On Cluelessness

Patrick Williamson

August 2022

A thesis submitted for the requirements of the MPhil Philosophy of the Australian National University

Except where it is otherwise acknowledged in the text, this thesis represents my own original work. No part of this thesis has previously been submitted for any degree, or is currently being submitted for another degree.

Signed:

Date:

15/08/2022

For Dugald and Rosemary Williamson,
who taught me how to write.

Acknowledgements

What must have been five years ago now, Katie Steele generously agreed to supervise me for an undergraduate reading course in normative ethics. I didn't know, at the time, that Katie would go on to be my supervisor from then until now. But I also didn't know Katie would prove to be the very best. Because of Katie, it feels, the whole world of philosophy has opened up for me. For that, and for all the many hours of chatting, I can't find enough words of thanks.

Many others have read many parts of this thesis too. Without them, I'd still be clueless. They include: Garrett Cullity, Brian Hedden, Christoph Lernpaß, Kirsten Mann, Will Moisis, Pamela Robinson, Nicholas Southwood, Hayden Wilkinson, and Timothy Luke Williamson. Brian and Nic's aid on my panel was invaluable, especially in pushing me to consider possible lines of reply to the argument I give in Chapter 2. I've also benefitted tremendously from discussions on these and related topics with Alan Hájek, Conor Leisky, Jonathan Tjandra, and Brandon Yip. Many of these conversations caused important thoughts to bubble to the surface. Thank you to all of you, and apologies to anyone I've missed. The ANU School of Philosophy is just about the best place there is to be a student – I miss it already, and haven't even left yet.

Thanks to Mum, to Dad, to Tim.

And to Beth.

Abstract

This thesis explores the significance of our *cluelessness* for the general project of moral philosophy. In the first chapter I continue a tradition which uses the facts of our cluelessness to argue against consequentialist accounts of right action. In the second chapter I develop a new cluelessness argument against recently popular relevance approaches to claims aggregation, approaches under which agents are required to maximise the strength-weighted satisfaction of relevant claims upon their conduct. In the third chapter I respond to the Paralysis Argument, a novel objection developed by Andreas Mogensen & William MacAskill which uses the facts of our cluelessness to undercut the traditional non-consequentialist distinction between reasons for doing versus allowing harm. In responding to the Paralysis Argument, I offer a refined version of the doctrine of doing and allowing harm, one which gives intuitively plausible verdicts in cases of risk and uncertainty. In the fourth chapter I examine whether we might sometimes interpret cluelessness arguments as *action-guidingness objections*: under action-guidingness objections, a particular moral principle is said to be incorrect insofar as that principle cannot be used by suitably motivated agents in regulating their conduct. I argue against the general merits of action-guidingness objections. I suggest that cluelessness arguments against consequentialism, for instance, can instead be given a more fruitful *epistemic* reading, a reading I defend in closing.

Table of Contents

Introduction

Mr Bean isn't the sort of person you'd trust to read a map the right way up in Grand Central Station – nor to boil coffee without burning his hands, nor to tie his shoelaces without tripping. Mr Bean, in sum, is clumsy. And that is why we might properly call him *clueless*.

In this thesis, I'll explore a fairly different sense in which we as moral decision makers might properly be called clueless. You and I are clueless in the following sense: it turns out that even the most seemingly trivial of our decisions are likely to exert a radical influence over the course of near and distant future history, a radical influence of which we are ignorant at the moment of choice. This thesis explores the significance of our cluelessness, in this second sense of the word, for the general project of moral philosophy.

That we are clueless in this second sense is not a recent insight. In *Reasons and Persons*, for instance, Derek Parfit emphasises that procreation is a highly fragile event: by inadvertantly influencing the exact moment in which others procreate by even the tiniest of margins, you alter the genetic composition of their future children.[1] But by altering the genetic composition of a future child, you change the immediate and not so immediate future in a myriad of other ways too – not least since such a change is likely to alter the identities and behaviours of many other future persons.

What *is* still in the early stages of debate, however, is the overall significance of our cluelessness for the general project of moral philosophy: in particular, whether our cluelessness might play some important role in arguments against particular moral

[1] Derek Parfit, *Reasons and Persons* (Oxford University Press: Oxford, 1984), 352.

principles or doctrines. Some, like James Lenman, have taken our cluelessness to play a distinctive role in arguments against consequentialist accounts of right action.[2] Others, like Andreas Mogensen & William MacAskill, have taken our cluelessness to play a distinctive role in arguments against *non*-consequentialist commitments, such as the traditional distinction between moral reasons for doing versus allowing harm.[3]

In this thesis, then, I examine whether the empirical facts of our cluelessness can be fruitfully put to use in arguing against particular moral principles or doctrines. Here is a brief sketch of its overall structure. In Chapter 1, drawing on a historical literature, I develop a cluelessness argument against objective and subjective consequentialist accounts of right action. I first draw attention to the fact that we regularly have conflicting evidence as to the longterm effects of our acts, where the correct resolution to this evidential conflict remains unclear. I then argue that, in light of this evidential conflict, subjective consequentialists cannot affirm certain deeply intuitive judgements of comparative betterness across the options. In particular, subjective consequentialists are faced with the following dilemma: either admit that the ideally (or nearly-ideally) rational ranking of our acts in terms of their expected value is typically epistemically inaccessible, or admit that 'anything goes' for the vast majority of decisions made within our moral lives.

In Chapter 2, I develop a new cluelessness argument against recently popular *relevance approaches* to claims aggregation, under which agents are expected to maximise the strength-weighted satisfaction of relevant claims upon their conduct. I first draw attention to the fact that even the most seemingly trivial of our decisions change which presently existing persons die from painful and premature death. I then suggest that, in light of this fact, those who believe in a relevance constraint on claims aggregation must nearly always deny that minor claims can bear on the deontic status of our options. This result, I suggest, is deeply implausible. I consider a range of possible replies to this argument, but suggest that none are entirely successful. One benefit

[2]James Lenman, "Consequentialism and Cluelessness," *Philosophy and Public Affairs*, 29, no.4 (2000):342–370.

[3]Andreas Mogensen & William MacAskill, "The Paralysis Argument," *Philosophers' Imprint*, 21, no.15 (2021):1-17.

of this discussion is that it allows us to explore, in new and interesting ways, how relevance approaches to claims aggregation might operate under situations of risk and uncertainty.

In Chapter 3, I examine Mogensen & MacAskill's Paralysis Argument, a novel argument against the traditional non-consequentialist distinction between reasons for doing versus allowing harm. The core of the Paralysis Argument goes as follows: if you really did have greater reasons against doing some harm x than against merely allowing some harm x, and if some standard framework for deontological decision making under risk were true, then you would have a preponderance of subjective moral reasons in favour of voluntarily entering a state of paralysis as opposed to ever doing anything at all. This is because voluntary paralysis risks only *allowing* various future harms; by contrast, active behaviour risks *doing* various future harms. I reject certain preliminary responses to the Paralysis Argument before advocating a possible alternative to the doctrine of doing versus allowing, one which gives intuitive verdicts in cases of risk and uncertainty, but which still undercuts the Paralysis Argument's central conclusion.

In Chapter 4, I conclude by considering a more general issue: whether cluelessness arguments against particular moral principles ought to be understood as *action-guidingness objections*. Under action-guidingness objections, particular moral principles are said to be incorrect insofar as they cannot be used by conscientious agents in regulating their conduct. I suggest that the answer to this question is *no*, and I argue against the general merits of action-guidingness objections. I instead suggest an epistemic interpretation of certain cluelessness arguments: cluelessness arguments against consequentialism, for instance, simply show that consequentialist accounts of right action force us to deny certain deeply intuitive claims of comparative betterness across the options that we take ourselves to already know.

A few central themes develop fairly naturally in the course of these four chapters. The first theme is that cluelessness arguments are probative: they do indeed give us reasons for rejecting particular moral principles or doctrines as incorrect. I take the

cluelessness arguments I develop, for instance, to give us at least some grounds for rejecting consequentialist accounts of right action and for rejecting standard articulations of the relevance approach. The second theme is that, regardless of whether or not one ultimately takes cluelessness arguments to be successful, such arguments almost always press us to consider how established moral principles or doctrines might be applied to situations of decision making involving radical risk and uncertainty. This is, I take it, the case in Chapter 3's discussion of the doctrine of doing versus allowing harm. And the third theme is closely related: often, it seems that the best way to *escape* cluelessness arguments is to undergo a subjective shift in our moral principles or doctrines. Thus we begin saying things of the following sort: an act's rightness hinges its effects *in expectation*, a claim is generated upon my conduct if I can *benefit someone's prospects*, and I have *pro tanto* reasons against imposing *additional risks of harm* upon others. Whether or not subjective shifts like these always and so easily resolve the woes of cluelessness is a recurring theme in the following pages.

We are, in the end, small creatures in a large world. What is far more surprising, though, is that each of carries the weight of future history in the palm of our hands and in the smallest of choices. This thesis attempts to make progress, in light of this stunning fact, in figuring out the exact nature and shape of our moral obligations.

Chapter 1

Cluelessness Redux

In this chapter, I defend a certain style of cluelessness argument against consequentialist accounts of right action. I begin by outlining an objection from cluelessness historically presented against objective consequentialism. I then suggest the following: we regularly have relevant evidence as to the sort of distant future consequences that are likely to follow our acts, where this evidence conflicts, and where the way in which this evidential conflict ought to be resolved remains unclear. In light of this fact, as I will argue, agents who endorse subjective consequentialism must either admit that the ideally rational betterness ordering across the options is usually inaccessible or admit that, for the vast majority of decisions made within our moral lives, anything goes. Both results, I suggest, are implausible.

1.1 Cluelessness and Birdwatching: An Introduction

1.1.1

Sunday afternoon, 4.P.M. I am out birdwatching in a patch of local wetlands and have just spotted an Endangered Plains-Wanderer (or perhaps it was only a common Button Quail) when I hear a terrible shout. Startled, I pack away my binoculars and my flask of tea and go to investigate.

A fellow birdwatcher is sinking in a nearby patch of mud. I must decide whether

or not to save his life. The risk to myself is negligible, since I need only throw in a large stick and he will be saved. By contrast, if I leave the birdwatcher alone he will certainly die a slow and painful death.

Should I save him or should I leave him? Fortunately enough, I always carry an ethics textbook in my backpack for situations just like this one. I flip to the section on objective consequentialism, and read that one act is better than another if it brings about the better total consequences.[1] This makes the analysis of my situation fairly simple. Since the positive consequences of saving the birdwatcher's life clearly outweigh the negative consequences of letting him drown, it is better to save him. Thus, I throw in the stick and go on my way.

1.1.2

Later that night, though, I begin to worry that something is wrong. I worry whether it really *was* better to save the imperiled birdwatcher. My reasoning goes as follows. Objective consequentialism says that comparative betterness across acts hinges straightforwardly upon their total consequences. But a moment's reflection reveals that, as I stood beside the mud pool listening to the birdwatcher's cries, I was in fact *clueless* about the total consequences associated with each available act. By objective consequentialism, then, I should have been clueless about which act was better than which.

Why think that, as I have suggested, I was clueless about the total consequences associated with each of my available acts? Simply put, it turns out that the birdwatcher's life and death was only a needle in the total haystack of consequences at stake in my decision. And in making my decision, although I took account of the needle, I failed to take account of the haystack.[2]

First note that my decision extended the birdwatcher's life by many years, allowing him to make various future decisions and to exhibit various future behaviours that he

[1] The target of this paper is act consequentialism.

[2] Here, I follow closely James Lenman's important paper, "Consequentialism and Cluelessness," *Philosophy and Public Affairs*, 29, no.4 (2000):342–370. Like Lenman, I motivate our cluelessness by drawing heavily on the identity-affecting nature of even the most seemingly trivial of acts.

would not have otherwise had the chance to make or exhibit. It is not unlikely that the birdwatcher, for instance, will now go on to have future children, each of which would not have existed if their father had perished in the pool of mud. My act will be a counterfactual difference maker as regards those various future consequences, good and bad, likely to obtain in the course of those future children's lives.

The problem is worse than this, though, since the saved birdwatcher (call him Darlington) can now go on to marry Sandra, who would have married Charlie if she had not married Darlington, and Charlie would have married Aubrey if she had not gone on to marry Alex ... and so on. Quite plausibly, by saving Darlington, I have changed the social structure of many dozens of families, and the subsequent identities of each of their children. These are the sorts of complex consequences that follow when someone lives a life they otherwise wouldn't have lived, and when someone goes on to marry someone they otherwise wouldn't have married.[3]

Already, it is not difficult to see that my act will change the identities of many dozens of persons even over the course of my own lifetime. My act will be a counterfactual difference maker as regards all those consequences (good and bad) liable to obtain in the course of these many dozens of lifetimes. That, however, is only the beginning. Most of the children I have inadvertantly brought into existence will go on to have children and grandchildren of their own. Each child, in other words, can be seen as the start of a family tree, or as a modification to an existing family tree. And each family tree will send consequences ricocheting through future history, multiplying as they go.

I was, in the end, quite literally *clueless* as to the total consequences of my decision to save Darlington rather than leave him be. By objective consequentialism, then, I should have been clueless about whether or not it was better to save him. Importantly, note that the problem does not only emerge when making weighty and pressing decisions like whether or not to save a stranger's life. The problem emerges for even the most seemingly trivial and innocuous of decisions, such as whether or not to bike to the shops or to have an extra slice of cake. Why? Well, here are two reasons. First, it turns out

[3]Lenman, 347.

that even the most seemingly trivial of our decisions influence the identities of future persons in virtue of influencing the exact moment or manner in which others procreate.[4] But second, it turns out that even the most seemingly trivial of our decisions serve as causal antecedents for far weightier decisions such as whether or not to save a drowning birdwatcher. Every causally antecedent act which caused me to save Darlington, in other words, *also* brought about the multitude of unforeseen consequences associated with his life being saved. These causally antecedent acts included my decision to take Sunday afternoon off, my decision to pursue birdwatching as a hobby, which was in turn sparked many years ago by my uncle's decision to buy me a pair of binoculars, which in turn was sparked by his holiday in Scotland as a child, and so on. Every single one of these causally antecedent acts changed the future course of history, hiding a haystack of unforeseen consequences.

1.1.3

There are different possible ways in which we might spell out this initial 'cluelessness concern' into a deductive argument against the truth of objective consequentialism. Here are two options. The first option would be to argue that objective consequentialism must be incorrect since it cannot be *used as a guide to action*. You might be tempted by this first option if you think, more generally, that it is an essential function of moral principles that those principles can be used by agents in fruitfully regulating their conduct. This first option is appealing because it does indeed seem that, given what we've said so far, objective consequentialism can't be used as a fruitful guide to action: agents typically cannot act out of a desire to conform to objective consequentialism, with any kind of justified belief that *really do* conform to the requirements of objective consequentialism.[5]

The second option would be to argue that objective consequentialism must be

[4]Lenman, 346; Hilary Greaves, "Cluelessness," *Proceedings of the Aristotelian Society*, 116 (2016):311-339; Derek Parfit, *Reasons and Persons* (Oxford University Press: Oxford, 1984), 352.

[5]This is broadly the sense of the word 'use' employed in Holly M. Smith, *Making Morality Work* (Oxford: Oxford University Press, 2018), 16, and also by R.M Hare in *Freedom and reason* (Oxford: Clarendon Press, 1963), 31-33.

incorrect since, if objective consequentialism *were* a true moral principle, we would not be able to know about comparative betterness across our acts in a way that we clearly *do* know about comparative betterness across our acts. If objective consequentialism were true, for instance, we would not be able to know that it is better to save the helpless birdwatcher rather than to leave him be. But clearly it *is* better to save the birdwatcher rather than to leave him be, and we can know as much at the moment of choice.

For reasons I explore in greater depth elsewhere, my preference is to argue in this second vein.[6] My preference, in other words, is to argue the following: there must be something wrong with objective consequentialism given that its truth would imply that we can never know about comparative betterness across our options in the way we regularly *do* know about comparative betterness across our options.[7]

1.1.4

In a straightforward deductive format, this second argument reads:

The Cluelessness Argument Against Objective Consequentialism

1. If objective consequentialism were true, then agents would never be capable of knowing, of any two distinct options ϕ and ψ available to them, whether ϕ is better than ψ or whether ψ is better than ϕ.

2. Agents do know, of at least some distinct options ϕ and ψ available to them, whether ϕ is better than ψ or whether ψ is better than ϕ.

C. Objective consequentialism is not true.

[6] See "A Guide to Action-Guidingness Objections," where I evaluate both styles of argument in depth. Past discussion of our cluelessness has touched on both concerns. N.B. Greaves, 312; Shelly Kagan, *Normative Ethics* (Boulder: Westview Press, 1998), 64; Lenman, 350. For an alternative reading of the cluelessness argument as a violation of the ought-implies-can constraint, see Frances Howard-Snyder, "The Rejection of Objective Consequentialism," *Utilitas*, 9, no.2 (1997):241–248.

[7] Now, you might think that there is a difference being between clueless about comparative betterness across our options (i.e., about whether it is better to save the birdwatcher), versus being clueless about our deontic obligations (i.e., about whether I *ought* to save the birdwatcher). I simply assume here, however, that the objective consequentialist straightforwardly determines our deontic obligations as a function of comparative betterness across the options. If the objective consequentialist is clueless about comparative betterness, in other words, it follows that they are clueless about our deontic obligations.

This argument is framed in terms of our ability to *know* about the comparative status of our acts. But note that not too much hangs on the use of this word. If it was your preference, I take it that you could also run the argument in terms of credences: surely us moral decision makers are capable of having certain especially high levels of rational credence that some particular available act ϕ is indeed better than the alternatives, but this possibility would seem to be precluded by the truth of objective consequentialism.

Now, we have already spelled out the reasoning behind Premise *1* in some depth: in brief, if objective consequentialism were true, then we could not have a clue about comparative betterness between any of the options since (a) under objective consequentialism, comparative betterness across the acts hinges solely on total actual consequences, but since (b) we are clueless as to the total actual consequences of the options.

