequentialism is false, I think Vranas and Stocker might be onto something, but I don't think consequentialism is false. As long as we assume consequentialism, as I do in this thesis, I don't think these counter-examples cut the mustard. Consequentialism is simply too closely connected to the notion that obligation is tied to what is currently possible, so the Professor loses the debate as far as OIC is concerned. (But it does not follow that he should accept the paper.) Additionally, I don't understand how failing to live up to one's obligations means you lose the obligation makes a joke of morality, since you are rendered blameworthy by breaking them, and you cannot demand a new obligation in their stead.

4.11 Recap

I have argued that mixing OIC with determinism seems to get us into trouble no matter what. *Ellie's Brain Scan* cuts across theories, no theory can accept the hypothetical sense of 'can' as the basis for OIC. So does the argument that if psychological determinism is the product of physical determinism, then all theories lead to wrongness being obsolete, as all theories accept that motives *exist* and have causal force. All theories accept psychological determinism if they accept physical determinism, and if so they have the same problem with OIC as global consequentialism does, which is that wrongness becomes obsolete if you ought to do the best you can with your *actual* motives.

Parfit's argument appears to be a bit too clever, because we end up in a quagmire from which there is no easy escape. However, as long as we are clear about the distinction between objective and subjective rightness I think there are some ways to escape the worst of these troubles, and in a way that does not entail evaluative conflict. The general form of my argument, then, is that we get into trouble no matter what, and so, even if it might initially sound somewhat counter-intuitive, it is not really that strange if global consequentialism tells us to give up on objective rightness. (This is especially true since we should give up on objective rightness even if determinism is false, and even if global consequentialism is false.)

Two final alternatives remain. One claim that psychological determinism obtains, but not physical determinism does not, which I find *ad hoc*, and it also just shift wrongness to the level of motives if it should happen to be true. Or, one could reject OIC outright, which I find to work poorly for consequentialism in general, and I do not think the counter-examples Vranas mentions hold water since it would not make a joke of morality to accept that

obligations disappear once they are not fulfilled. Next on the agenda is to see how well this works for other focal points than motives and acts.

5 UNIVERSALIZING MASON'S DEFENSE

All of my responses to the proposed problem of evaluative conflict have used Mason's reply as a template, but up until now I have mostly considered one type of evaluative conflict, cases where the right *motives* lead to the wrong *acts*. But what about evaluative conflict between other evaluative focal points; will Mason's reply work for those cases too? I cannot go through all possible types of evaluative conflict, but I can assess evaluative conflict between decision procedures and acts, and between rules and acts. Along with motives, these are arguably the most important evaluative focal points. I would have liked to also consider character traits, virtues, laws, etc. but it will simply not be possible to do so in one hundred pages.

5.1 Decision Procedures and Acts

In *Hooker's Paradox*, a decision procedure conflicts with an act.

Suppose, on the whole and in the long run, the best decision procedure for you to accept is one that leads you to do act x now. But suppose also that in fact the act with the best consequences in this situation is not x but y . So global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. That seems paradoxical.¹⁸⁷

I would argue that something like Mason's reply holds in this case too, since you either *can* or *cannot* accept a decision procedure *in conjunction* with not doing the act that the decision procedure picks out. If you can, then what is paradoxical about it? If you cannot, then why would a global consequentialist say you ought to both follow the decision procedure and not do the act?

If you are able to do so then you ought to do so, that is, accept the best decision procedure and do the best act you can do, which is not one that the decision procedure picks out. This is not paradoxical, though it might sound weird. Or at least, if it is paradoxical it is paradoxical because it is somehow possible to accept a decision procedure and not do what it picks out, *not* because it gives the wrong answer to what you ought to do given the fact that it is paradoxically possible to follow a decision-procedure and not do the act it oicks out. As

¹⁸⁷ Brad Hooker, "Rule Consequentialism," § 5.

long as you really can do so you ought to do so, no matter how paradoxical it might happen to be.

I don't know *how* you could accept a decision procedure and not do the act it picks out; it is difficult to ferret out what accepting it means if it does not mean to follow. The reason for this is that one important connection between a decision-procedure and an act is logical, and not causal. Presumably, the way you would be unable to follow a decision procedure and not do the act it picks out is that accepting the decision procedure somehow *means* using the procedure, and if you use the procedure you would end up with a decision, as it is a decision-procedure. The question then becomes whether you can decide to do something, then not do it, and still qualify as following the decision-procedure. This is a question I do not know the answer to. (There are causal connections as well of course, but this seems the biggest difference between the motive/act cases and decision-procedure/act cases.) There *is* a difference between acceptance and following, so it might be that accepting merely means *believing* it is the best decision procedure to follow, for example. If so there is no obvious logical problem, but if accepting it implies following it, then there might be a logical problem.

I do not know the answer, but I think the point is moot if you can do so. If you really can accept a procedure yet not do the act it picks out, then you ought to do so simply because it is possible and best.

What should the global consequentialist say if you *cannot* accept a decision procedure and not do the act it picks out? That is, if following the decision procedure *means* doing the acts perhaps, or if doing the acts necessarily results from following the procedure. Well, then the so called paradoxical feature of Hooker's Paradox will not be possible. It will be false that global consequentialism tells you to accept the decision procedure and do act *y*. Rather, global consequentialism will tell you to choose one of the two, or if you already accept the decision procedure it will tell you that you ought to do act *x*. This is obviously parallel to Mason's defense.

5.2 Rules and Acts

What about rules and acts; does global consequentialism lead to evaluative conflict between rules and acts? First of all, it is impossible to *comply* with a set of rules without doing the acts

that the rules pick out. It just cannot happen logically, since complying with a set of rules *means* doing the acts it picks out. That means that you can either comply with the set of rules or do some other act, a global consequentialist would then say that the best option is what you are obligated to.

But compliance is not the only relation you can have with a set of rules. Rules can be complied with, flouted, accepted, etc.¹⁸⁸ If you accept a set of rules it may not follow that you will always act in compliance with it. Unsurprisingly though, a global consequentialist will then tell you to accept them if it has the best possible consequences, and then do the best possible acts within that framework. If it will maximize value you should break the rules as often as you can while still accepting the set of rules.

I could go on, but at some point this reiteration of the same point will become tedious and unproductive. It should be clear that my response will be that for any evaluative focal point, and any relation you could have with that evaluative focal point, a global consequentialist should have, do, be, follow, comply with, etc. whatever mix of evaluative focal points that will maximize value. But it is important to note that this also means that if evaluative focal point x will maximize value, then it follows from this that doing, having, being, following, or complying with evaluative focal point y cannot both be incompatible with x and maximize value. As a matter of logic, if y makes x impossible, then y will not maximize value, because x maximizes value. The easy way to say this is that one should x and y only if the conjunction of x and y -ing will maximize value, but the two statements are compatible.

¹⁸⁸ Accepting a rule roughly means believing that it ought to be followed or complied with.

6 CONCLUSION

As the flowchart below indicates, there are a number of ways for a global consequentialist to avoid evaluative conflict. None perfect, but there seems to be no good ways to interpret OIC if determinism is true.

I started out by arguing that Mason is right in claiming that Molly from *My Malevolence* does not have an obligation to refrain from torturing people after taking the motive-altering pill, that is, *if* she really *cannot* refrain from doing so. This is explained by the fact that some form of torture would be the *best* thing she could possibly do. This holds as long as we are considering her subsequent actions in the appropriate comparison class; that is, among those actions she can do as long as she has the motives she is morally *obligated* to have. In other words, it is explained by the fact that not having the motives and not torturing people would be *worse* than having them and torturing people. I also argued that this is essentially identical to how a classical consequentialist would reason if she were faced with a choice between two *acts* that cannot both be done. But Parfit argues that this understanding of OIC leads to absurdities.

OIC tells us that if I could not have done otherwise, then I ought not have done otherwise. But Parfit argues that this cannot mean (2) that acting differently would be impossible given my *actual* motives, because that would make wrongness obsolete as long as psychological determinism is true. He proposes that the claim is rather (3) that you can appeal to OIC if acting differently would have been impossible even with different motives.