In this paper I'll take Premise *2* to be fairly uncontroversial, and so shall say little about it. I'll simply assume that we at least sometimes *do* know about comparative betterness across our options as we go through life making moral decisions of various kinds – deciding, for instance, whether it is better to rescue or abandon a drowning stranger at no cost to oneself, to crash a bus into a crowded mall versus press the brakes, or to feed a starving puppy rather than menace and torture it.[8]

In this general form the argument complains that, given the truth of objective consequentialism, agents could *never* know about comparative betterness between any two acts. This language may seem unnecessarily strong. For instance, assume objective consequentialism were true. Then consider a possible world with a single (lonely) inhabitant whose actions affect none other than herself. Such an agent, it seems, might still have a firm grasp on the total morally significant consequences associated with her acts. Such an agent might sometimes come to know, it seems, which of her acts are better than which. I take it, then, that the 'never' featuring in the argument is a never of a fairly qualified sort. The claim is only that agents like you and I, who find ourselves in the middle of a causal history, and whose actions are associated with

[8]I explore this issue, too, in greater depth in "A Guide to Action-Guidingness Objections", especially in Section 3 of that paper. One relevant issue is whether the objective consequentialist can still affirm certain hypothetical or counterfactual judgements of comparative betterness between options, and whether such hypothetical or counterfactual judgements are, in the end, enough.

"massive causal ramifications", are never capable of knowing which of our options are better than which.[9] I take it that this qualification doesn't make the argument any less persuasive: the concern, after all, was whether agents like you and me can know about comparative betterness across the options in the way we intuitively do.

1.2 The Defence from Subjective Consequentialism

1.2.1

The cluelessness argument sketched a moment ago complained that objective consequentialism could not facilitate a common-sense knowledge of comparative betterness across the options. Subjective consequentialists, however, use *expected moral value* to determine comparative betterness between acts. And since subjective consequentialists are not clueless about the expected moral value associated with their available acts, subjective consequentialists need not be clueless about comparative betterness between their available acts. Thus goes the standard subjective consequentialist defence against the woes of cluelessness.[10]

As an example, take again the choice of whether or not to save the drowning birdwatcher. My two options are *Leave* and *Save*. Each act is associated with an expected value *EV(Leave)* and *EV(Save)*. In determining which course of action is better than which, the subjective consequentialist need only know which act has the greater expected value: whether *EV(Leave)* or *EV(Save)* is the larger.

One way of proceeding then goes as follows. Expected value is simply the probability-weighted sum of the possible values in the possible outcomes associated with each individual act, given the epistemic state (perhaps: the ideally rational epistemic state) of the decision maker at hand. If we restrict our attention to only the 'foreseeable' consequences at stake – those immediate consequences for which I have

[9] Lenman, 347.

[10] Greaves, 317. See also Elinor Mason, "Consequentialism and the Principle of Indifference," *Utilitas*, 16, no.3 (2004):316–321. Mogensen terms this the *Naive Response* in Andreas Mogensen, "Maximal Cluelessness," *The Philosophical Quarterly* 71, no.1 (2021):141–162.

relevant evidence, such as Darlington's impending doom – then *Save* is clearly associated with the greater expected value. After all, *Save* brings about a saved life, whereas *Leave* brings about a painful death. These outcomes, given the case as described, are sure-things.

Importantly, though, *Save* will have the greater expected value *simpliciter* if it then turns out that 'unforeseeable' consequences are incapable of influencing the ranking of our available acts in terms of their expected value. (Such unforeseeable consequences, as we saw in the previous section, include those consequences that might follow my act in the haystack of future history, but for which I have no relevant evidence either way at the moment of choice.) Put simply, *Save* will retain the lead in terms of expected value if Greaves' EVF thesis holds:

> **EVF.** The expected value of an action is determined entirely via its foreseeable effects.[11]

If EVF is true, then *EV(Save)* outranks *EV(Leave) simpliciter*, since it outranks *EV(Leave)* with respect to the foreseeable, and since that is all that matters. The subjective consequentialist can thus conclude that *Save* is the better act.

1.2.2

Luckily for the subjective consequentialist, there is a strong argument in favour of EVF.[12] The argument relies upon the principle of indifference, and upon certain assumptions regarding our evidential situation when it comes to the distant future consequences of our acts. For any two acts A_1 and A_2, and for any two unforeseeable effects E_1 and E_2, and by the very definition of 'unforeseeable,' we have no evidence as to whether $(A_1 \mathbin{\square\!\!\rightarrow} E_1 \ \& \ A_2 \mathbin{\square\!\!\rightarrow} E_2)$ or whether $(A_1 \mathbin{\square\!\!\rightarrow} E_2 \ \& \ A_2 \mathbin{\square\!\!\rightarrow} E_1)$. The principle of indifference therefore tells rational agents to assign equal credence to these to possibilities. In other words, the principle of indifference rationally entails that $\mathrm{Cr}(A_1 \mathbin{\square\!\!\rightarrow} E_1 \ \& \ A_2 \mathbin{\square\!\!\rightarrow} E_2) = \mathrm{Cr}(A_1 \mathbin{\square\!\!\rightarrow} E_2 \ \& \ A_2 \mathbin{\square\!\!\rightarrow} E_1)$.

[11] Greaves, 318

[12] See Greaves *ibid.* for an extended statement of this argument.

But if this is the case, then the unforeseeable consequences E_1 and E_2 cannot sway the comparative expected value ranking of either A_1 or A_2. For either unforeseeable consequence is just as likely to tether itself to either available act. Perhaps E_1 would be a great or terrible thing, but I cannot use this fact to change the comparative expected value of A_1: for every degree to which E_1 might change the EV(A_1), there is an equal and counterpart degree to which E_1 might change EV(A_2). Perhaps, too, E_2 would be a great or terrible thing, but I cannot use this fact to change the comparative expected value of A_2: for every degree to which E_2 might change the EV(A_2), there is an equal and counterpart degree to which E_2 might change EV(A_1). In general, then, we see that given the principle of indifference, and given certain assumptions about our evidential situation as regards the "unforeseeable", the unforeseeable cannot change the expected value ranking of any given act. In this article I do not dispute the principle of indifference; I simply take it as a given that, if one has no relevant evidence bearing on the truth or falsity of each member of a set of mutually exclusive and exhaustive propositions, one ought assign equal credence to them all.[13]

1.3 Weakly-Foreseeable Consequences

1.3.1

My aim in the following sections is to show that the defence from subjective consequentialism does not quite succeed. Subjective consequentialists, too, are beset by the woes of cluelessness. My starting point is the implicit distinction between foreseeable consequences and unforeseeable consequences themselves.

Our actions have consequences. This much everyone can agree to. So far we have been carving up these consequences in rather a crude way, namely, into the categories 'foreseeable' and 'unforeseeable.' This distinction was necessary in articulating the standard subjective consequentialist defence against cluelessness.

[13]This is to put aside the problem of partitioning, discussed in Greaves, 319-322. I put aside partitioning issues since they are, as it were, on my side: they threaten the standard subjective consequentialist response to cluelessness.

We have taken foreseeable consequences, roughly speaking, to be those consequences about which we can have, or can reasonably hope to gain, *relevant evidence*. In the case of Darlington the birdwatcher, we took the key foreseeable consequence to be a saved life. By contrast we have taken unforeseeable consequences, roughly speaking, to be those consequences about which we do not have, and cannot reasonably hope to gain access to, relevant evidence. In the case of Darlington the birdwatcher, an unforeseeable consequence might conceivably consist of a freak tornado coming about several thousand years from now as a result of my action disturbing the air currents in *just* the wrong way.

We have implicitly accepted this distinction all along, but it is problematic for the following reason. The simple fact is, we often have conflicting evidence that certain sorts of consequences will follow our acts in the long run, where the way in which this evidential conflict ought to be resolved remains unclear. Consequences like these fit comfortably into neither of the two previous categories – they are neither 'foreseeable' nor 'unforeseeable' in the ordinary sense of either word. Perhaps we could call such consequences the *weakly-foreseeable*.[14]

Now, much more needs to be said here. After all, in trying to discern the truth or likelihood of particular propositions, we almost *always* have to trade-off between conflicting pieces of evidence. The weather forecaster predicted rain but clouds are nowhere to be seen; my friend says she is good at darts but has already missed the first three shots; the painting looks like that of a pipe but the inscription says otherwise; and so on. We normally assume that, when faced with conflicting evidence, agents are capable of making the sensible evidential trade-offs and evaluations that an agent ought to be able to make before ultimately coming to a sensible and considered level of credence in the relevant proposition. (It will probably not rain; my friend is almost certainly bad at darts; the painting is definitely that of a pipe rather than that of a

[14]For a discussion of this issue of conflicting evidence, see Greaves, 323 – my discussion in the following section is indebted to some of the important distinctions drawn in that paper. To emphasise, this term, *weakly-foreseeable*, does not reference the level of credence one has that a particular consequence will obtain: it instead references the underlying fact that in trying to adopt a particular credence *vis-à-vis* such consequences, one is faced with relevant but conflicting evidence.

banana; and so on.)

In my mind, though, there is something especially significant about the category of 'weakly foreseeable consequences.' These are consequences for which we not only have conflicting evidence, but for which the *way* in which this evidential conflict ought to be resolved remains unclear. The way in which the evidential conflict ought to be resolved remains unclear because these are scenarios in which there exists no past data on how the evidential trade-offs might reliably be made. These are scenarios, rather, in which we hypothesise about novel and one-off events, about novel social trends, or about the extremely long-run effects of our actions in the distant future. In none of these cases is there past data on how the evidential trade-offs might reliably be made, nor are there mechanisms for checking whether we have given each piece of evidence the weight it deserves to have in our deliberative processes.[15]

1.3.2

We cannot get much further without turning to consider particular examples: cases in which our acts are associated with 'weakly-foreseeable' consequences for which we have conflicting evidence, where the correct resolution to this evidential conflict remains unclear at the moment of choice. Having shown that such cases are pervasive, it will be a short step towards vindicating a certain variety of the cluelessness argument against subjective consequentialism.

1. First, consider an example provided by Greaves in her discussion of *complex cluelessness*.[16] Suppose I am deciding whether or not to donate to an effective charity, *Aid*. By giving to *Aid* I save lives and yield an immediate future population that is larger than it otherwise would have been. By not giving to *Aid* I allow others to die and hence yield an immediate future population that is smaller than it otherwise would have been.

[15]You might think there are other tools (besides past data) that we can use as an aid in making evidential trade-offs: for instance, we might draw on analogous cases, draw on intuition, or draw on some kind of *a priori* reasoning. But such tools are not obviously of help in the sorts of cases I consider shortly. Perhaps intuition is, but the question is then whether intuition is a reliable or robust guide to the resolution of evidential conflicts in moral decision making.

[16]Greaves, 323.

I have conflicting evidence, in such a case, that each of my options will lead to an especially favourable *future social context* – a future social context, that is, which is especially conducive to the flourishing of future generations. This is a weakly-foreseeable consequence associated with each of my acts.

Why? Well, there are reasons for thinking that an increase in population size would be systematically better, instrumentally speaking, for the flourishing of distant future generations. After all, more persons plausibly means a greater rate of technological, scientific, and medical advancement; the production of more beautiful works of art and music than otherwise would have existed; the opportunity for more meaningful relationships, commitments, and shared projects, and access to more collective social resources and goods; in other words, an increase to population size means that more will have access to more of the sorts of things that often make life worth living. But there are, however, conflicting reasons for thinking that a *decrease* in population size would be systematically better, instrumentally speaking, for the welfare of future generations, particularly distant future ones. The suggestion here is that decreases to the present population size mitigate against long-run risks of overpopulation and resource depletion. As Greaves puts it:

> ... Assuming for the sake of argument that the net effect of averting child deaths is to increase population size, the arguments concerning whether this is a positive, neutral or negative thing are complex. But, callous as it may sound, the hypothesis that (overpopulation is a sufficiently real and serious problem that) the knock-on effects of averting child deaths are negative and larger in magnitude than the direct (positive) effects cannot be entirely discounted.[17]

It turns out, then, that whichever act I perform, I have complex and conflicting evidence for thinking that my act will be particularly and especially favourable as regards the distant future social context. And it is not obvious that there exists, as I mentioned previously, any clear way forward in resolving this evidential conflict. Plausibly, there

[17] Greaves, 325.

exists no clear way forward in resolving this evidential conflict since this is not the sort of case in which we can use past data to act as a guide as we attempt to weigh the evidence as it ought to be weighed. We are talking, rather, about the novel effects our actions will have on the shape of the distant future.[18]

2. A structurally analogous problem arises for cases *in general* in which we have some firm reason for thinking that our act will influence the immediate population size in a certain way – either increasing or decreasing it in comparison to what it would have otherwise been. The problem, in other words, is not limited to situations of effective giving. Think, here, of decisions about whether to procreate, to kill, to rescue, to vote for one party rather than another, to plan cities one way rather than another, to wage war, to build hospitals, and so on, so long as in each case one has concrete reasons for thinking that the relevant act will have some bearing upon the proximate population size.

In each of these ubiquitous population-size affecting cases, just as before, a weakly foreseeable consequence associated with each act is that it will bring about an especially favourable future social context. One has conflicting evidence for thinking that each available option will bring about a social context which is particularly conducive to the overall flourishing of future generations. And just as before, given the novel nature of the case, the correct resolution to this evidential conflict remains unclear.

3. These initial examples all gained traction because the options in question had some measurable influence over the proximate population size. But this need not be the case. Consider the following and rather different sort of case in which weakly-foreseeable consequences are present: *pathway-style cases.*

In pathway-style cases, just as before, one has conflicting evidence for thinking that each available option will lead to an especially favourable future social context. This is a weakly-foreseeable consequence associated with each act. But the reason,

[18]In Greaves' language, in defining cases of complex cluelessness, we say that each act leads to systematically favourable unforeseen consequences (see for instance Greaves, 323). But this is a difficult phrase since, if a consequence is genuinely unforeseen or unforeseeable, it is unclear how we can know such firm facts about it – for instance, that it is systematically favourable. Instead, I think we simply ought to say that each act is associated with a weakly-foreseeable consequence: we have conflicting evidence that each act will lead to a favourable future social context.

here, has nothing to do with population size. In pathway-style cases, rather, one must decide in which way to make the world a better place, where making gains along one pathway of social improvement demands sacrifice along another. One has reasons, in such cases, for thinking that the pursuit of *each* pathway of social improvement will lead to particularly favourable long-run social conditions.

A concrete example will help. Suppose that I am a philanthropist. I must choose between funding climate change mitigation in order to better the welfare of future generations versus funding basic educational resources for those who desperately need them now. I cannot fund both causes, since my resources are limited.

I have conflicting evidence, in a pathway-style case like this, for thinking that each available act will bring about an especially favourable future social context. The first act is especially preferable insofar as it directly benefits future persons, mitigating the impact of natural disasters, pollution, environmental degradation, ecological collapse, and so on. I also have reason for thinking that this benefit is of primary importance, given the sheer number of future persons likely to benefit from my act. But the benefits of the second act are extremely difficult to quantify, and when viewed in a certain light, seem even better than those of the first. In choosing the second act, not only do I directly benefit present persons by improving their quality of life, but in some important sense I seem to *indirectly* benefit future persons as well. After all, this second act encourages present generations to make wiser and more informed choices which may well in turn improve the future social and environmental context. I can't have a clear grasp of just how likely these indirect benefits are to come about, given that we are talking in a fairly speculative manner about the exact way in which long run benefits might filter down across many generations, but the possibility of such indirect benefits coming about is a serious hypothesis that I need to include in my deliberations going forward.

There are, then, conflicting pieces of evidence for thinking that each pathway will bring about the most favourable future social context in the long-run. But the way in which I ought to weigh these conflicting pieces of evidence remains unclear. It remains

unclear because the evidence is not the sort of thing that permits of an easy resolution: it draws on speculative (though serious) arguments about the ways in which distant future persons are likely to benefit from the performance of my act either way.

In the philanthropic example just given, the two possible 'pathways' of social improvement between which we had to choose were direct climate change mitigation versus the funding of educational resources for those who need them now. I do not mean to imply, though, that these are the *only* two dimensions we could have drawn upon in constructing the case. In fact, it strikes me that such pathway cases are utterly ubiquitous. Pick any two pathways of possible social improvement – investment in the arts, investment into scientific and technological advancement, investment into the testing of economic theories, investment into research on the most effective forms of governance – and it is difficult to discern, at the best of times, which strategy or combination of strategies will bring about the most favourable social context in the long run. There usually exists conflicting and unresolved evidence for thinking that, in the long run, *each* pathway of social improvement will prove especially favourable.

Pathway-style cases, even more so than population-size affecting cases, are ubiquitous in the course of our daily moral lives – we face them knowingly or knowingly whenever we make a decision about how to spend our money, where to work, whether or not to have a family and with whom, which social causes to pursue, which social causes not to pursue, which candidates and policies deserve our vote, and so on. This is because, in all such cases, we tend to directly or indirectly promote some avenues of social improvement at the cost of others. In all such cases, the thought goes, we have conflicting evidence as to the long-term favourability of our options.[19]

Of course, someone might at this point emphasise that we *do* regularly trade-off the evidence in favour of different pathways of possible social improvement – we weigh the evidence concerning the favourability, say, of funding climate change mitigation

[19]These 'pathway' cases are closely related to Greaves' complex cases of cluelessness; perhaps they simply *are* further examples of complex cluelessness since, as before, these are cases in which we have reasons for thinking that each available option is going to prove particularly conducive to human flourishing in the long-run. They key point, in any case, is that cases with a pathway-style structure are ubiquitous, and that the problem is not limited to those cases involving changes to the population size.

versus funding educational programs. We subsequently pursue one strategy of social improvement in lieu of the other. Which pathway is worth pursuing will depend on the precise empirical details of the case at hand. However, the fact that we do regularly come to a choice in pathway cases is of less significance than we might think. The question is instead whether we regularly do make the *rational and all-things-considered* choice in pathway cases, taking into account not only the immediate consequences of our acts for which we have evidence, but also taking into account the distant-future and weakly-foreseeable consequences for which we have conflicting evidence, where the resolution of this evidential conflict remains unclear. *This* is the issue at stake: and this is a question, I suspect, with which we grapple comparatively rarely.[20]

1.4 The Precise Credences Horn of the Dilemma

1.4.1

Having shown that even the most mundane of our acts are regularly associated with 'weakly-foreseeable consequences' – in particular, our acts are associated with weakly-foreseeable consequences whenever we influence population size, whenever we trade-off between possible dimensions of social improvement, and whenever we make decisions which will in turn foreseeably influence our decisions in such cases – we are now well-placed to vindicate a variety of the cluelessness argument against subjective consequentialism.[21]

Now, recall that subjective consequentialists maintain that the better act is the

[20]Indeed, on this point, climate change is an especially pertinent example. As emphasised by John Broome, "Should We Value Population," *The Journal of Political Philosophy*, 13, no.4 (2005):399–413, discussions in climate policy typically bracket off the possible effects of our actions beyond a time horizon of, say, the next century. This does not imply, though, that in pursuing the particular climate policies we do, we have taken account of all those consequences relevant to the decision-making process and about which we have relevant evidence.

[21]One might question whether I really have shown that 'even the most mundane of our acts are regularly associated with weakly-foreseeable consequences.' This is a strong claim. But against this concern, note that our decisions are associated with weakly-foreseeable consequences whenever we have some concrete reason for thinking they will *influence* our behaviour in population-size and pathway-style cases. And my suspicion is that trivial and mundane acts hold at least some influence over the development of our characters and dispositions in one way rather than another, and thus over our choices in, say, pathway-style cases.

one associated with the greater expected value, where expected value is simply the probability-weighted sum of the possible values in the possible outcomes that might follow the performance of my act. It seems that, when contemplating acts associated with weakly-foreseeable consequences, subjective consequentialists have at least two options. Neither is friendly.

The first option is to bunker down and insist that the ideally rational agent still ought to assign a single and precise set of probabilities to the possible outcomes associated with each of their available acts, given the evidence at hand.[22] It follows, on this first option, that each available act can still be represented in terms of expected value. It follows, too, that we can use expected value to rank acts in terms of comparative betterness. On this first option it is neither here nor there that, as I have been labouring, there often exists complex and conflicting evidence as to whether certain consequences will follow our acts in the long-term, where the correct resolution to this evidential conflict remains unclear.

The problem with this first option, though, is as follows. As I have argued, population-size affecting choices and pathway-style choices are ubiquitous in the course of our moral lives. It is hard to think of a decision scenario which *doesn't* directly or indirectly fall under one of these two banners. But it is plausible to think that in all such cases involving weakly-foreseeable consequences, the ideally (or nearly ideally) rational ranking of my acts in terms of their expected value will prove epistemically inaccessible to me as a moral decision maker. Why? Well, it is plausible to think that in all such cases the ideally (or nearly ideally) rational probabilities that ought to be assigned to the possible outcomes in light of my evidence will *similarly* be epistemically inaccessible to me as a moral decision maker.

Of course, this answer only pushes the buck back further. We are in want of a reason for thinking that the ideally or nearly-ideally rational probabilities that ought to be assigned to the possible outcomes, given my available evidence, will be epistemically inaccessible. The particular reason I have in mind goes as follows: giving a proper

[22] See Greaves, 327.

evaluation of the complex and conflicting evidence involved in cases involving weakly-foreseeable consequences is not impossible, but it is a Herculean task, demanding of tremendous time and tremendous epistemic resources. This is a task, we might think, which is typically beyond the constraints of those decision making scenarios in which we find ourselves.[23]

It will be most useful, here, to motivate the point with a particular example. Take my decision as regards the drowning birdwatcher. Now, I have complex and conflicting evidence for thinking that an increase to population size will bring about a particularly favourable future social context; I also have complex and conflicting evidence for thinking that a decrease to population size will bring about a particularly favourable future social context. Perhaps the tension between these conflicting pieces of relevant evidence really can be resolved, and perhaps these conflicting pieces of evidence really do specify an ideally rational and precise probability distribution over the possible outcomes associated with each act. But crucially, to resolve this evidential conflict and arrive at the rational probability distribution mandated by my evidence would seem to be a Herculean task. One would have to synthesise and appraise arguments, hypotheses, and theories on the long-term effects of population growth on social outcomes: on the way in which population size influences resource depletion, technological development, environmental degradation, economic and governmental stability, and so on, to name a few. One would have to weigh each piece of evidence as it deserves to be weighed. Perhaps, here, one could draw on external expertise as an aid, but not unless one themselves already had the tools and abilities needed in finding and appraising external sources of expertise for oneself. And as I say, even if one *could* manage the Herculean task of resolving the evidential conflict as it deserves to be resolved, and therefore assigning probabilities to the possible outcomes as they ought to be assigned, there is the further question of whether individual agents can undergo an

[23]You might worry that rational epistemic probabilities for you *have* to be epistemically accessible. But I simply assume, here, that a claim of the following form is sensical: there may be a fact of the matter as to the ideally (nearly ideally) rational probabilities you ought to assign to the outcomes given your evidence, even if you are typically unable to arrive at those ideally rational probabilities within the constraints of the decision making scenarios with which you typically find yourself.

even remotely adequate evidential evaluation of this sort within the decision-theoretic constraints with which they typically operate.

An immediate response might emphasise that subjective consequentialism only requires agents to maximise expected value *as best they can* – to maximise expected value, for instance, using the most rational probabilities assignments agents are capable of allocating to the possible outcomes in light of their available evidence and in light of the given time constraints. If this response is right, though, then it seems we have little guarantee that agents like you and me who attempt to maximise expected moral value 'as best we can' are doing so in a manner that approximates that of our ideally rational counterparts. We have no guarantee, for instance, that the probability assignments we endorse are anywhere close to those that would be endorsed by our ideally rational counterparts who possessed the same set of evidence as us. If it is unclear whether agents like you and me maximise expected value in a manner which even broadly approximates the manner in which our ideally rational counterparts would maximise expected value, it is less clear why we should place such great weight on expected value maximisation as being the locus of right action in the first place.

1.4.2

At this point a very familiar subjective consequentialist line of response beckons. The response goes that, often, the expected value maximising thing is to *refuse* to calculate expected value. After all, spending our whole lives contemplating a single act can lead to terrible outcomes. As an alternative, one should simply choose an act based on norms or rules or instinct, where these more basic and generalised decision procedures themselves are directed towards the maximisation of the good. This is, as it were, to maximise expected value the short way round.[24]

This point seems to dissolve the problem I have only just outlined. It no longer matters that there exists complex and conflicting evidence as to the long-term con-

[24]In particular see Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics*, 101, vol.3 (1991):461–482. As Jackson, 469, puts it, "ducking, swerving, smiling, playing a drop shot, and the like are commonly best done straight off as the spirit moves one and without further ado."

sequences of our acts – that it would take a lifetime to evaluate and weigh off such evidence, coming to the most appropriate probability distributions demanded by that evidence. Given the difficulties involved in synthesising such complex and conflicting evidence, one should simply act based on norms or rules or instincts which are *themselves* directed towards the maximisation of good consequences. In the case of Darlington the birdwatcher, one should simply pick up the stick and throw it in. Such an act is simply grounded in some norm or rule of rescue, a norm or rule which is itself directed towards the promotion of good consequences in the long-run.

But note that this style of response is uniquely circular when employed in this context. For the question at stake is *precisely this question*, of which actions, and which patterns of behaviour, will maximise expected value in the extremely long run given that we have conflicting evidence about the sorts of longterm consequences liable to follow our acts. To insist that one ought to give up on expected value calculations when it is infeasible to weigh the complex and conflicting evidence *vis-à-vis* the longterm consequences – and thereby maximise expected value the short way round – is to beg the question. We are in want of an argument for the claim that certain rules, norms, and action-guiding instincts will themselves maximise expected value in the long-run, given that our evidence currently conflicts in complex ways. That is the issue at stake.

Of course, some collective courses of behaviour have treated us favourably thus far. But it would be inductively dubious to insist that therefore, without having weighed the complex and competing evidence either way, those same courses of action will treat us favourably indefinitely into the future.[25] That precludes the possibility, for instance, that we find ourselves at a tipping point in history, where, in order to promote the welfare of future persons, drastic changes and revisions are demanded of our present behaviour. For consequentialists who take such possibilities seriously, it is not enough to simply defer to established rules, norms, or instincts when the evidential going gets tough.

[25]For an importantly related discussion, see Joanna Burch-Brown, "Clues for Consequentialists," *Utilitas*, 26, vol.1 (2014):105–119.

1.4.3

I previously said that, when confronted with the fact that our acts are regularly associated with weakly-foreseeable consequences, the subjective consequentialist has two options. So far we have only dwelt on the first option: the subjective consequentialist might maintain that *even though* there exists complex and conflicting evidence as to the long-term consequences associated with our acts, this evidence still specifies a precise and ideally rational set of probabilities over the set of outcomes associated with each act.[26]

I have claimed that even if the complex and conflicting evidence *does* specify a precise and uniquely rational set of probabilities over the outcomes, we should usually expect these ideally rational probabilities to be inaccessible to us, given that we will usually lack the epistemic resources required in a proper evaluation and appraisal of this complex and conflicting evidence. We should therefore expect the ideally rational betterness ordering to be inaccessible.

I take this result to be a vindication of the original concern from cluelessness. The subjective consequentialist is left admitting that the rational betterness ordering mandated by their available evidence is usually epistemically inaccessible. And this means that subjective consequentialists will typically be unable to affirm certain commonsense judgements of comparative betterness across the options – the judgement, for instance, that it is better to save a stranger from painful and premature death at no cost to oneself. At least, the subjective consequentialist will be unable to justify such judgements on the grounds that the rational ranking of acts in terms of their expected value mandates those judgements. This should leave us admitting that the concern from cluelessness has scored a serious point against subjective consequentialism.

[26] I am assuming here that, whatever else is true, one's evidence determines a *unique* spread of probabilities over the outcomes. I have not considered, here, what we might say about subjective consequentialism if some kind of epistemic permissivism were true: a topic as large as this, however, would be deserving of a paper in its own right.

1.5 The Imprecise Credences Horn of the Dilemma

1.5.1

This leaves the consequentialist with the second option. On the second option, one emphasises not only that the evidence conflicts in complex ways, but also that precisely *because* the evidence conflicts in complex ways, we may not even be warranted in adopting precise credences over the possible outcomes associated with our available acts. Instead, our credences over the outcomes ought to be imprecise.

Our credences ought to be imprecise, the more general thought goes, when our evidence as to the likelihood of outcomes conflicts, when we have not yet properly evaluated the evidence due to the great difficulties and demands in doing so, and when our uncertainty is so severe that we cannot justify one possible probability distribution over others *even given* our evidential base.[27]

Now, if our credences over the possible outcomes are imprecise then our acts can no longer be represented as expected value lotteries over the possible states of nature, where one act is better than another if and only if is associated with the greater expected value. After all, in order to represent acts as expected value lotteries in the first place, we needed precise probabilities by which to weight the possible outcomes. Instead, then, we are left in want of an alternate decision rule that will give us, at the very least, criteria of moral permissibility given imprecise credences.[28] Such a decision rule would tell us which acts are morally permissible even when there is no fact of the matter, in precise terms, about how likely we should think the possible outcomes associated with each available act.

My aim in this section is not to exhaustively search for the most plausible decision rule that might govern rational choice given imprecise credences. My aim is only to consider one plausible and widely-discussed decision rule, the maximality rule,

[27] Richard Bradley, *Decision Theory with a Human Face* (Cambridge: Cambridge University Press, 2017), 225-226. Also James M. Joyce,"A Defense of Imprecise Credences in Inference and Decision Making," *Philosophical Perspectives*, 24 (2010):281–323.

[28] Greaves p.328.

and show that, if subjective consequentialists wish to fall back upon the maximality rule, they will inadvertantly vindicate a particular kind of cluelessness worry against consequentialism. What is this worry, in exact terms? The upshot of the discussion will go as follows: for the vast majority of moral decisions made within our moral lives, the maximality rule will end up telling us that anything goes. Such a result is deeply implausible, and serves as a vindication of the original concern from cluelessness.[29]

1.5.2

To begin with, a definition of the maximality rule. Suppose that, rather than endorsing a single and precise distribution of probabilities across the possible outcomes, I entertain a range of probability distributions over the possible outcomes: call the set of the entertained probability distributions my *representor, R*. To entertain multiple alternative probability distributions across the outcomes just is what it is for me to be in a state of imprecise credence over the possible outcomes.

The maximality rule then says the following: one act A is only preferred to another act B in the particular case that A has greater expected value than B as calculated with respect to every probability distribution entertained within the representor. If this criterion is not met, then both A and B are permissible. Neither act is preferable to the other.

Now, one upshot of the maximality rule, as Mogensen rightly notes in the particular context of effective giving, is that more acts will be permissible than we may have initially thought.[30] Giving to a seemingly ineffective charity will be permissible so long as the seemingly ineffective act is associated with greater expected value under certain probability distributions contained within the representor. And it is plausible to think that the ineffective charity *will* be associated with the greater expected value under certain probability distributions contained within the representor, given the severe depths of our uncertainty about the effects of effective giving upon distant

[29]My discussion in this section draws on Mogensen *ibid*, who similarly provides an analysis of the maximality rule in the context of moral decision making. As will become apparent shortly, my discussion differs from Mogensen's in certain key respects.

[30]Mogensen, 154.

future – given the severe depths of our uncertainty, for instance, as regards whether saving many or fewer lives now will improve the distant future social context.[31]

For those who are invested in effective giving, then, there is something unsettling about the maximality rule. The maximality rule cannot vindicate certain orthodox strategies employed within effective giving movements. But we can here, however, press a more general concern. The more general concern is that it seems subjective consequentialists who fall back upon the maximality rule will rarely *ever* be able to say, of a given act, that it is impermissible. We can call this the problem of *blanket permissibility*. Given maximality, the concern goes, it will turn out that for the vast majority of decisions made within the course of our moral lives, anything goes.

The reasoning goes as follows. As I have emphasised, it is not only in cases of effective giving in which we have profoundly conflicting evidence as to the long-term favourability of our actions. We have profoundly conflicting evidence as to the long-term favourability of our actions whenever we exert an influence over the size of the immediate population, whenever we choose between competing pathways of possible social improvement, and whenever we make choices liable to *influence*, in some measurable way, our behaviour in population-size cases and pathway-style cases.

When applied to this extremely wide range of cases, the maximality rule says the following: some option A is better than some other option B only so long as that option A turns out to have the greater expected value under every probability distribution contained within the representor. But this, however, is exactly the sort of result we should expect *won't* hold when applied to this extremely wide range of cases. We should rather expect that, for any two options A and B, A will come out as having the greater expected value under some probability distributions and B will come out as having the greater expected value under others. This is for the exact same reason that Mogensen highlights in cases of effective giving: these are cases in which there exists a deep and profound uncertainty as to the longterm favourability of each available act – as to the longterm favourability, say, of pursuing one pathway of social improvement at

[31] *ibid*.

the expense of another, or of engaging in a course of behaviour which will measurably *influence* our choice in pathway-style cases.

We may ultimately have to conclude, then, that anything goes in this extremely wide range of cases, just as we had to conclude that anything goes in the case of effective giving. We won't be able to say definitively, in the evaluation of any such cases, that any particular option A is to be preferred to any particular option B.

There is, I suspect, something deeply unsettling about this result. To emphasise as much, let's dwell on some particular examples. Suppose that as I stand beside the mud-pool, wondering whether or not to save the drowning birdwatcher or leave him be, I am in a state of imprecise credence over the long-term consequences associated with my available acts. I am in a state of imprecise credence due to the fact that I have complex and conflicting evidence as to the longterm favourability of either of my options. (After all, the case is a population-size affecting one. In fact, suppose there exist several hundred helpless young birdwatchers drowning horrifically as a result of their raft sinking, and that I can save them all with the slightest push of a stick.) *Ex hypothesi*, suppose that I also endorse a kind of maximality consequentialism: I maintain that in cases with this structure, one act is better than another if and only if that act has greater expected value given every probability distribution contained within the representor.

Given these facts, it is plausible that I ought to conclude the following: *it is neither better nor worse to save Darlington and his fellow birdwatchers than to let them drown.* Both courses of action are equally permissible. After all, one act is strictly preferred to another only if it contains greater expected value given every probability distribution contained within the representor. And I am in just the sort of case where this condition is likely to fail. Conflicting and unresolved evidence leads me to think that the long-term consequences of both acts will be systematically better than those of the other, where the resolution to this evidential conflict remains unclear.

I take it that such a result would fairly straightforwardly follow from maximality: I take, however, such a result to be deeply implausible. Clearly it *is* better to

save the several dozen birdwatchers from painful and premature death at no cost to oneself, and clearly we are capable of affirming these facts at the moment of choice. If maximality consequentialism forces us to deny as much, then there is something deeply implausible about this principle as an account of right action under risk and uncertainty.[32]

1.6 Cluelessness for Non-Consequentialists?

I have tried to present a version of the cluelessness argument has force against subjective consequentialism whether our credences over outcomes are precise or imprecise. The exact concern varies depending upon the exact way in which we cash out our credences; subjective consequentialism coupled with a single set of precise credences over the outcomes plausibly entails that the ideally rationally betterness ordering across our acts is typically beyond us, while subjective consequentialism coupled with imprecise credences over the outcomes (and a popular imprecisionist decision rule like maximality) plausibly entails blanket permissibility.

On both horns of the dilemma, subjective consequentialists find themselves unable to affirm certain commonsense and deeply intuitive judgements of comparative betterness across the options – they find themselves unable to affirm, for instance, that it is better to save the drowning birdwatcher at no cost to oneself. I do not know which of these horns is more or less desirable. Each horn seems a mark against the all-things-considered plausibility of subjective consequentialism, just as the original *Cluelessness Argument* served as a mark against the all-things-considered plausibility of objective consequentialism.

But before closing, though, one final concern needs to be addressed. I have framed the previous discussion as a dilemma *for (subjective) consequentialists*: either subjec-

[32] As far as I can see, my own treatment of maximality is distinct from Mogensen's in the following way. Mogensen's conclusion seems restricted to the concern that the maximality rule will be unable to vindicate orthodox strategies adopted within effective giving movements. But my concern is a more general one: given our general evidential situation as regards the long-run future, it seems that the maximality rule will tell us, for the vast swathe of moral decisions we encounter within the course of our moral lives, that anything goes.

tive consequentialism entails that the betterness ordering is inaccessible or subjective consequentialism entails blanket permissibility. One natural reply to this argument, however, is to note that *many* normative theories care about and place importance upon the consequences of our acts. If that is the case, then isn't the above dilemma one that affects many normative theories equally? And isn't there then something dubious in using the above dilemma to argue *in particular* against subjective consequentialist accounts of right action?