I have several objections to this. There is currently no consensus as to whether determinism is true, since quantum physics may be fundamentally non-deterministic. Nor is there any consensus as to whether this potential non-determinism affects our decisions. In other words, it might be true that our actions are determined and it might not. Traditionally, quantum randomness has been seen as just as problematic for free will as classical determinism, so this lack of consensus does not make much of a difference in the free will debate. But, it does matter for whether basing OIC on your *actual* motives renders wrongness

obsolete. If some of our acts are partially random, for example, then wrongness is not necessarily obsolete, because there will be more than one possible outcome.¹⁸⁹

Based on this we can say that if determinism is false, then Parfit's argument does not appear to work. Parfit considers non-determinism, but he thinks that it will still be impossible to act against one's motives all the time, essentially re-introducing a form of local determinism. I remain agnostic concerning whether breaking the link between motives and acts all the time is possible or not, but I argue that it either is possible to do so, or it is not. If it is then there is no reason *not* to demand that one always has the best motives *and* does the best acts, no matter how difficult this may be. If it is not possible to break the link all the time, as Parfit believes, the upshot may be evaluative conflict. But there will only be evaluative conflict as long as we accept Parfit's hypothetical interpretation of OIC, and I argue that this is a bad assumption as long as determinism is not true. There would be absolutely no reason to think so, since it would not render wrongness obsolete to base OIC on what one is *actually* able to do, which is his entire argument for the hypothetical interpretation of OIC.

This shows that *global* psychological determinism is not necessary in order to mount Mason's defense; some acts may be determined and some not. If so, then you can still appeal to the categorical interpretation of OIC, and you can do so without rendering wrongness obsolete. This is essentially the same argument as the previous one; it just makes it clear that global determinism is not necessary for Mason's defense. When Parfit considers non-determinism he presents his case without distinguishing between whether *no* acts are determined, or some are and some are not. He assumes that we cannot break the link between motives and actions all the time, meaning he essentially assumes the latter, but it seems that no matter which is true, there will be no evaluative conflict.

If we *do* assume global determinism, and the hypothetical interpretation of OIC, then Parfit runs into problems no matter what type of rightness he is talking about. If Parfit's claim (3) is that we *subjectively* ought to do the best we can, whatever our motives might have been, we run into *Ellie's Brain Scan*. I believe *Ellie's Brain Scan* shows that his solution fails for the subjective sense, because it would mean that we subjectively ought to do something we *know* we cannot do, which makes no sense. Luckily, Prichard's interpretation of OIC provides at least some respite. His solution may not be everything I would like it to be, but it shows

¹⁸⁹ I say partially because it is unreasonable to think that me going to the post-office is *entirely* random; some physical and mental events are foreseeable.