In answer to this objection, let's note the following. Although it is certainly true that every plausible normative theory assigns at least *some* importance to consequences, it would be equivocating to therefore conclude that every plausible normative theory cares about consequences in the same way. Since non-consequentialists might not care about or attach weight to consequences *in the same way* that consequentialists do, we cannot simply assume that non-consequentialists are subject too to the dilemma I have outlined in the previous sections.

One possibility, for instance, is that the non-consequentialist simply takes unforeseeable consequences to be irrelevant as far as right action goes.[33] I take it that this is James Lenman's view in response to the woes of cluelessness, when he suggests that whether or not we act rightly hinges on whether or not we engage in our "local projects" and in doing so live "virtuously, with dignity and mutual respect."[34]

Now, it is not immediately obvious whether a view like this solves the problem of so-called "weakly foreseeable consequences" to which I have devoted so much time and attention. Plausibly, if the non-consequentialist has complex and conflicting evidence as to whether a certain weakly-foreseeable consequence x will follow their act, x might still be importantly relevant in determining the deontic status of their options. The non-consequentialist cannot simply say that x is not of 'moral concern.'[35] (For instance: suppose two options A and B are equally meritorious in every respect,

[33]Lenman, 363. Also see Andreas Mogensen & William MacAskill, "The Paralysis Argument," *Philosophers' Imprint*, 21, no.15 (2021):1–17, esp. 7. This response also features in Gerald Lang, "Consequentialism, Cluelessness, and Indifference," *The Journal of Value Inquiry*, 42 (2008):477–485.

[34]Lenman, 364.

[35]Lenman, 363.

except that you have complex and conflicting evidence that A will lead to a terrible nuclear explosion killing billions of persons in the distant future. Regardless of your normative theory, it seems commonsense to say that B is the more choiceworthy option, since B does not risk the weakly-foreseeable consequences associated with act A.)

I suspect, then, that a more plausible non-consequentialist strategy might go as follows. The non-consequentialist might admit that she has conflicting evidence as to the long-run consequences of two of her available options, A and B. (We can suppose that A and B involve giving versus not giving, respectively, to an effective charity.) But the non-consequentialist might then say something along the following lines. The immediately foreseeable satisfaction of non-consequentialist constraints or duties, or the immediately foreseeable promotion of good consequences, has a *special kind of priority* when it comes to determining the deontic status of our acts. The immediately foreseeable satisfaction of such non-consequentialist constraints or duties, and the immediately foreseeable promotion of certain consequences, plays an especially fundamental role in determining the deontic status of the acts *even if* the non-consequentialist also affirms that distant future consequences are of moral significance and are such that they might play some role in determining the deontic status of our options. It is a view along these lines, I suspect, which will allow the non-consequentialist to give the right answer in the case with which we started: that it is better to save the drowning bird-watcher at no cost to oneself, rather than to leave them be. Given that our evidential situation as regards the long-run future really is so perplexing, I do not know how else such a verdict might be preserved.

1.7 Conclusion

I began this paper by noting that we are clueless as to the total actual consequences of our acts. This result would seem to spell trouble for objective consequentialists, since objective consequentialists determine comparative betterness across the options via reference to total actual consequences. To put the problem a little more sharply:

if objective consequentialism were true, then it seems that we would be unable to affirm certain deeply intuitive and commonsense judgements of comparative betterness across the options that we regularly *do* affirm and take ourselves to know.

A common consequentialist response to the woes of cluelessness, we saw, involves retreating to subjective versions of consequentialism. According to subjective versions of consequentialism, comparative betterness across the acts hinges upon their *expected moral value*. We are not, it seems, clueless when it comes to expected moral value. And so subjective consequentialists need not be clueless when it comes to the comparative status of their acts.

I have argued, though, that this retreat to subjective consequentialism is far from a simple fix. This is because we regularly have complex and conflicting evidence as to whether certain sorts of future consequences are likely to follow our acts, where due to a lack of past data and precedent, the exact way in which this evidential conflict ought to be resolved remains unclear. I have suggested that, if this complex and conflicting evidence really does mandate a uniquely and ideally rational probability distribution over the possible outcomes, the rational probability distribution will typically remain elusive to us moral decision makers, and we will have no guarantee that our attempts to promote expected value will approximate those of our ideally rational counterparts. If this complex and conflicting evidence mandates an imprecise spread of credences over the possible outcomes, and if some plausible imprecisionist decision rule like maximality is true, then it is likely that, for the vast majority of decisions made within our moral lives, anything goes. We face, as I labelled it, the problem of blanket permissibility. Neither result is appealing: under neither result can we affirm, as was the original goal, certain deeply intuitive and commonsense judgements of comparative betterness across the options – not least the judgement that it is better to save the drowning birdwatcher at no cost to oneself. These results, I take it, serve as serious marks against the all-things-considered plausibility of subjective consequentialism as an account of right action.

Chapter 2

Cluelessness and Relevant Claims

In this chapter, I present a new cluelessness argument against a popular family of views in ethics concerning claims aggregation: those views under which agents ought to maximise the strength-weighted satisfaction of *relevant* claims upon their conduct. A claim is said to be relevant if it is sufficiently strong compared to those claims with which it competes. The core of the cluelessness argument will go as follows. In nearly all of our decisions, we change which groups of presently existing individuals are spared severe harms such as painful and premature death. In light of this fact, those who believe in maximising expected relevant claim satisfaction must nearly always ignore minor claims below the threshold of 'death relevance.' This result, I suggest, cannot be right. In the latter half of the paper I consider possible routes of escape for the relevance theorist, concluding that none are entirely successful.

2.1 A Rock and a Hard Place

2.1.1

Suppose that you can spare only one of two groups, X or Y, from suffering severe harm. Suppose that each group contains 100 persons, and that you have not a clue as to the identity of any particular group member. Suppose, in other words, that you find yourself in the following choice scenario:

Choice 1. A Rock and a Hard Place. Perform exactly one of the following.

X-sparing Act: Spare 100 X-members from severe harm.

Y-sparing Act: Spare 100 Y-members from severe harm.

It is worth being upfront about the details of this choice scenario as I am envisioning it. I am assuming that, in making your decision, you do not have a clue as to the downstream causal effects of your choosing to spare one group over the other. I am also assuming that it is beyond your power to *learn* any such facts prior to making your decision. Here is one thing you do know, though: the severe harm which will befall each individual member of the group who suffers is not nice. It involves painful and premature death.

Each individual member of X and each individual member of Y has a *claim* upon you to be spared. These claims conflict insofar as they cannot be jointly satisfied. You must, after all, perform *either* the X-sparing act or the Y-sparing act. Of course, these claims are not the sorts of things 'present in the minds' of the X members or the Y members. In the case as I am envisioning it, the X members and the Y members do not even know that you hold their lives in the balance. But these individuals still have claims on your behaviour insofar as they have a standing interest in not coming to harm, and insofar as this interest would be satisfied under the performance of some of your options but not others.[1]

If these really are the only two options, and if this is the end of the story, then it seems permissible to perform either option. It would be fine to spare the X members, but it would also be fine to spare the Y members. This is because, when all is said and done, you can only spare one group, and because there is no further information waiting in the wings that might serve as a guide to your choice. Of course, it is extremely tempting

[1] I follow Alex Voorhoeve, "How Should We Aggregate Competing Claims?" *Ethics*, 125, no.1 (2014):64–87, in assuming that an individual has a claim *vis-à-vis* your choice if they stand to be counterfactually benefitted by your choice. This fits in with a more general tradition which maintains that all it takes to generate a claim on your conduct is for a potential victim's well-being to vary across the possible outcomes at stake in your choice, where the strength of a claim is determined by the level of variance in welfare. See Matthew Adler, *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis* (New York: Oxford University Press, 2012) and Thomas Nagel, *Equality and Partiality* (New York: Oxford University Press, 1995).

to posit some third option which would fare better against a criterion of fairness: flip a coin, perhaps, in deciding between the options. Admitting that such a third option would be better is fine by me, and cuts against nothing I wish to say in the following sections.[2]

2.1.2

Intuitions become less clear, however, if we modify the case. Suppose that, if you perform the *X*-sparing act, a small Bichon Frise named Spot, belonging to a harbourmaster in Cherbourg, will be spared a cut upon his left hind leg.[3] In other words, suppose you find yourself in the following choice scenario:

> **Choice 2. Rock and a Harder Place.** Perform exactly one of the following.
>
> *X-sparing Act:* Spare 100 *X*-members from severe harm. Additionally, spare Spot a small cut upon his left hind leg.
>
> *Y-sparing Act:* Spare 100 *Y*-members from severe harm.

In Choice 2, just as in Choice 1, we face competing claims on our conduct. In Choice 2, however, our intuitions about the comparative status of the options are liable to conflict in the following way.

One possible view is that, in a case like Choice 2, one ought to choose the option which maximises the total strength-weighted satisfaction of claims *simpliciter*. For each act, that is, figure out what are the claims its performance would satisfy, weight each claim by its respective strength, and sum these values. Choose the act which fares best in this respect, and that is the end of the story.

[2]Note that in speaking of individuals' *claims* in this way, I am not precluding the truth of popular moral theses like utilitarianism or prioritarianism. Such theses can be vividly construed in person-affecting terms or given a person-affecting defence: see Matthew Adler & Nils Holtug, "Prioritarianism: A Response to Critics," *Politics, Philosophy & Economics,* 18, no.2 (2019):101–144.

[3]Spot's injured leg is introduced as an example in James Lenman's "Consequentialism and Clue-lessness," *Philosophy & Public Affairs*, 29, no.4 (2000):342–370. Lenman's original Spot case attempts to demonstrate a fairly particular point about the strength of moral reasons with objective consequentialists might justify their acts.

If this view is correct, then one ought to perform the *X*-sparing act. This is the act that wins out in terms of strength-weighted claim satisfaction *simpliciter*, since it satisfies 100 extremely strong claims for the *X*-members in addition to a further and fairly minor claim presented by Spot. One wouldn't satisfy this further minor claim for Spot by choosing the *Y*-sparing act.

This, however, is not the only possible treatment of the case. Many others think that, when faced with conflicting claims, one should choose the option which maximises the strength-weighted satisfaction of *relevant* claims.[4] A claim is relevant only insofar as it is *sufficiently strong* relative to those other claims against which it competes. Famously, claims against headaches are not relevant when competing with claims against premature death, although claims against headaches might still be relevant when competing with, say, claims against a lost finger.[5] Following Alex Voorhoeve, call this alternative view the *Aggregate Relevant Claims* (ARC) thesis.

Plausibly, when faced with Choice 2, the defender of ARC says the following: the X-sparing act and the Y-sparing act are still both permissible options. Why? Well, note that Spot's claim to be spared a cut competes against the far stronger claims of *Y*-members to be spared from painful and premature death. Given this fact, Spot's claim is plausibly not *relevant* as regards the deontic status of the options. The only claims that are relevant in determining the deontic status of the options are those belonging to the equally balanced *X*-members and *Y*-members.

In setting up Choice 2, I specified that Spot's minor claim was to be spared a cut on his left hind leg. But it is worth emphasising that Spot could have in fact been threatened with a fairly wide range of harms – up to and including a broken leg, perhaps – where Spot's claims to be spared from those harms would *still* fail to qualify as relevant, given that these claims would still conflict with other claims against

[4]Most recently, see Voorhoeve, "How Should We Aggregate?"; also see Kirsten Mann, "Relevance and Nonbinary Choices," *Ethics*, 132, no.2 (2022):382–413 and "The Relevance View: Defended and Extended," *Utilitas*, 33, no.1 (2021):101-110. Other notable defences of relevance include Thomas Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 229-241; Frances Kamm, *Morality, Mortality, I: Death and Whom to Save from it* (New York: Oxford University Press, 1993).

[5]Alastair Norcross, "Comparing Harms: Headaches and Human Lives," *Philosophy & Public Affairs*, 26, no.2 (1997):135–167.

painful and premature death. This need have nothing to do with Spot's being a dog. It is simply to do with the fact that, when a painful and premature death is on the line for some, the claims of others are usually required to be fairly strong in order to count as relevant. I'll address this point in greater depth later; for now, it will not hurt to simply assume that the threshold of relevance (when death is on the line for others) sits somewhere around the point of a lost finger.[6]

2.1.3

Those, then, are two different ways in which we might think about the deontic status of your options in Choice 2. In short, those who believe in maximising the satisfaction of strength-weighted claims are likely to think that the *X*-sparing act is the better one, since the *X*-sparing act satisfies a further and additional claim for Spot. Those who believe in maximising the satisfaction of strength-weighted *relevant* claims are likely to think that the available options (or a fair lottery between them) all remain permissible. This is because Spot's minor claim to be spared a scratch is not relevant: it competes with the (much stronger) claims of *Y*-members against painful and premature death. For now let's just note that, when applied to the Spot case, the relevance approach gives a verdict which is not implausible: in particular, the relevance view seems to capture one intuitive way in which we might pay special respect and attention to those for whom everything is on the line in our choice.

Choice 2 might have struck you as a contrived decision scenario. However, I will now suggest that nearly all of our moral decisions are structurally analogous to Choice 2 in an important respect: in nearly all of our moral decisions, we change which presently existing persons suffer painful and premature death and which are spared. This stunning point may seem like the stuff of science fiction. It turns out, however, that a conclusion along these lines is one we ought to embrace given the empirical facts (upon which I'll dwell shortly).

In light of this result, proponents of relevance views will be forced to make an

[6] See Voorhoeve, 81.

unusual confession. Under relevance views, it will turn out that the sorts of minor claims which we *usually* take to be relevant and evaluatively significant in our moral decision making are in fact almost *never* relevant or evaluatively significant in our moral decision making. Not just Spot's leg, but scrapes and grazes, cuts and scratches, stubbed toes, sprained ankles, broken fingers, dislocated shoulders, ruptured cartilage, burns of varying (but probably not third) degree, ear infections, seasonal flus, tonsillitis, and even, perhaps, lost toes and lost earlobes: typically, claims to be spared from all such harms will not bear on the deontic status of options in moral decision making. This result strikes me as deeply implausible; I will take it to serve as a *reductio* of the relevance view as it stands.

2.2 Chaos and Death Everywhere

2.2.1

The key result I need to demonstrate is that our choices in binary decision scenarios nearly always change which presently existing people suffer severe harm such as painful and premature death. There may be multiple ways of demonstrating this result, but the simplest involves noting some of the most obvious ways in which various physical systems with which we interact are *chaotic*.

When we say that a system is chaotic, we are referencing the fact that even minute changes to the initial conditions can result in substantial changes to its later states. Earth's atmospheric systems, for instance, are famously chaotic: most commonly discussed is the example of Edward Lorenz, who once provided a dynamic model of common convection patterns which entailed just such a sensitivity to initial conditions.[7]

[7]Edward Lorenz, "Deterministic Nonperiodic Flow", *Journal of the Atmospheric Sciences*, 20 (1963):130–141. To say only this much is to miss much of what was groundbreaking in Lorenz' original paper. Lorenz showed that, given variables representing the initial state of a two-dimensional rotational flow (call them x, y and z), and given idealised differential equations modelling the change of these initial variables over time, the future values taken by x, y, and z turn out exhibit incredibly complex behaviour under the so-called *Lorenz Attractor* model. As a result, longterm prediction becomes impossible given even the smallest margins of measurement error in the present. For an accessible treatment of the Lorenz case,

How sensitive, exactly, are the later states of the atmospheric system to its initial conditions? This is not an easy question to answer in any precise terms, not least since Lorenz' original work on convection made a range of simplifying assumptions (to do with the uniformity of air density, for instance) in order to get the model off the ground.[8] In an even more famous lecture, Lorenz seriously entertained the suggestion that a butterfly flapping its wings in Brazil could determine whether or not a tornado obtains in Texas.[9] But as others have noted, the case of the butterfly in a difficult one: the answer may depend on whether or not air viscosity in Brazil is sufficiently thick to dampen the perturbations caused by the butterfly's wings.[10]

Fortunately, we do not need to resolve the issue of butterflies in this discussion. As John Broome and Hayden Wilkinson both note, even seemingly trivial decisions (my choice, for instance, of whether or not to turn on a fan or to release emissions by driving) cause perturbations to airflow many millions, billions, and trillions times greater (with regards to kinetic energy released) than those perturbations caused by the flap of a butterfly's wings. Such perturbations can indeed, it seems, change the initial conditions of the atmospheric system in the relevant (non-dampened) way. Broome's estimate is that we should expect global weather patterns to be entirely altered within a period of mere decades contingent on my decision *vis-à-vis* going for a joyride.[11]

We must take very seriously, then, the possibility that even seemingly trivial individual actions bring radical change to the later states of the atmospheric system. Such seemingly trivial actions may influence, for instance, the exact behaviour and distribution of storms and natural disasters going forward, including which sorts of storms and natural disasters occur precisely where, when, and in what order. Storms and natural disasters, however, are the sorts of things that kill people. And a storm or

see Peter Godfrey-Smith, *Explaining Chaos* (Cambridge: Cambridge University Press, 1998), 9–16.

[8]Godfrey-Smith, 11–13.

[9]Edward Lorenz, *The Essence of Chaos* (Seattle: University of Washington Press, 1995), 181–184.

[10]John Broome, "Against Denialism", *The Monist*, 102, no.1 (2019):110–129; Hayden Wilkinson,"Chaos, add infinitum" (unpublished manuscript), 7. Lorenz also worried that different atmospheric conditions around the equator might confine the butterfly's influence to the Southern Hemisphere – see Lorenz, *Chaos*, 184. If that is true, then perhaps it would be better to speak of coyotes in Texas causing hailstorms in Maine.