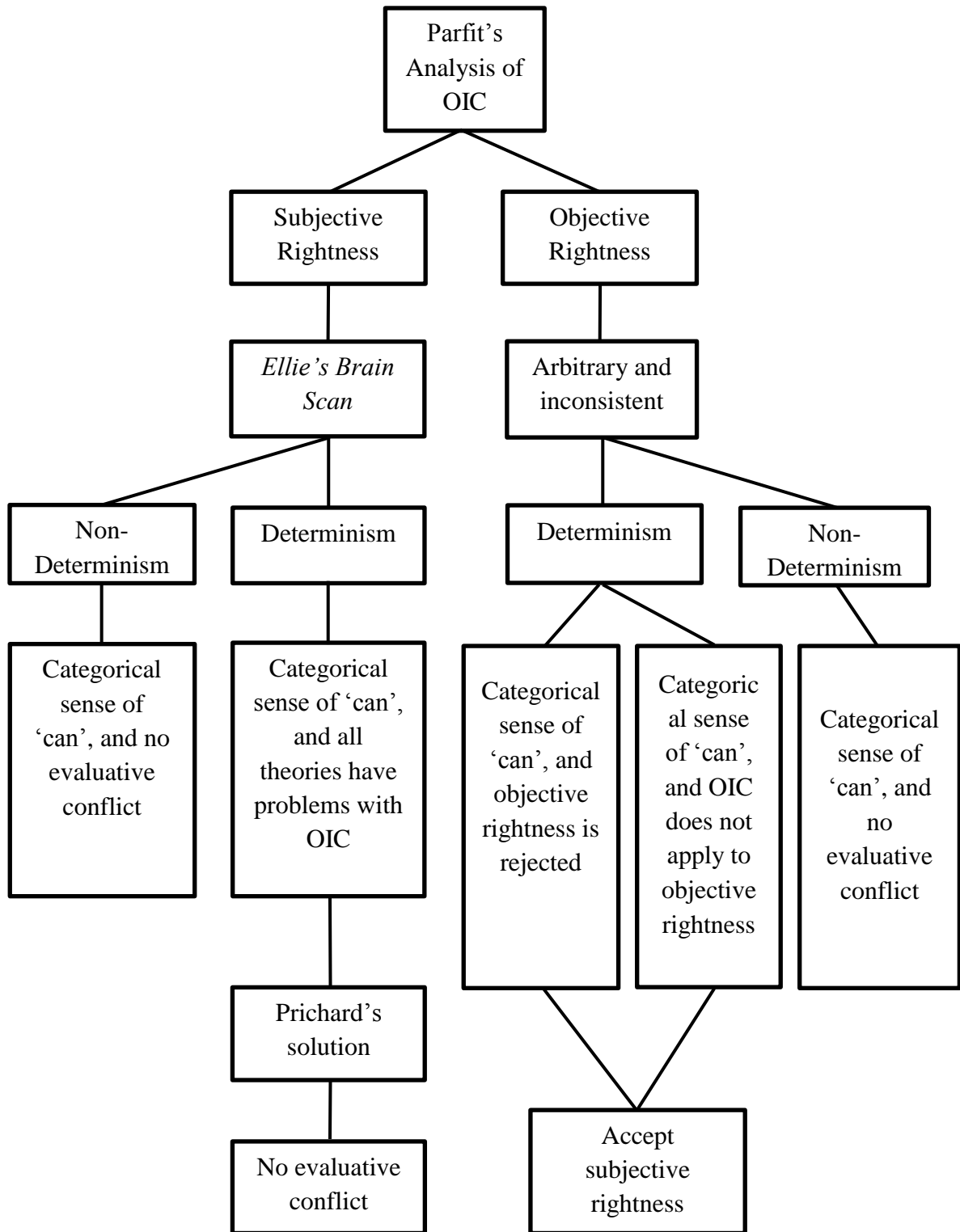
that it is possible to evade *Ellie's Brain Scan* and still retain a conception of rightness. Furthermore, it does not lead to evaluative conflict.

If Parfit's claim (3) is that we *objectively* ought to do the best we can whatever our motives might have been, I have argued that this is inconsistent with how we think about OIC and *physical* determinism. As long as we are assuming that psychological determinism is true *because* physical determinism is true, we are disregarding the fact that the same argument that Parfit is making about psychological determinism, can be made about physical determinism, and it would apparently succeed for the exact same reason, namely that wrongness would be obsolete. This objection cuts across all ethical theories.

I think this puts OIC in a rather awkward position, but it seems premature to abandon it, as it seems to be necessary in some form or another. Therefore I have proposed some alternative solutions. The most promising, in my view, are either rejecting objective rightness, or denying that OIC applies to objective rightness whilst still accepting that it applies to subjective rightness, both of which would mean that wrongness is not obsolete. I am not claiming that what I have argued is ideal. In fact almost every part of it irks me, but when it comes to philosophical questions as slippery as the one I have been discussing, that is to be expected. And even if my solution should fail, I believe I have shown that the hypothetical interpretation that Parfit believes we should accept is not viable, and if so we have a new problem on our hands: how should we interpret OIC given that determinism is true? Furthermore, as long as all theories have these problems with OIC given determinism, global consequentialism does not have a special problem that needs to be addressed.

I also argue that if psychological determinism is somehow *not* true because of physical determinism, but for some other reason, this merely shifts what we can call right and wrong to the level of motives (or whatever evaluative focal point that is not determined).

Finally, I have tried to show that the general form of Mason's defense can be used against potential evaluative conflict between decision procedures and acts, and between rules and acts. Though this is not enough to show that *no* evaluative conflict can occur between any set of evaluative focal points, I believe it shifts the burden of proof further toward the opposition.