[11]Broome, 5-6; Wilkinson, 7.

natural disaster that occurs in Springfield, Michigan, kills different people to a storm or natural disaster that occurs in Springfield, Illinois. And so we arrive at the result that even the most seemingly innocuous of our acts, such as driving or turning on fans, are liable to change which *particular* presently existing people suffer severe harms such as painful and premature death.

Importantly, the claim here isn't that these seemingly innocuous acts involving cars and fans will cause different *future* people to die in horrible, painful, premature deaths in natural catastrophes. The claim is narrower: namely, we must take seriously the possibility that such acts involving cars and fans reshuffle the distribution of horrible, painful, premature deaths across many *present* people.

2.2.2

Another example of a chaotic system by which our actions regularly reshuffle the distribution of harms across presently existing persons is that of traffic.[12] For quite some time now, there has existed an extensive literature on both theoretical and experimental approaches to the study of traffic flow. One influential method of modelling traffic flow for large queues in urban areas, for instance, involves *kinematic waves*, mathematical devices used elsewhere in modelling a much broader range of natural phenomena including ocean currents, mudslides, avalanches, and precipitation run-off.[13] The central point for our purposes is that under the kinematic wave model of congested traffic, "at any point of the road the flow q (vehicles per hour) is a function of the concentration k (vehicles per mile)."[14] The speed of the congested traffic system as a whole depends, in other words, on the exact number and spacing of cars *within* the traffic system. The same goes for the exact timing of "kinematic shock waves" at which traffic experiences sudden drops in velocity.[15]

[12] Wilkinson, 6.

[13] MJ Lighthill & G Whitham, "On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 229, no.1178 (1955):317-345; more recently, GF Newell, "A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory," *Transport Research-B*, 27, no.4 (1993):281–287.

[14] Lighthill & Whitham, 319.

[15] *ibid.*

Such kinematic models imply that the overall states of congested traffic systems are highly sensitive to individual commuter decisions and behaviour: whether or not, for instance, I decide to hold up traffic for five minutes by stealing a stranger's car in rush hour. And since the overall states of congested urban roads turn out to be highly sensitive to the decisions and behaviour of individual commuters, so too are a range of other phenomena which are causally dependant on the exact state of urban roads at any given moment in time (not least, for instance, the behaviour of freeways which take such urban roads as their entry or exit points).[16]

Since individual urban commuter decisions exert a radical influence on the later states of the traffic system, we should expect that individual urban commuter decisions influence the distribution of fatal and serious crashes that will obtain *across* the traffic system in the following hours and days (influencing, for instance, the exact time and location at which crashes occur, and hence, the identities of the particular victims who die or find themselves seriously injured in those crashes). We should also expect the decisions of individual urban commuters to hold lives in the balance in other ways. For instance, many serious harms either come about or are prevented depending whether or not ambulances, police cars, fire trucks, and other emergency vehicles make their destinations on time under tight deadlines; but whether or not these vehicle *do* make their destinations on time at least partly hangs on the overall states of the traffic system at any given point in time.

It will not hurt to substantiate all this with some numbers. In Los Angeles County in 2017, for instance, an average of over 250 road accidents occurred every day: this included an average (per day) of multiple fatalities and around 10 severe injuries, where severe injury includes harms such as paralysis, damage to skull and chest, as

[16]You might worry that under chaotic models of traffic flow we aren't justified in inferring of *any particular commuter* that their choice is a counterfactual difference maker as regards some future harm event. In answer to this worry: the same point could be motivated by drawing on standard models of acceleration and deceleration for congested areas. Under, for instance, the car-following model discussed in GF Newell, "A Simplified Car-Following Theory: A Lower Order Model," *Transportation Research Part B Methodological*, 36, no.3 (2002):195–205, the *nth* vehicle's trajectory follows the *n-1th* vehicle's trajectory translated in space and time. But under such models, individual commuters (in queues, for instance) can still exercise an enormous influence over the commuting trajectories of various others. I'm indebted to Bryce Huebner for emphasising this point.

well as second and third-degree burns covering more than 10% of the body.[17] If Los Angeles County is like any other congested traffic system, then its overall states are highly sensitive to the decisions of individual commuters. And so the exact distribution of these severe harms (including painful and premature death) on any given day will be highly sensitive to the behaviour and decisions of individual Los Angeles County commuters.

2.2.3

I have focused now on two ways in which even the most mundane of our daily moral decisions turn out to be analogous, in an important respect, to Choice 2: in even the most mundane of our daily moral decisions, we in fact determine which groups of presently existing individuals suffer severe harms such as painful and premature death and which groups are spared. In light of this fact, we are now placed to press some uncomfortable results for defenders of relevant claims aggregation. To begin with, consider the following two cases:

> **The Picnic.** You have just arrived at your favourite picnic spot, only to find that it has been taken by a stranger. Your options are to (i) do nothing and drive home now, or (ii) spare the stranger an extremely painful broken finger, at no cost to yourself, by alerting her to the enormous branch which is about to fall. If you choose (ii), you will drive home slightly later than you would have otherwise done.

> **Late for Work.** You are running five minutes late for work. You happen to live on a busy road in Los Angeles County. Your two options are (i) to take your bike like normal, or (ii) to drive. You are utterly indifferent between these two options. By driving, you'll make it to the first intersection *just* in

[17] This data is from the California Highway Patrol's 2017 Statewide Integrated Traffic Records System Report, and can be found at https://www.chp.ca.gov/programs-services/services-information/. It is extremely rare for California to go a day without traffic fatalities: this has occurred a total of five times in the years 2009, 2013, and 2015.

time, and thus be able to save someone from suffering a broken nose and financial loss in a minor car crash.

Of course, these two choice scenarios are highly idealised ones, but the particular idealisations I have made – for instance, including only two of many salient possible options in each case – do not matter for our purposes here.

In **The Picnic**, how should we evaluate the comparative status of option (i) versus option (ii)? Well, it would be a hard bullet to bite if someone suggested that options (i) and (ii) were just as good – or to put it slightly differently, that you should choose either of them. After all, option (i) neglects to save the stranger from a painful broken finger for no good reason. If you don't share this intuition, simply add to the number of picnickers: surely if there were *twenty* or *two-hundred* picnickers, each about to suffer a painful broken finger, or first degree burn, or broken nose, and so on, it would be better to spare them from such a fate at no cost to yourself. Similarly, it would be a hard bullet to bite if someone suggested that the best thing to do involves deferring to a fair lottery between (i) and (ii). That would be to treat the options as if they were on a par, which clearly they are not.

However, given what I have said in the previous section, this is the conclusion at which defenders of relevant claims aggregation must arrive. The defender of relevant claims aggregation must say that (i) and (ii) are both equally good (that you should pick either of them), or perhaps, that it is best to decide between the two options with a fair lottery. Why? Well, the stranger certainly has a *claim* for you to spare her finger by selecting (ii). But typically, claims against broken fingers are not relevant when those claims compete with claims against painful and premature death. And it turns out in **The Picnic** that the stranger's claim against a broken finger *does* compete with claims against painful and premature death. So, the stranger's claim is not relevant.[18]

[18]Why, in **The Picnic**, does the stranger's claim compete with claims against painful and premature death? Well, in this particular case you make your decision in the open air of the park, and your body will cause different perturbations in the air depending on which option you perform. Your car will cause different perturbations in the air depending on whether or not you drive home now or later, and the same goes for the stranger's car. Additionally, depending on your choice either way, the stranger's next few days and weeks are likely to change in various important ways. All these changes will cause various small and large perturbations in the initial conditions of the atmospheric system, and *arguendo*,

In **Late for Work**, how should we evaluate the comparative status of option (i) versus option (ii)? As before: it would be a hard bullet to bite if someone suggested that options (i) and (ii) were just as good – or to put it slightly differently, that you should choose either of them. After all, option (i) neglects to save a stranger from substantial pain and discomfort for no good reason. If you don't share this intuition, as before, simply add to the number of commuters. Surely if there were *twenty* or *two-hundred* commuters, each of whom you could save from broken noses and financial loss at no cost to yourself, then you should spare them from this fate. Similarly, it would be a hard bullet to bite if someone suggested that the best thing to do involves deferring to a fair lottery between (i) and (ii). That would be to treat the options as if they were on a par, which they are clearly not.

However, given what I have said in the previous section, this is the conclusion at which defenders of relevant claims aggregation must arrive. The defender of relevant claims aggregation must say that (i) and (ii) are just as good (that you should pick either of them), or perhaps, that it is best to decide between the two options with a fair lottery. Why? Well, the stranger certainly has a *claim* for you to spare him from moderate discomfort. But plausibly, claims against broken noses are not relevant when those claims compete with claims against painful and premature death. And it turns out in **Late for Work** that the stranger's claim against a broken nose *does* compete with claims against painful and premature death. So, the stranger's claim is not relevant.[19]

These conclusions are deeply unnerving. It turns out that when faced with even the most seemingly mundane of decisions, defenders of relevant aggregation cannot take account of minor (but non-trivial) claims on our conduct. Such minor (but non-trivial) claims are not relevant when it comes to determining the permissibility of the options. This is despite the fact that we usually take claims against, say, broken fingers and

such perturbations eventually go on to change who lives and who dies.

[19]Why, in **Late for Work**, does the stranger's claim against a broken nose compete with other claims against painful and premature death? Well, the overall states of the urban sprawl of Los Angeles County throughout the course of the day are highly sensitive to the decisions and behaviour of individual commuters just like you. The exact distribution of fatal car crashes within this urban sprawl are hence *also* highly sensitive to the decisions and behaviour of individual commuters just like you. Your decision will, in the end, determine which groups of presently existing persons suffer serious injury and premature death on the road in the course of the following days.

broken noses to bear on the matter of what we ought to do.

2.3 Five Responses

Let us take stock. I have suggested the following. In many even seemingly mundane decision scenarios, it turns out that our decisions change which presently existing persons suffer severe harms such as painful and premature death and which are spared. This is due to the chaotic nature of the physical systems with which we regularly interact. In light of this fact, defenders of relevant claims views are forced into an unusual confession. Many of the smaller claims which we *usually* take to be straightforwardly relevant in moral decision making (a stranger's claim to be spared a broken finger or a broken nose) turn out to be deontically irrelevant. Such claims turn out to be deontically irrelevant because they turn out to always compete against other and much stronger claims against painful and premature death.

We might take this result to show, by *reductio*, that relevant claims views are false: after all, claims against broken fingers and broken noses clearly *do* factor into the deontic status of our options, and they do so all the time. If relevance approaches to claims aggregation ask us to deny as much, then these approaches must be false.

One could instead, however, take the *tollens* as a *ponens*. Under this strategy, the central insight behind relevance views (that only relevant claims can factor into the deontic status of our acts) remains as true as it ever was. It just turns out that, as a surprising empirical matter of fact, minor claims such as those to be spared from broken fingers *always do* compete with far, far, stronger claims against painful and premature death. And so it turns out that the range of minor claims that we usually take to be relevant to deontic status, claims such as those to be spared broken fingers, are not in fact relevant to deontic status after all.

I do not think that this attempt to bite the bullet, as it stands at least, will have many subscribers. It *just seems* that broken fingers can and do bear on the comparative status of our acts in a range of decision-making scenarios, for instance, in **The Picnic**

and **Late for Work**. It would be revisionary to much of our moral thought and practice to suggest that they in fact never do. And so instead, I will now turn to consider some of the other possible responses that the defender of relevant claims aggregation might give by way of resisting the implications of the argument just outlined. One benefit of the discussion, as we will see, is that it offers a chance to think in depth about how relevance approaches to claims aggregation might operate in situations of risk and uncertainty.

Option 1: Dispute the empirical facts

Defenders of (say) ARC might push back at the way in which I have presented the empirical facts: they might push back, for instance, against my suggestion that individual commuter decisions in urban areas almost always alter the distribution of serious harms upon present people in motor vehicle crashes.

I do not place much hope in this line of thought. Take traffic: popular models of urban congestion make it highly plausible that the distribution of serious harms that occur *across* urban traffic systems are highly sensitive to individual commuter choices such as those featuring in **Late for Work**. Perhaps this pushback *vis-à-vis* the empirical facts is more plausible in the case of the weather: there is still a lot about atmospheric systems that we do not know.[20]

Note, however, that I have considered only two possible sources of physical chaos in the previous section. There are, I suspect, many more such sources of chaos which very speedily reshuffle the distribution of severe harms across present persons. For instance, even the most seemingly trivial of our decisions are *identity-affecting*: they influence the identity of immediately future persons in virtue of influencing the exact moment at which parents conceive.[21] But when our decisions influence the identities of immediately future persons, they also reshuffle the distribution of severe harms across *present* persons, since immediately future persons make all kinds of decisions

[20]For a recent discussion of the butterfly case, see TN Palmer, A Döring & G Seregin, "The Real Butterfly Effect", *Nonlinearity*, 27, no.9 (2014):R123.

[21]Lenman, 346; Wilkinson, 5; Hilary Greaves, "Cluelessness", *Proceedings of the Aristotelian Society*, 116 (2016):311–339. Also see Derek Parfit, *Reasons and Persons* (Oxford University Press: Oxford, 1984), 352.

which have various causal impacts upon the welfare of people who presently exist and will continue to do so (decisions which include, not least, whether or not to drive in urban traffic). I take, in light of this further fact, scepticism about the empirical facts to be untenable.

Option 2: The 'commuters don't have claims' response

A fairly different style of response goes as follows. You might think that those random commuters stuck in the chaos of traffic, those whose fate hinges on my mode of transport, do not have *claims* on my conduct. You might think that these commuters do not have claims on my conduct because, somehow, of the causal or epistemic facts about the way in which my choice relates to their well-being.

Here is one way of spelling out this thought. You might think that those random commuters stuck in the chaos of traffic do not in fact have claims on my conduct because the causal link between my action and their welfare is somehow *too weak*. After all, I am only one commuter out of millions. Which particular set of victims gets harmed in the traffic on any given day will be a function of not only my decision, but of the decisions of these millions of other commuters too. Since I am only one of the many commuters who together causally determine which particular set of victims gets harmed in traffic, no particular victim of harm can have claims of much strength on my conduct. Or so the thought goes.

On reflection, though, we should abandon this first-pass suggestion. It is certainly true that the later states of the urban traffic system are sensitive to not just *my* individual decisions on any given day but also to the decisions made by countless *other* drivers who each navigate their own daily commute. But this point alone does not undercut the fact that my individual decision on its own will be a counterfactual difference maker of great importance: my individual decision will determine which groups of presently existing commuters will be spared and which will not. This fact should be enough to generate claims on my conduct: those individuals who might suffer in traffic accidents have a standing claim on me to bring about the counterfactual state of affairs in which

they are spared.[22]

Here is a second way of spelling out this thought. These random commuters do not in fact have claims on my conduct because I am *ignorant of their particular identities*, and also incapable of learning such facts. This second strategy amounts to imposing a fairly strict epistemic constraint on which particular claims can bear on the deontic status of options: in order for x's claim to bear on the deontic status of my options, I must know or at least be capable of reasonably learning x's particular identity.

But a moment's reflection should lead us to abandon this second-pass view too. It is not at all obvious that I must know x's identity in order for x's claims to play a role in determining the deontic status of my options. I might know that a stranger, somewhere in Los Angeles County, will be horrifically electrocuted in the following days if I fail to give my nose the smallest of scratches. The fact that I am not capable of learning this stranger's identity does not mean that their claim to be spared is irrelevant to my conduct and to the status of my options. What seems to matters in figuring out whether the stranger has a claim, rather, is whether the stranger (whoever they are) would be a counterfactual beneficiary of my choice.

The discussion in this section has been fairly condensed, but there is a reason why. Voorhoeve's canonical statement of ARC relies on a fairly weak conception of what it is for someone to have a claim on your conduct: someone has a claim on your conduct if they would be benefitted given your conduct relative to a baseline outcome in which you *hadn't* acted, where the strength of a claim is grounded in the difference in that person's well-being across the two outcomes.[23] The ARC, in adopting this fairly weak conception of the word 'claim', follows a more general tradition of seeing some individual x's claims as being generated when x's well-being would differ across the possible outcomes at stake in your choice.[24] If this is what it takes for claims to be generated upon your conduct, then it should not be too controversial to say that random commuters stuck in traffic have claims on your conduct. That is because, as

[22]This quick-and-easy answer becomes complicated if one thinks that claims on our conduct depend on, or are generated by, the *ex ante* risk of harm we impose upon others. More on this point later.

[23]Voorhoeve, 66.

[24]See Adler, *Well-Being*; Nagel, *Equality*.

I have argued, your choice of transport either way is going to make a counterfactual difference as to which commuters suffer and which are spared in the following days.

Option 3: Lower the relevance threshold

Alternatively, defenders of ARC might attempt to lower the threshold of 'relevance' for minor claims when those claims are found to compete with the claims of others against death. Suppose, for instance, that claims against broken fingers always *do* turn out to be relevant when competing with claims against painful and premature death. If that were the case, then defenders of relevance views could give the right answer in **The Picnic**. Defenders of relevance views could maintain that, in **The Picnic**, (i) is the better option insofar as (i) satisfies an additional and relevant claim from the picnicking stranger and her broken finger.

I do not think this strategy will work either. Recall that the original motivation behind the relevance view was the thought that *no* number of claims against headaches (for instance) can compete with any number of claims against painful and premature death. If we substantially lower the relevance threshold such that headaches, for instance, are now relevant to painful and premature death, we give up on this original motivation. It will turn out that, in a range of decision scenarios, claims against headaches can outweigh claims against death. And this was the exact result that defenders of relevance wanted to avoid.

One more precise framing of this objection would go as follows. The defender of relevance views could object that, in the above examples, I have consistently set the relevance threshold *a little too high*. We do not need to lower the relevance threshold to the point of slight headaches, but it may still be the case that claims against broken fingers, very minor crashes, and so on, can still be relevant even when death is on the line for others.

I have been fairly careful, though, to use examples in which the minor harms at stake are non-trivial, while also being such that claims to be spared them are likely to fall below the relevance threshold given plausible views of *how* we go about setting the

relevance threshold. One of Voorhoeve's central insights in his original defence of ARC, for instance, was that the relevance threshold is tied closely to the level of self-interest that individuals, from their first-personal perspective, could or could not legitimately demonstrate when faced with claims from others.[25] If someone were facing premature death, I could, in the name of my own first-personal self-interest, legitimately refuse to save them on the grounds that it would cost me two of my limbs. My claim against having two limbs sacrificed hence sits *above the relevance threshold* even when others face death. If someone were facing premature death, I could not, in the name of my own first-personal self-interest, legitimately refuse to save them on the grounds that it would cause me a headache. My claim against having a headache thus sits *below the relevance threshold* when others face death.