BIBLIOGRAPHY

- Adams, R.M. "Motive Utilitarianism." *Journal of Philosophy* 73, no. 14 (1976): 467–481.
- Allhoff, Fritz. "An Ethical Defense of Torture in Interrogation." In *Ethics of Spying: A Reader for the Intelligence Professional*, edited by Jan Goldman. Oxford: Scarecrow Press, inc., 2006.
- Anomaly, Jonny. "An Argument Against External Reasons." *Sorites* 18 (2007): 56–59.
- Bishop, Robert C. "Chaos, Indeterminism, and Free Will." In *The Oxford Handbook of Free Will*, edited by Robert Kane. New York: Oxford University Press, 2002.
- Broad, C.D. *Ethics*. Edited by C. Lewy. Dordrecht: Martinus Nijhoff, 1985.
- Crisp, Roger. "Utilitarianism and the Life of Virtue." *The Philosophical Quarterly* 42, no. 167 (April 1992): 139–160.
- Crisp, Roger. "Well-Being." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. (Summer 2016 Edition, forthcoming.)
<http://plato.stanford.edu/archives/sum2016/entries/well-being/>.
- Driver, Julia. "Global Utilitarianism." In *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller. New York: Cambridge University Press, 2014.
- Eggleston, Ben and Dale E. Miller. Introduction to *The Cambridge companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller. New York: Cambridge University Press, 2014.
- Feldman, Fred. Introduction to *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht: D. Reidel Publishing Company, 1986.
- Foot, Philippa. "Utilitarianism and the Virtues." *Proceedings and Addresses of the American Philosophical Association* 57, no. 2 (November 1983): 273–283.
- Goldman, Holly S. "Doing the Best One Can." In *Values and Morals*, edited by Alvin I. Goldman and Jaegwon Kim. Dordrecht: D. Reidel Publishing Company, 1978. See also Smith, Holly and Smith, Holly M.
- Goldman, Holly S. "Dated Rightness and Moral Imperfection." *The Philosophical Review* 85, no. 4 (October 1976): 449–487. See also Smith, Holly and Smith, Holly M.
- Hodgson, David. "Quantum Physics, Consciousness, and Free Will." In *The Oxford Handbook of Free Will*, edited by Robert Kane. New York: Oxford University Press, 2002.
- Hooker, Brad. "The Demandingness Objection." In *The Problem of Moral Demandingness: New Philosophical Essays*, edited by Tim Chappell. London: Palgrave Publishers, 2009.

- Hooker, Brad. "Rule Consequentialism." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. (Winter 2015 Edition.)
<http://plato.stanford.edu/archives/win2015/entries/consequentialism-rule/>.
- Howard-Snyder, Frances. "The Rejection of Objective Consequentialism." *Utilitas* 9, no. 2 (1997): 241–248.
- Howard-Snyder, Frances. "Response to Carlsen and Qizilbash" *Utilitas* 11, no. 1 (1999): 106–111.
- Jackson, Frank and Robert Pargetter. "Oughts, Options, and Actualism." *The Philosophical Review* Vol. 95, no. 2, (1986): 233–255.
- Kagan, Shelly. "Evaluative Focal Points." In *Morality, Rules, and Consequences*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller. Edinburgh: Edinburgh University Press, 2000.
- Kant, Immanuel. *Religion Within the Boundaries of Mere Reason*, translated by George di Giovanni. In *Religion and Rational Theology*, translated and edited by Allen W. Wood and George di Giovanni. Cambridge: Cambridge University Press, 1996.
- Louise, Jennie. "Right Motive, Wrong Action: Direct Consequentialism and Evaluative Conflict." *Ethical Theory and Moral Practice* 9, no. 1 (2006): 65–85.
- Mason, Elinor. "Against Blameless Wrongdoing." *Ethical Theory and Moral Practice* 5, no. 3 (2002): 287–303.
- Mason, Elinor. "Objectivism and Prospectivism About Rightness." *Journal of ethics & Social Philosophy* 7, no. 2 (March 2013): 1–21.
- Mason, Elinor. "Objectivism, Subjectivism, and Prospectivism." In *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller. New York: Cambridge University Press, 2014.
- McKenna, Michael and Justin D. Coates. "Compatibilism." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. (Summer 2015 Edition.)
<http://plato.stanford.edu/archives/sum2015/entries/compatibilism/>.
- Melville, Herman. *Moby Dick or The White Whale*. Boston: The St. Botolph Society, 1892.
- Mill, John Stuart. *Utilitarianism*, edited by Roger Crisp. New York: Oxford University Press, 1998.
- Mulgan, Tim. *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations*. Oxford: Oxford University Press, 2006.
- Iris Murdoch. *The Bell*. New York: Penguin Books, 1958.

- Norcross, Alastair. "The Scalar Approach to Utilitarianism." In *The Blackwell Guide to Mill's Utilitarianism*, edited by Henry West. Malden: Wiley-Blackwell, 2006.
- Ord, Toby. "Beyond Action." PhD diss., University of Oxford, 2009.
<http://www.amirrorclear.net/academic/papers/beyond-action.pdf>.
- Ord, Toby. "How to be a Consequentialist About Everything." Lecture performed at ISUS X (2008): 1–24. <http://www.amirrorclear.net/academic/papers/everything.pdf>.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Parfit, Derek. *On What Matters vol. 1*. Oxford: Oxford University Press, 2011.
- Peterson, Martin. *The Dimensions of Consequentialism: Ethics, Equality and Risk*. Cambridge: Cambridge University Press, 2013.
- Pettit, Philip and Michael Smith. "Global Consequentialism." In *Morality, Rules, and Consequences: A Critical Reader*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller. Edinburgh: Edinburgh University Press, 2000.
- Portmore, Douglas W. "Foundational Consequentialism and its Primary Evaluative Focal Point." Working paper, accessed June 28, 2014.
<https://dl.dropboxusercontent.com/u/14740340/Foundational%20Consequentialism.pdf>.
- Prichard, H.A. "Duty and Ignorance of Fact." In *Moral Writings*, edited by J. MacAdam. Oxford: Clarendon Press, 2002.
- Prichard, H.A. "Duty and Ignorance of Fact." In *Moral Obligation and Duty and Interest*. Oxford: Oxford University Press, 1968.
- Ross, W.D. *Foundations of Ethics*. Oxford: Clarendon Press, 1939.
- Russell, Bertrand. "The Elements of Ethics." In *Readings in Ethical Theory*, edited by Wilfrid Sellars and John Hospers, 2d edition. New York: Appleton-Century-Crofts, 1970.
 Originally published in *Philosophical Essays*. London: Longmans, Green & Co., 1910.
- Sinnott-Armstrong, Walter. "Are Any Moral Beliefs True?" In *Moral Skepticisms*. Oxford: Oxford University Press, 2006.
- Smith, Holly M. "Subjective Rightness." *Social Philosophy and Policy* 27, no. 2 (2010): 64–110.
- Smith, Holly M. "The 'Prospective' View of Obligation," discussion note in *Journal of Ethical and Social Philosophy*. (February 2011): 1. See also Smith, Holly and Goldman, Holly S.
- Smith, Holly. "Culpable Ignorance." *The Philosophical Review* 92, no. 4 (October 1983): 543–571. See also Smith, Holly M. and Goldman, Holly S.

- Sobel, David. "The Impotence of the Demandingness Objection." *Philosophers' Imprint* 7, no. 8 (2007): 1–17.
- Stocker, Michael. "'Ought' and 'can'." *Australasian Journal of Philosophy* 49, no. 3 (1971): 303–316.
- Smart, J.J.C. "Extreme and Restricted Utilitarianism." *The Philosophical Quarterly* 6, no. 25 (October 1956): 344–54.
- Smart, J.J.C. and J.J. Haldane. *Atheism and Theism*, 2nd edition. Malden: Blackwell Publishing Ltd., 2003.
- Timmons, Mark. Introduction to *Conduct and Character: Readings in Moral Theory*, 6th edition. Boston: Wadsworth Cengage Learning, 2006.
- Vranas, Peter B. M. "I Ought, Therefore I Can." *Philosophical Studies* 136, no. 2 (2007): 167–168.
- West, Henry R. "Mill's 'Proof' of the Principle of Utility." In *An Introduction to Mill's Utilitarian Ethics*. Cambridge: Cambridge University Press, 2004.