On this model, it strikes me that the sorts of minor harms made salient in **The Picnic** and **Late for Work** clearly aren't relevant; they clearly fall below the threshold of 'death relevance.' For instance, if someone were facing painful and premature death, I could not, in the name of my own first-personal self-interest, legitimately refuse to suffer a broken finger in order to save them. So, it seems clear to me that my claim against suffering a broken finger falls below the relevance threshold when others face death.

Voorhoeve's suggestion, about the relevance threshold being tied to the particular level of sacrifice that individuals could legitimately refuse in the name of self-interest, is not the only means by which we could attempt to set the relevance threshold. But it strikes me as a good a means as any, and on models such as this, when competing with claims against death, the sorts of minor claims I have described in **The Picnic** and **Late for Work** plausibly do not qualify as relevant.

Option 4: Endorse a lexicographic version of ARC

One possibility, in light of the difficulties faced so far, would be for defenders of ARC to fall back upon a tiebreaker or lexicographic model. In broad strokes: when both relevant and irrelevant claims are on the line, say that the former have lexical priority

[25] Voorhoeve, 71.

over the latter with regards to the moral status of acts. An act is permissible if and because it fares better with respect to the satisfaction of strength-weighted relevant claims; in the special case that two acts ϕ and ψ fare just as well with respect to the satisfaction of strength-weighted relevant claims, ϕ is permissible if and because ϕ fares just as well as ψ with respect to the strength-weighted satisfaction of *irrelevant* claims.

Such a lexicographic view says of Choice 2 (in which one can spare either 100 X-persons or 100 Y-persons) that one should choose the option which satisfies an additional minor and irrelevant claim for Spot. Such a minor claim for Spot can indeed act as a tiebreaker as regards the deontic status of your options, despite the fact that claims to be spared from scratches are not relevant when compared with claims to be spared from death.

Some defenders of relevance have historically rejected tiebreaker models such as these on the grounds that it would be inappropriate to settle high-stakes claim ties by turning to consider the comparatively trivial claims of others.[26] Here is one fairly compelling way of putting the point, broadly following some suggestions made by Bastian Steuwer.[27] Plausibly, when we say that x's claim is irrelevant to y's, we are saying something like the following: in deciding whether or not to satisfy y's claim, it would be disrespectful to employ x's claim anywhere in our deliberative process *regarding* y and their claim. To even "consider" x's claim when deliberating about y and their claim would seem to trivialise y's claim which was, from the start, of a much graver nature.[28] If that is the case, though, then the possibility of using x's claim as a tiebreaker when y*'s* claim competes against some further relevant claim z is precluded. That would be to use x's claim in our deliberative process concerning whether or not to satisfy y's claim.

[26]See, for instance, Kamm, 146, who explicitly rejects irrelevant interests as tiebreakers in the *Principle of Irrelevant Utilities.* Note that Lazar's non-contractualist model of limited aggregation does incorporate this lexical structure; see Seth Lazar, "Limited Aggregation and Risk," *Philosophy & Public Affairs*, 46, no.2 (2018):117–159, esp. 125.

[27]Bastian Steuwer, "Aggregation, Balancing, and Respect for the Claims of Individuals," *Utilitas*, 33, no.1 (2021):17-34.

[28]Steuwer, 24.

In any case, supposing that a tiebreaker model *is* compatible with the justifications and rationales usually given in support of a relevance constraint, the tiebreaker model I just outlined a moment ago does not yet solve a case like **Late for Work.** That is because **Late for Work** involves a certain level of risk: although you know that your choice either way is going to change who suffers and who is spared, you do not know, for instance, that your choice either way will spare the *same* number of persons from suffering severe harm. The tiebreaker model, in order to be of any use in a case like this one, would need to be expanded.

For choices in which one is uncertain how many and which relevant claims an option might satisfy, we might endorse the following lexicographic view. An act is permissible if and because it fares better with respect to the strength-weighted and probability-weighted satisfaction of relevant claims. An act is permissible, in other words, if it satisfies the greater sum of strength-weighted relevant claims *in expectation*. In the special case that two acts ϕ and ψ fare just as well with respect to the satisfaction of strength-weighted relevant claims in expectation, ϕ is permissible if and because ϕ fares just as well as ψ with respect to the strength-weighted satisfaction of *irrelevant* claims in expectation.

Under a lexicographic view like this, the **Late for Work** case is solved. Both options fare just as well with respect to the expected satisfaction of relevant claims, but only one option is permissible, since only one option wins out in terms of the expected satisfaction of irrelevant claims (namely, the option in which you save a stranger from the trauma of a very minor car crash). This was the intuitive result we wished to vindicate all along.[29]

If a suggestion somewhere in this vicinity were correct, then we might see all this as a strong argument that relevance views need to go lexical: this in and of itself would be a non-trivial result given that, as I already emphasised, certain defenders of relevance

[29]Why do both options fare just as well with respect to the expected satisfaction of relevant claims? This suggestion relies on a form of indifference reasoning which I will not try to dispute here. The thought is simply that, in the absence of any relevant evidence, you should think each option just as likely to spare just as many random commuters from just as severe crashes in the coming days. Hence, each option is associated with the same level of relevant claim satisfaction in expectation). For a related discussion see Hilary Greaves, "Cluelessness", *Proceedings of the Aristotelian Society*, 116 (2016):311–339.

already take the use of tiebreakers to be inimical to the spirit of their view. That being said, I'll now turn to consider a more general response to the argument of the previous sections, one which draws further on this central idea of risk, without presupposing a lexicographic approach to the satisfaction of relevant and irrelevant claims.

Option 5: Ground claims in the *ex ante* risk of harm

Finally, defenders of ARC might fall back upon an *ex ante* model of claims aggregation. Under the *ex ante* model, an individual only has a claim on my conduct insofar as, at the moment of choice, in virtue of my choosing one particular option rather than another, that particular individual is given *worse prospects* than they otherwise would have had. (To figure out an agent's prospect's under one of your acts, do the following: ask how well-off that agent would be under each possible outcome given the performance of your act, weight this value by the likelihood of that outcome obtaining given the performance of your act, and then sum these values. I'll assume that the likelihoods are epistemic ones: they are the likelihoods that epistemically rational agents would assign to a particular outcome at the moment of choice, given any and all relevant evidence.)

If one believes in both the *ex ante* model, as well as in a relevance constraint, then one might arrive at the following view: decision makers should choose the act associated with the greatest expected satisfaction of relevant claims, where those claims arise in virtue of potential victims receiving worse prospects than they otherwise would have had.

An *ex ante* relevant claims model such as this one offers to defuse the argument of the previous sections in the following way. In **Late for Work**, for instance, my choice between (i) and (ii) does not increase the *expected* number of deadly crashes within the swirl of chaos on any given day. And so, at the moment of decision, my choice between (i) and (ii) does not raise the *ex ante* chances *for any particular* commuter that they will end up dead in a freak crash in the course of the coming days.

In light of this fact, it seems that no *particular* commuter stuck in the swirl of chaos

of traffic has a claim on my decision between (i) and (ii). The only person with a claim on my choice in **Late for Work**, then, ends up being the poor stranger whose nose is at stake. I thus ought to spare his nose.

There is something deeply intuitive in this response: it simply seems *true* that, since each of my options brings about the same *ex ante* risk of harm to each random commuter in traffic, none of these random commuters in traffic have a claim on my conduct either way at the moment of choice. This may have seemed, as it were, the simple solution all along.

I will now suggest, though, that things are not quite so simple. First, it is important to note that *ex ante* claims models along these lines have recently been the subject of sustained and damaging criticism. These criticisms include: the fact that *ex ante* claims models are improperly biased towards identified rather than merely statistical lives[30]; the fact that *ex ante* claims models do not permit many lower risks of (relevant) harm to aggregate such that they outweigh fewer higher risks of (relevant) harm[31]; and the fact that *ex ante* claims models give implausible or inconsistent verdicts in sequential choice.[32]

In light of this fact, this fifth defence, even if successful, lands us in something of a dialectical dilemma. On the one hand, solve the cluelessness argument I have outlined by adopting an *ex ante* approach to claims aggregation: in doing so, however, open the path to objections from identified versus statistical lives, harms aggregation, and sequential choice. Or on the other hand, solve the problems of identified versus statistical lives, harms aggregation, and sequential choice by abandoning the *ex ante* model: in doing so, however, open the path to the cluelessness argument of the previous sections. As things stand, neither fork of the dilemma presents itself as

[30]Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs,* 43, no.3 (2015):175–223. Frick's response to this problem is to scale back the contractualist project: to say that it is simply a higher-order wrong-making property of an act that it is prohibited by contractualist principles.

[31]See S.D. John, "Risk, Contractualism, and Rose's 'Prevention Paradox'," *Social Theory and Practice*, 40 (2014):28–50, esp. 46-47; Joe Horton, "Aggregation, Complaints, and Risk," *Philosophy & Public Affairs*, 45, no.1 (2017):54–81. Lazar, 138, coins this a violation of the principle *Aggregate Against Risks of Relevant Harms*.

[32]Christoph Lernpaß, "A Diachronic Argument Against the Ex Ante Complaint Model" (unpublished manuscript).

particularly appealing. I wish to conclude, however, by making matters worse: by raising what strikes me as a yet undeveloped worry for proponents of an *ex ante* relevance approach, and particularly, a worry for those who would see the *ex ante* relevance approach as a remedy to the cluelessness argument I've outlined in the previous sections.

The Shark Tank I

Imagine a morbid game show with 100 contestants. Each of the contestants is assigned a different number between 1 and 100. Each contestant also happens to find themselves balanced precariously above a vat of hungry sharks. Earlier, in secret, the presenter rolled two 100-sided dice and covered each die with a box. As the guest host, you must choose which box to open. If you open the first box, then whichever number the first die reads, that contestant will be spared: all the others will be dropped into the shark-infested water. If you open the second box, then whichever number the second die reads, *that* contestant will be spared: all the others will be dropped into the shark-infested water.[33]

Since this is the structure of the show, note the following facts. First note that, at the moment of choice, regardless of whichever box you choose to open, the prospects for each individual contestant on the podium remain unchanged: specifically, each individual contestant faces a 99% chance of being eaten alive by hungry sharks. (As I mentioned previously, I take the probabilities relevant for determining prospects to be the epistemically rational ones held by decision makers at the moment of choice.) This would seem to indicate, under the *ex ante* rule just considered, that none of the contestants have claims upon your conduct either way: none of them have a claim for you to pick one box rather than the other.

But second note that, in choosing to open one box rather than another, you almost certainly change which particular contestants live and die: this is because two 100-

[33]The case as I am imagining it is not one in which you *inflict harms* upon the ninety-nine others. Imagine, for instance, that you have been forced to play the game at gunpoint by a television executive, and that should you refuse to play every single contestant will die.

sided dice are exceptionally likely to give two different integers on any two given rolls. It is overwhelmingly likely that by opening the first box rather than the second, or vice versa, some *different* contestant will be let off the hook than would have otherwise been.

In light of this second fact, you might find yourself reasoning with the television executives behind the camera in the following way: "Hold on a second. You and me both know at the moment of choice that my decision is almost certainly going to change who lives and who dies. But note what follows from this. It follows that there almost certainly exists some contestant on the stage, standing right in front of us, who is properly picked out by the definite description *The Contestant Who Survives Under the First Box Only*. This contestant will survive if I open the first box, but will perish if I open the second. Their *ex ante* prospects in fact *change greatly* depending on which box I pick. They are the exactly the sort of contestant who can thus have a claim on my conduct."

This point complicates the story. Contestants are going to have invariant *ex ante* prospects across your options if we designate them in some ways (using labels like *Contestant 73*, for instance) but not if we designate them in other ways (using labels like *The Contestant Who Survives Under the First Box Only*, for instance).[34]

Now, in this initial version of the shark tank game show, we can sidestep this complication fairly easily. You don't know *which* contestant is picked out by the designator *The Contestant Who Survives Under the First Box Only*. You also don't know which contestant is picked out by the designator *The Contestant Who Survives Under the Second Box Only*. You should assign, then, equal credence to any and every contestant that they will fall under either designator. If you care about maximising the expected satisfaction of relevant claims, then, you will do just as well by picking either box. That is the first puzzle solved.

[34]Some of these issues of designation are foreshadowed in Bastian Steuwer, "Contractualism, Complaints, and Risk," *Journal of Ethics and Social Philosophy*, 19, no.2 (2021):111–147. See also Michael Otsuka, "Risking Life and Limb: How to Discount Harms by their Improbability," in *Identified versus statistical lives: an interdisciplinary perspective*, eds. Glenn Cohen, Norman Daniels, and Nir Eyal (Oxford University Press: Oxford, 2015), 77–93.

The Shark Tank II

But then suppose you learn the following. If you choose the second box, a further benefit will arise: an innocent puppy named Spot will be spared a small cut to his left hind leg. The rest of the game show remains unchanged. In this second version of the puzzle, assuming the *ex ante* relevance model, which box ought you choose?

Well, you might argue that Spot is the only one with a claim of any strength on your conduct, and that as a result, you ought to choose the Spot-friendly option. Just as before, you might argue, none of the individual contestants on the podium have a claim on your conduct since none of the individual contestants' prospects change depending on your choice.

But just as before, we might reply to this point with an emphatic *not necessarily*. If we had instead designated the contestants using labels like *The Contestant Who Survives Under the First Box Only*, then it is *not* true that the contestants have invariant *ex ante* prospects of harm. A contestant such as *The Contestant Who Survives Under the First Box Only* has an extremely strong claim on you to pick the first box under which they are saved. And such a claim emerges a in direct competition with the satisfaction of a much smaller claim for Spot. This would seem to indicate that Spot's claim comes out as deontically irrelevant: it is defused by the much stronger claim presented by *The Contestant Who Survives Under the First Box Only.*

You might reply to this point: we don't know, at the moment of choice, *which* contestant is picked out by a designator like *The Contestant Who Survives Under the First Box Only*. We don't know, for instance, that contestant's proper name. But note that this isn't the issue at stake in this second puzzle. The issue at stake in this second puzzle is whether *Spot's* claim can here bear on the issue of what you ought to do. And as I've just outlined, given the problem of designation, there is an argument for thinking that it cannot. Spot's claim competes in a counterfactual competition with the much graver interests of another contestant whose *ex ante* prospects emphatically aren't invariant across your options.

The argument here is fairly intricate, so it is worth summarising what we have

seen. We first assumed that the *ex ante* relevance approach was correct. Then we noted the following. Since your choice in the shark tank show is overwhelmingly likely to change who lives and who dies, contestants be put under certain designations at the moment of choice (*The Contestant Who Survives Under the First Box Only*) such that their *ex ante* risks of harm are not invariant across your option sets. When designated in this way, the contestants still have extremely strong claims on your conduct. If Spot were to present a minor claim on your conduct – a claim, say, to select one box rather than another – then it would emerge in a direct competition with these extremely strong claims, belonging to individuals' whose prospects emphatically aren't *ex ante* invariant across your options. Plausibly, in light of this fact, Spot's claim would count as irrelevant.

To put all this in more general terms: *even assuming* that the *ex ante* relevance approach is the way to go, this alone does not get out of the argument of the previous sections. Since our choices almost always change who suffers and who is spared from severe harms, we can almost always designate the possible beneficiaries of our acts using the labels *The Person Who is Spared Under the X-act* or *The Person Who is Spared Under the Y-act*. When possible victims and beneficiaries are specified in this way, their *ex ante* prospects are not invariant, and minor claims which *compete* with these much stronger claims are going to come out as deontically irrelevant. This was the deeply unpalatable result from which we have been trying to save the relevance view all along.

It is immediately tempting to dismiss a definite description such as *The Contestant Who Survives Under The First Box Only* as somehow gerrymandered, and as a result, to dismiss this whole line of thought as somehow improper. But this temptation should be resisted. Given the empirical details of the case at hand you do know, at the moment of choice, that this definite description almost certainly refers to one of the one-hundred candidates standing on the podium in front of you. And this definite description is as clean and well-behaved as they come: it is not as if the referent of the definite description will change, for instance, depending on whichever option you select (as would be the case for some choice-dependent definite description such as

The Contestant Who Survives Given My Choice).

Sympathetic Identification and Multiple Designators

All this points, in the end, to a question that to the best of my knowledge no defenders of a relevance constraint have explicitly considered. This is the general methodological question of whether a minor claim comes out as irrelevant so long as (i) that minor claim comes out as irrelevant under *some* legitimate description of the competitors against whose claims it competes, or so long as (ii) that minor claim comes out as irrelevant under *every* legitimate description of the competitors against whose claims it competes.[35]

Although I doubt I have the words here to spell out a detailed answer to this general methodological question, I would like to conclude by tentatively entertaining the following thought. Perhaps Spot's claim should come out as deontically irrelevant *simpliciter* given that it comes out as deontically irrelevant under one particularly salient description of the competitor against whose claims it competes – given that it comes out as deontically irrelevant when we use the label *The Contestant Who Survives Under The First Box Only*.

The reasoning, here, goes as follows. One rationale commonly offered in support of a relevance constraint is the thought that the treatment of claims must be somehow grounded in a process of *sympathetic identification*.[36] The thought here is that, when faced with conflicting claims, one must attempt to take on the first-personal perspective of each potential victim whose interests are at stake in your choice. A claim is then said to be *relevant* if it is the sort of claim that I, as a victim, given my first-personal perspective, could legitimately satisfy for my own sake at the expense of the interests presented by others.

Let's indeed suppose that the process of sympathetic identification underpins, in some important way, a relevance constraint on claims aggregation. Well, Spot is

[35]These, admittedly, are not the only two options. For the issue of descriptors in the context of *ex ante* views, see Anna Mahtani, "The Ex Ante Pareto Principle," *The Journal of Philosophy*, 114, no.6 (2017):303–323.

[36]Voorhoeve, 70-72.

extremely confident, *ex ante*, that he presents his claim in a directly competitive context. He is extremely confident, *ex ante*, that by satisfying his own claim against a cut, he will undercut the claim of *The Contestant Who Survives Under The First Box Only*. Spot can sympathetically identify with this contestant: he can put himself in their shoes. And Spot can ask whether he, Spot, could reasonably spare himself a minor cut when doing so would undercut the much stronger claim of *The Contestant Who Dies Under the First Box Only*. It strikes me as implausible to think that Spot could satisfy his own interest in not receiving a cut at the expense of this particular contestant's much weightier interests in not dying painful and premature death. But this means that Spot's interest, in the shark-tank themed game show, comes out as deontically irrelevant. Spot's interest does not pass the test of sympathetic identification.

Importantly, it does not seem to make a difference, here, that Spot *does not know* at the moment of choice which particular contestant is picked out by the definite description *The Contestant Who Survives Under the First Box Only*. Why? Well, it is perfectly possible for Spot (at the moment of choice) to sympathetically identify with this contestant, *The Contestant Who Survives Under the First Box Only*, *even if* Spot does not know the particular proper name of this particular contestant. This is simply the nature of sympathetic identification: one uses imaginative faculties to *suppose* that you occupy the shoes of such-and-such a person, where this process of supposition only requires a rudimentary grasp of particular basic facts about such-and-such a person's circumstances, and where this process of supposition does not demand knowledge of such-and-such's proper name. One can sympathetically identify with *The Bank Teller Who Works Mondays*, *The Winner of the Race*, and *The Next Governor of California*, regardless of whether one knows certain further facts about these individuals – facts including, for instance, the names by which their friends call them at the pub.[37]

The suggestions in this final section, as I have emphasised, are only tentative. But the view I have outlined strikes me as plausible as any. In the shark-tank themed

[37]This discussion is closely related to Steuwer, 126. The more general thought here is that, in figuring out whether our actions are justifiable to others, we do not need to know (at the moment of choice) the particular identities of those who are 'doomed' by our actions.

game show, if you know *ex ante* that your choice will change who lives and who dies, then you also know *ex ante* that there exists some contestant on the podium picked out by the definite description *The Contestant Who Survives Under the First Box Only*. It is possible for you and Spot to sympathetically identify with this contestant: to ask whether either of you could spare yourself a minor wound, if doing so meant certain death for *The Contestant Who Survives Under the First Box Only*. The answer to this question seems to me to be no, since sympathetic identification can occur even when a potential victim's proper name remains a mystery. And so, the process of sympathetic identification should lend support to the thought that Spot's claim still comes out as deontically irrelevant *simpliciter*.

2.4 Conclusion

Those, then, are five possible responses that may tempt defenders of relevance views. I've suggested that none are entirely successful. And so you might think that we arrive back where we started, with defenders of relevance views admitting that minor claims are almost always deontically irrelevant. Minor claims (recall: claims to be spared cuts, scratches, stubbed toes, sprained ankles, broken fingers, dislocated shoulders, ruptured cartilage, burns of varying but not third degree, ear infections, seasonal flus, tonsillitis, lost toes and perhaps lost earlobes too) can barely ever bear on the deontic status of our options. This is because claims to be spared from such harms almost always compete with the much stronger claims of others to be spared painful and premature death. And *this* is the case because it turns out that we regularly interact with complex and chaotic systems, thereby inadvertantly changing the distribution of severe harms across present persons.

The final two responses were, I think, the most promising. Indeed, one reading of all this is that defenders of relevance views must either go lexicographic or go *ex ante* or go both. I've tried to suggest, however, that neither of these manoeuvres are entirely simple. As I mentioned previously, for instance, it is not obvious that the lexicographic

model always pays proper respect to the claims presented by potential victims, in at least one possible sense of the word 'respect.' And as I mentioned previously, *ex ante* models of claims aggregation have recently been the target of sustained and damaging criticism. Even for those who endorse *ex ante* models of relevant claims aggregation, there is still the lingering question of whether such models do entirely resolve the riddle of this paper. Given that your decision in a case like **Late for Work** changes who lives and who dies, there are salient ways of designating those individuals who might suffer severe harms under your act such that their *ex ante* chances of harm are *not* invariant across your option set, such that those individuals might present weighty claims on your conduct, and such that those weighty claims might defuse minor claims.

Earlier, I suggested there are two ways of taking the overarching result of this paper – the overarching result that, for defenders of relevance views, minor claims almost never count. The first option is to say that, by *reductio*, relevance views of claims aggregation must be false, or perhaps, must stand in need of radical revision. This is the conclusion at which I have arrived. The second option is to bite the bullet that minor claims to be spared from harm never count: one simply accepts this as a necessary consequence of relevance views. As I said earlier, this would be a surprising result. It would mean radical revisions are needed in much of our moral thought and practice, though I have not attempted to spell out here what such revisions would involve.

If the argument of the previous sections is successful, then we have a cluelessness argument that bruises the relevance view. Till now at least, cluelessness arguments have typically been used in arguing against standard subjective and objective consequentialist theories of right action. But I have shown that the this is not the whole story; the standard narrative around cluelessness needs to be reframed. Given the chaotic world in which we live, the features of the relevance view which make it uniquely appealing are also those which make it uniquely vulnerable to the problems of cluelessness.

Chapter 3

Risky Doings and the Doing of Risk: A Reply to the Paralysis Argument

Recently, Andreas Mogensen & William MacAskill have defended a novel argument against non-consequentialists who maintain a traditional distinction between moral reasons for doing versus allowing harm. The Paralysis Argument, as they coin it, attempts to demonstrate the following point: if the doing versus allowing distinction were true, we would have greater subjective reasons against active behaviour than we would have against voluntarily entering a state of paralysis. In this paper, I explore and reject some preliminary replies to the Paralysis Argument. I then suggest a replacement of the doing versus allowing doctrine, one which gives intuitive verdicts in cases of risk and uncertainty, and which does indeed undercut the Paralysis Argument's central conclusion.

3.1 The Doctrine of Doing and Allowing

Many non-consequentialists endorse an intuitively appealing distinction between moral reasons for doing versus allowing harm.[1] Call this distinction the *Doctrine*

[1]The canonical discussion is Philippa Foot, "The Problem of Abortion and the Doctrine of Double-Effect," in *Virtues and Vices* (Oxford: Basil Blackwell, 1978), 26. See also Jonathan Bennett, *The Act Itself* (Oxford: Oxford University Press, 1995); Alan Donagan, *The Theory of Morality* (Chicago: The University of Chicago Press, 1977); Frances Kamm, *Morality, Mortality, II: Rights, Duties, and Status* (Oxford: Oxford University Press, 1996); Warren Quinn, "Actions, Intentions, and Consequences: The Doctrine of Doing

of Doing and Allowing (DDA). Under the DDA, all else being equal, the *pro tanto* reasons you have against doing some harm x are stronger than the *pro tanto* reasons you have against allowing some harm x. My reasons against breaking your foot are, all else being equal, of a weightier sort than my reasons against merely allowing you to suffer a broken foot as a result of natural causes or as a result of the active intervention of another.

The DDA is appealing insofar as it promises to explain and justify our intuitions in a wide range of cases – perhaps most famously, in competing interests cases of the following sort:

> **Hospital I.** You can save the lives of five patients. In order to do so, you must (painlessly) kill a stranger and harvest five of their organs.
>
> **Hospital II.** You can save the lives of five patients. In order to do so, you must delay tending to a sixth critically injured patient. This delay will result in their (painless) death.[2]

Intuitively, in **Hospital I**, it is impermissible to save the five at the expense of the one. Intuitively, in **Hospital II**, it might still be permissible to save the five at the expense of the one. How are these differing intuitions best explained? One popular answer is found in the DDA. Since our reasons against doing harm are of a much weightier sort than our reasons against merely allowing harm, it might be permissible to allow a death in order to save the five, but it is impermissible to *do* a death in order to save the five.

Here is one clarification that will become important shortly. In their discussion of the doing versus allowing distinction, Andreas Mogensen & William MacAskill suggest that we reject a possible corollary of DDA: the further suggestion that our *pro tanto* reasons in favour of *providing* some benefit x to another are greater than our *pro tanto* reasons in favour of merely *allowing* some benefit x to befall another. This

and Allowing," *The Philosophical Review* 98, no.3 (1989):287–312; and Fiona Woollard, *Doing and Allowing Harm* (Oxford: Oxford University Press, 2015).

[2]Foot, 27.

is because this corollary – call it the *Inverse Doctrine of Doing and Allowing* (IDDA) – would seem to give intuitively implausible verdicts in cases of the following sort:

> **Hospital III.** You could drastically improve a patient's vision by using on them a particular vial of medicine. Your colleague, who works in a different ward, could drastically improve the vision of *five* similar patients given that same vial of medicine. Unfortunately, you do not have access to her ward. Neither of you have special relationships or special duties of care to patients in either ward.[3]

If the IDDA were true, then you might plausibly have greater moral reasons in favour of actively benefitting the one – depending, perhaps, on the exact numbers involved in the case. But this result seems wrong: a benefit is a benefit, and if your colleague really *could* benefit many more patients with the vial, then it seems like giving her the vial is the thing you ought to do. The fact that *you* would benefit a single patient under the alternative option does not seem to change this fact. A case like **Hospital III** draws our attention, then, to an interesting asymmetry. Our reasons against doing harm seem much stronger than our reasons against merely allowing harm, but our reasons for providing benefit are not obviously stronger than our reasons for merely allowing benefit.

3.2 The Paralysis Argument

The Paralysis Argument, a novel objection defended by Mogensen & MacAskill, attempts to force non-consequentialist proponents of the DDA into absurdity. More precisely, the Paralysis Argument attempts to force non-consequentialist proponents of the DDA into the following conclusion: we have a preponderance of moral reasons in favour of doing as little as possible, where "doing as little as possible" consists of voluntarily entering a state of paralysis. The point of the Paralysis Argument is not to

[3] Andreas Mogensen & William MacAskill, "The Paralysis Argument," *Philosophers' Imprint*, 21, no.15 (2021):1–17, esp. 5. I reframe the case in terms of more explicit benefits, following Charlotte Franziska Unruh in "Doing and Allowing Good," *Analysis*, (2022):1–9.

show that we *should* voluntarily enter a state of paralysis. The point is rather to show that a traditional non-consequentialist distinction like the DDA will have to go.

My aim in this section is to sketch the details of the Paralysis Argument. Before sketching these details, though, we must first set out some of the framing assumptions on which the Paralysis Argument rests.

Let's first assume that the subjective choiceworthiness of options for non- consequentialists is determined in the following way. One has subjective reasons to perform an option when that option fares *well in expectation*.[4] In order to figure out how well an option fares in expectation, do the following. First, list the possible outcomes associated with that option. Second, attach probabilities to each possible outcome associated with that option.[5] Third, rank these possible outcomes in terms of their interval-scale measurable *objective choice-worthiness*: the objective choice-worthiness of an outcome is a function of all the morally relevant features that would obtain if that outcome were the case, whether "the breaking of a promise, the intending of a harm, or the using of a victim as a mere means."[6] Fourth, take every possible outcome, weight it by both its probability and its objective choiceworthiness, and sum these values. This gives the value of the option in expectation. The degree to which an option fares *well* in expectation will be a comparative matter: it will depend on the expected value of the other options within the choice set.

That, then, is a brief sketch of a framework for deontological decision making under uncertainty. I've deliberately left certain details blank – for instance, the exact permissibility rules that might govern a set of options with varying degrees of subjective choiceworthiness – because these further details don't matter for our purposes here.[7] What crucially matters for our purposes here is that under a framework like the one just sketched, the subjective choiceworthiness of an option will be informed by factors

[4]The approach sketched here is given an extended rationale in Seth Lazar, "In Dubious Battle: Uncertainty and the Ethics of Killing," *Philosophical Studies*, 175 (2018):859–883. It is also the approach assumed in Mogensen & MacAskill, 5.

[5]I'll assume here that the probabilities are epistemic ones; they represent the levels of credence that rational agents ought to hold in an outcome obtaining, given their available evidence.

[6]Mogensen & MacAskill, 5

[7]For instance: I don't want to preclude satisficing views by assuming that agents ought always choose the option with the greatest degree of subjective choiceworthiness.

including, for instance, the extent to which that option *risks* doing or allowing harm to others.

Second, the Paralysis Argument draws on the empirical assumption that the indirect harms and benefits brought about by our actions are likely to be many indeed. This empirical assumption is plausible given the *identity-affecting nature* of our acts.[8] Even the most seemingly trivial of our acts influence the exact moment and manner in which others procreate; in doing so, however, these seemingly trivial acts inadvertantly influence the identities of future persons. Insofar as you influence a future person's identity, you also influence the various decisions they make and the various behaviours they exhibit over the course of their lifetime. But such effects ramify over time: by changing a future person's decisions and behaviours in this way, you will in turn influence the exact identities of *further* future persons who are yet to be born. It is plausible, then, that before very long at all the indirect harms and benefits caused by your choices are going to be many indeed – likely to vastly outnumber, for instance, any directly foreseeable harms and benefits that might be at stake in your choice.[9]

With these two assumptions stated, we are now placed to consider the details of the Paralysis Argument. Imagine, then, some set H_1–H_n, which represents the set of all possible unforeseeable and indirect future harms that might follow any given one of your acts. A particular harm H_1 might consist of, for instance, a distant future commuter named Eddy dying prematurely in a terrible car crash. A particular harm H_2 might consist of a distant future person named Sarah suffering a broken arm in a freak tornado. And so on. For any particular harm within the set – call it H_i – we don't have any relevant evidence that H_i is going to follow any given one of your options rather than another. So, you should plausibly think H_i just as likely to come about given the performance of any single one of your available options as another.

Now imagine that you have two options, A and D. If you perform A, and any particular indirect harm H_i comes about, you will count as merely having *allowed* that

[8]See James Lenman, "Consequentialism and Cluelessness," *Philosophy and Public Affairs*, 29, no.4 (2000):342–370; Derek Parfit, *Reasons and Persons* (Oxford University Press: Oxford, 1984), 352.

[9]Mogensen & MacAskill, 6.

harm to befall the one on whom it falls. If you perform D, and any particular indirect harm H_i comes about, you will count as having *done* that harm to the one on whom it falls.

This does not bode at all well for option D. Why? Well, we have established that an option which risks *doing* some harm H_i fares worse in expectation, all else being equal, than an option which risks to the same degree merely *allowing* that harm. Insofar as D risks doing H_i, but A risks merely allowing it, D is going to fare worse in expectation than A. Insofar as D risks doing H_{i+1}, but A risks merely allowing it, D is going to fare worse in expectation than A. Insofar as D risks doing H_{i+2}, but A risks merely allowing it, D is going to fare worse in expectation than A ... and so on, iterating to your heart's content for the enormously large set of possible indirect harms $H_1 - H_n$. It seems, then, that you are going to end up with far stronger subjective reasons against the performance of D than you have against the performance of A.

Mogensen & MacAskill take an action like A to involve voluntarily entering a state of paralysis such that your behaviour 'does' as little as possible as regards the passage of future history; by contrast, an action like D would involve active *doings* of the sort with which we are all familiar – writing philosophy papers, going for runs, boiling rice, brushing teeth, and so on. If this is how we conceive of the A and the D options, then it seems we are going to have far stronger subjective reasons in favour of voluntarily entering a state of paralysis than in favour of ever doing anything at all.

That, then, is the core of the Paralysis Argument. Here are a couple of initial clarifications, which I'll couch in the form of objections. One might initially object to the Paralysis Argument in the following way: "But surely, there is some set of *positive benefits* (call them $P_1 - P_n$) which might come about in distant future history as a result of each of your available options. In the absence of any relevant evidence, you should have equal credence that any member of the set $P_1 - P_n$ will follow either one of your options. If you perform D, then you will count as having *done* these positive benefits; if you perform A, then you will count as having merely *allowed* these positive benefits. This is going to improve the expected value of D compared to that of A. (It is better, after

all, to *do* (chance doing) benefit to others rather than merely *allow* (chance allowing) benefit to others.) Given all this, you might not in the end have greater subjective reasons in favour of performing *A simpliciter.*"

This response fails, however, if we reject the IDDA, as do Mogensen & MacAskill, and as seems plausible given **Hospital III**. As a general rule, all else being equal, it seems that our reasons in favour of *providing* a benefit (of some magnitude x) to another are not greater than our reasons in favour of *allowing* a benefit (of some magnitude x) to another. If there *is* a reasons asymmetry between these two possible ways of benefitting, it is doubtful that this asymmetry is as marked or as weighty as the asymmetry between our reasons against doing harm and our reasons against allowing harm. Given this fact, I'll put this concern to one side in the following sections.

Another initial objection to the argument might go as follows: "You said that A is going to have greater value in expectation than D, given that D risks *doing* the various future harms $H_1 - H_n$. But in coming to this verdict, you did not even consider the direct and immediately foreseeable consequences at stake in the decision making process. If you *had* considered these direct and immediately foreseeable consequences, then the jury might still be out as to whether A has a higher expected value than D *simpliciter*."

But we can set aside this objection, too, for reasons mentioned previously. The set of indirect harms that might follow either option A or D is going to be extremely large, given plausible assumptions about the length and size of the long-run future. Since you risk *doing* every member of this (extremely large) set of harms under D, but since you only risk *allowing* every member of this (extremely large) set of harms under A, this is going to skew the expected value (and hence the subjective choiceworthiness) of the two options extremely heavily in favour of A. So unless the immediately foreseeable stakes are *extremely* high, it is implausible that the immediately foreseeable consequences are likely to lead us to revise our initial verdict about the comparative choiceworthiness of A and D. For this reason I set aside this objection, too, in the following sections.[10]

[10] I simply grant the point in this paragraph for the sake of argument. For a statement of the reasoning with numbers, see Hillary Greaves & William MacAskill, "The Case for Strong Longtermism," *Global*

That, then, is the Paralysis Argument. I don't wish to imply here that I agree with every element of its framing. One worry about which I have written elsewhere is whether a simple distinction between the foreseeable versus unforeseeable consequences of our acts can adequately capture our evidential situation as regards the long-run future. And the Paralysis Argument seems to make implicit use of just such a distinction between foreseeable and unforeseeable consequences.[11] Another general concern to which I am sympathetic is whether we can sensibly speak of our actions as *harming* individual future persons whose existence is contingent on our choice. If you think that harm is an inherently comparative concept, for instance, then it is not immediately obvious that we can.[12] But I'll deliberately put aside these more general concerns in the following sections; doing so will allow me to develop the strongest possible response to the Paralysis Argument I can, and to focus on certain novel issues for the doing versus allowing distinction.

3.3 Responding to Paralysis

3.3.1

In the rest of this paper, I'll consider three possible responses to the Paralysis Argument. I'll take the first two responses to be intuitive but ultimately unsuccessful. The third response involves investigating whether we might make plausible alterations to the DDA such that it gives more intuitive responses in cases of risk and uncertainty, and such that we can avoid the central conclusion of the Paralysis Argument.

Here is a first-pass attempt at spelling out why the Paralysis Argument doesn't demonstrate its central conclusion. You might think that, as a general rule, if some harm x emerges as the result of chancy or chaotic processes in which you merely play

Priorities Institute Working Paper Series, 5 (2021):1–43.

[11]See "Cluelessness Redux." One riddle in particular for the Paralysis Argument is whether there *really is* any indirect harm H_i which we should think equally likely to follow any of our options. If one of my options is certain to imminently end the world by initiating an extinction event, then the set of indirect future harms that might come about given my choice would seem to be empty. This riddle comes about because the size of society's future is contingent on our choices.

[12]Mogensen & MacAskill respond to non-identity considerations on 10.

some contributory role, then you do not count as *doing* that harm x.

This initial hypothesis would seem to undercut the Paralysis Argument in the following way. Suppose that each of the indirect harms $H_1 - H_n$ can be properly understood as the outputs of chancy or chaotic systems in which you merely play some contributory role. If this is the case then, regardless of whichever course of action you choose, you will not count as *doing* these harms $H_1 - H_n$. The possible harms $H_1 - H_n$ are thus not going to furnish you with subjective reasons in favour of some courses of action rather than others: in favour, say, of voluntary paralysis over active behaviour.

This initial hypothesis has some intuitive appeal. (Do I really *do* the harms associated with a distant future tornado by turning on my fan, thus swirling the air currents in just the wrong way?) In this initial form, however, the response is untenable. The mere fact that a harm x comes about via chancy or chaotic processes in which you play some contributory role doesn't seem to imply, as a general rule, that you can't count as a *doer* of that harm x. Take the following case:

> **Rockets.** An evil villain releases several thousand explosive rockets into the atmosphere. The rockets travel on chaotic and chancy trajectories, but some of them land in urban areas, killing millions of people.

It is fairly intuitive to think that the villain *does* harm in **Rockets.** The fact that these harms emerge via chancy or chaotic processes doesn't seem to imply that he can't count as a doer of them. Similarly, imagine that you hold a gun to my head with only one of its many barrels containing a bullet. You program a chancy randomiser to select and fire a barrel, but the randomiser *just so happens* to select the barrel with a bullet. Surely, you should still count as *doing* me harm in the special case that I collapse to the floor with a bullet in my head.

To get around **Rockets**, then, the hypothesis needs to be further refined. Let's stipulate that you might still qualify as a doer of some harm x if you *initiate* a chaotic or chancy process which results in x coming about. But let's say that you fail to qualify as a doer of some harm x if you merely play some *other* contributory role in those chaotic

or chancy processes which result in x coming about. If this is the view, then we can still affirm that the evil villain does harm in **Rocket**, as seems intuitive. But we might still be able to say, as we originally wanted to, that regardless of whichever course of action you choose you don't qualify as a *doer* of the indirect future harms $H_1 - H_n$. This further result would hold so long as you didn't count as initiating the chaotic or chancy processes which led to those indirect future harms $H_1 - H_n$.

A view like this fits neatly with Philippa Foot's canonical analysis of the doing versus allowing distinction. Under Foot's analysis, the most standard way in which you *do* a harm is to initiate a process or sequence of processes which later results in a harm event.[13] By contrast, you merely *allow* some harm if you let a sufficiently independent process run its course, where that sufficiently independent process eventually brings about a particular harm event. If this is what it takes to do a harm, then it seems like merely playing some contributory role in the ongoing development of chaotic or chancy processes which themselves lead to harm events $H_1 - H_n$ would not necessarily qualify you as a doer of those harms.

A lingering issue for this line of response is what it would mean, in general terms, for someone to play a contributory but non-initiating role in chaotic or chancy processes which lead to harm events. This suggestion remains a little mysterious. I suppose I have in mind cases of the following sort. Suppose that a terrible tornado T comes about several centuries from now in virtue of the earth's chaotic atmospheric systems evolving in *just* the wrong way. In a case like this, many millions of agents might well each play some critical contributory role in developing the atmospheric system in just the wrong way such that T later obtains. It is hard, however, to think of any of these individual agents (putting aside perhaps the first) as the *initiators* of a process which leads to the harm event T. Take myself, for instance. Although it might certainly be true that by boiling rice for dinner I stir the air currents in just the wrong way such that T eventually comes about, it is hard to think of me as initiating the process leading to T. This is not least because whether or not T later obtains will depend on the activity

[13] Philippa Foot, "Killing and letting die," in *Killing and Letting Die 2nd Edition*, eds. Bonnie Steinbock and Alastair Norcross (New York: Fordham University Press, 1994), 280-289.

of millions of other agents *vis-à-vis* the atmospheric system, many of whom lived and died and enacted the relevant behaviours long before I was ever born.

Let us recap the reasoning of this section so far. I began by considering the plausible intuition that somehow, when a harm emerges via chaotic and chancy processes, I cannot qualify as a *doer* of that harm. This initial view quickly failed: in a case like **Rockets**, the villain seems to do harm even though the harm comes about via chaotic and chancy processes. The hypothesis would have to be refined, then, in the following way: I cannot qualify as doing some harm x when I merely play some *contributory (but non-initiating) role* in the chaotic and chancy processes which bring about x. Under this refined version of the hypothesis the villain counts as doing harm in **Rockets**, but plausibly, you fail to qualify as a *doer* of the various future indirect harms $H_1 - H_n$. And this, plausibly, is because you do not initiate the processes which lead to those harms $H_1 - H_n$.

This refined version of the hypothesis would still, in the end, undercut the Paralysis Argument in the following way. Imagine you have two courses of action: voluntary paralysis or active rice boiling. Each of these courses of action might certainly bring about future indirect harms $H_1 - H_n$. But regardless of which particular course of action you choose, it seems that you will fail to qualify as *doing* any of the harms $H_1 - H_n$ in the event that they eventually come about. This is because you will fail to qualify as initiating the chaotic or chancy processes which result in those harms. Hence, it will not be as if one of your acts *does* the harms $H_1 - H_n$ while the other act merely allows them. Hence, you won't have a preponderance of subjective reasons in favour of voluntary paralysis.

In the end, I am still agnostic about this more refined version of the hypothesis. Part of my agnosticism stems, I suspect, from the fact that there is a certain level of vagueness in talk of 'initiating processes.' In any case, it seems that this refined hypothesis still gives the wrong answer in particular cases. Imagine, for instance, that the villain from **Rockets** (with whom you happen to be friends) has invented a horrific contraption which will release devastating tornadoes upon the surface of

the earth – the tornadoes will cause death and destruction by travelling on chancy or chaotic trajectories. Suppose that the villain turns on the contraption and that it begins to produce tornadoes. Suppose too that, in order to keep the contraption and the tornadoes going, many buttons need to be held down all at once. The villain pressed down all of the buttons when he turned on the machine, but you later take over some of the buttons in order to give his fingers a rest. Hence, the tornadoes keep travelling and causing death and destruction rather than fizzling out to mere gusts of breeze. This is a case in which it is hard to think of you as *initiating* a chancy and chaotic process which results in harm. You seem to play, rather, some crucial contributory role in the chaotic and chancy processes initiated by another which themselves lead to harm events. But it still seems plausible to say that this is a case in which you *do* harm to those who suffer as a result of your actions. A verdict like this would be prohibited by the hypothesis at which we just arrived.

What we may want to defend, of course, is a range of more modest claims along the following lines. If a harm comes about via chaotic or chancy processes in which you play some contributory role, then we might not necessarily *blame* you for that harm, even if you qualify as a doer of it. We would not blame you for boiling your rice and thus inadvertantly bringing about a Texan tornado in the distant future. This is because the Texan tornado was unforeseen, unpredictable, and unintentional. And whether or not we blame someone for doing a harm is typically going to depend on facts such as whether or not that harm was unforeseen, unpredictable, or unintentional. Similarly, there is a long tradition of legal thought which says that we would not find you *legally responsible* for certain harms that came about via chaotic or chancy processes, if such processes somehow severed the link between the exercise of your agency and the harm event in question; otherwise, we would risk having to hold Adam and Eve legally responsible for my spilt soup, as well as for everything else that has ever gone wrong in the world.[14]

But none of these further (and very reasonable) points undercut the Paralysis Ar-

[14]See, for instance, John Gardner, "Complicity and Causality," *Criminal Law and Philosophy* 1 (2007): 127–141.

gument. The key upshot of the Paralysis Argument was, recall, that you have a preponderance of subjective reasons in favour of voluntary immobility, assuming the DDA, assuming a deontological decision theory of decision making under risk, and granting certain reasonable empirical assumptions about the way in which our actions proliferate indirect harms over time. Such an argument is not threatened by the reasonable acknowledgement that we need not blame me nor hold me legally responsible for some particular future indirect harm event H_i that comes about given the performance of my act.

3.3.2

Here is a second-pass attempt at explaining where the Paralysis Argument goes wrong. Note that each of the indirect and distant future harms $H_1 - H_n$ are *freakish.* They are freakish in the following sense. First, you don't have any particular reason for thinking that any particular harm H_i will come about given the performance of any particular one of your acts. But second, given this fact, you would be rational in thinking it exceptionally unlikely that any given harm H_i *would* come about given the performance of any particular one of your acts. (How likely should I think it that, by grilling this cheese sandwich, I will cause an unimaginably horrific tornado measuring 5 on the Fujita Scale to tear through Memphis at 3PM on a Tuesday four hundred years from now?)

Now, you might think that, at the moment of choice, it can be instrumentally rational to put aside certain possible outcomes which one takes to be exceptionally unlikely. (I can put aside the outcome in which I have credence 0.000000001, for instance, that my having this grilled cheese sandwich will cause an unimaginably horrific tornado measuring 5 on the Fujita Scale to tear through Memphis at 3PM on a Tuesday four hundred years from now.) You might particularly think that this is the case if accounting for such exceptionally unlikely outcomes would in the end demand that you pursue intuitively unpalatable courses of action: if accounting for such exceptionally unlikely outcomes, for instance, would force you to abandon certain

otherwise sure-thing goods at stake in the decision-making process.

If this is the case, then you might think it can be instrumentally rational, at the moment of choice, to put aside certain freakishly unlikely outcomes O_i containing certain freakishly unlikely harms H_i. The extremely unlikely outcomes containing such harms can be ignored by the instrumentally rational. Under a picture like this, an act which risks *doing* the exceptionally unlikely freak harm H_i will not fare worse in terms of expected value, all else being equal, than an act which risks merely *allowing* the exceptionally unlikely freak harm H_i. And the same result might hold for each individual freak harm in the set $H_1 - H_n$.

Of course, a proposal along these lines deviates from the textbook decision-theoretic approach to choice under risk we outlined previously. And a proposal along these lines might lead us into a more general discussion on whether normative decision theory ought to endorse *fanatical verdicts*: whether extremely low probability payoffs should sometimes be able to determine comparative betterness across the options in a way which seems intuitively unpalatable.[15] But I suspect, however, that we need not enter these broader discussions here. The response outlined in this section seems to fail on simpler grounds.

Suppose that the villain in **Rockets** releases only *one* explosive rocket. He might then use the response developed in this section to defend his conduct in the following way: "Admittedly, I am releasing a risky explosive rocket. But the rocket will travel on a chaotic and chancy trajectory. The chances that it lands on some *particular* person (a farmer in North Dakota called Adeline, say) is exceptionally low. And so I can put aside this possible outcome when it comes to the instrumentally rational calculation of expected value. The same goes for any *particular* person on whom the rocket might happen to fall. And so, in fact, my releasing the rocket doesn't fare worse than any of my other options (sitting on the sofa, say) in terms of expected value!"

Such a response is clearly inadequate. Something has gone wrong in the villain's reasoning. I take him to make the following mistake. Although, for instance, it might

[15] See recently, for instance, Hayden Wilkinson, "In Defense of Fanaticism," *Ethics*, 132, no.2 (2022):445–477.

be exceptionally unlikely that some *particular* person would die as a result of his rocket – say, a North Dakotan farmer named Adeline, a New York banker named Sarah, or a Californian orange farmer named Robert – it is still fairly likely that *someone* is going to die as a result of the villain's rocket. This fact is clearly relevant when it comes to the calculation of expected value. And this indicates to us that this is the relevant level of grain with which the outcomes need to be described. Or at least, that we cannot describe the outcomes in a level of grain which would *erase* the clear significance of this fact.

We can make the same point, I suspect, when it comes to the freakish set of harms $H_1 - H_n$. Although it is exceptionally unlikely, at the moment of choice, that your grilling a cheese sandwich will cause a *particular* tornado measuring 5 on the Fujita Scale to zoom through Memphis at 3PM on a Tuesday afternoon four-hundred years from now, you might still think it fairly likely that your act will, say, *indirectly bring about future tornadoes in general* or *indirectly bring about future harms in general via natural disasters*. And you might think that we can't describe the outcomes in a level of grain such that we would erase the clear significance of this fact for moral decision making.

Then, though, we are back to where we started. An act like voluntary paralysis is going to risk *allowing* tornadoes or natural disasters in general. And an active behaviour like boiling rice is going to risk *doing* tornadoes or natural disasters in general. These risks plausibly need to be taken into account in the calculation of expected value. And so, given the DDA, all else being equal, you may still end up having greater subjective reasons against active doings than against voluntary paralysis.

3.4 From Risky Doings to the Doing of Risks

3.4.1

I've considered, now, what strike me as two intuitive responses to the Paralysis Argument: first, denying that you *do* a harm if that harm is best understood as the output of chaotic or chancy systems in which you merely play some contributory role, and

second, denying that you need to take any particular freakish harm H_i into account in being an instrumentally rational calculator of subjective value. I've regrettably found both of these intuitive responses wanting. I haven't said that either of these responses are indefensible. But I've sketched what seem to me to be some of the most serious problems for those who wish to go down either route. In the final sections of the paper, I'll articulate and defend a fairly different style of response to the Paralysis Argument.

Here is one possible thought. You might think that what really matters, in a case like **Rockets**, is whether the villain imposes an *additional risk of harm* upon any particular individual in acting the way he does. You might think that we have a *pro tanto* reason against performing an act which imposes an additional risk of harm on some individual. And you might think that *this* is what we need to focus our attention on – that a view like this is the natural successor to the DDA, capable of giving more appropriate verdicts in cases of risk and uncertainty.

Let's be as precise about this idea as we can. Whether or not my act imposes an additional risk of harm upon some individual A is a matter of whether or not my act *worsens A's prospects* compared to their prospects under some specified baseline. To calculate A's prospects under a given act, do the following: ask how well-off A would be in each possible outcome associated with that act, weight A's well-being in each possible outcome by the chance of that outcome obtaining, and then sum these values. The word 'chance', here, should be given an epistemic gloss: the chance of an outcome obtaining given my act is the chance that *I as a decision maker* would reasonably assign to that outcome obtaining, in light of my available evidence and at the moment of choice.[16]

What is the relevant baseline against which my risk imposition upon A is to be measured? For our purposes, let's simply assume that the baseline consists of the prospects that A *would* have faced, given voluntary immobility on my part. In figuring

[16]The approach to risk imposition sketched here broadly follows Christian Barry & Garrett Cullity, "Offsetting and Risk Imposition," *Ethics*, 132, no.3 (2022):352–381; also see their "Do We Impose Undue Risks When We Emit and Offset? A Reply to Stefansson," *Ethics, Policy, & Environment* (forthcoming), for a defence of the claim that the relevant probabilities in risk imposition are the epistemically rational ones.

out whether the villain imposes an additional risk of harm upon anybody in **Rockets**, for instance, we must ask whether he gives any individual person worse odds than they would have otherwise have had if he had stayed immobile at home. Clearly, the villain does. And so, he imposes upon them an additional risk.

3.4.2

That, then, is a first pass replacement for the DDA which gives more plausible cases involving risky and uncertain harms: you have a *pro tanto* reason against the performance of some act if that act imposes additional risk upon another, relative to some specified baseline involving immobility on your part.

Now, we were initially attracted to the DDA because it gave intuitive and powerful verdicts in cases such as **Hospital I** and **Hospital II**. Because of the DDA we could uphold the intuitively correct verdict that your reasons against doing harm in **Hospital I** were of a weightier sort than your reasons against merely allowing harm in **Hospital II**. If we replace the DDA with a rule about us having *pro tanto* reasons against additional risk imposition, can we still uphold an intuitive verdict along these lines?

The answer is probably yes – if, as I have suggested, the baseline involves immobility on your part. Here is why. In **Hospital I** your decision to kill a stranger and harvest their organs would impose an additional risk of harm upon them. (If you had stayed immobile instead of chopping up this stranger, they would have had far better prospects.) In **Hospital II**, your decision to abandon the sixth patient in the waiting room plausibly *wouldn't* impose an additional risk of harm upon them. (If you had stayed immobile instead of treating the first five patients, the sixth would still have died.) Killing the stranger in **Hospital I** would thus involve the imposition of additional risk upon a particular individual, but abandoning the sixth patient in **Hospital II** would not involve the imposition of additional risk upon any particular individual. This gives you a *pro tanto* reason against killing in **Hospital I** not present in your choice in **Hospital II**. We can still preserve the intuitively correct verdict that there is a reasons asymmetry between these two cases.

Importantly for our purposes, if we replace the DDA with a rule about additional risk imposition along these lines, then the central conclusion of the Paralysis Argument is avoided. We need not fear having a preponderance of subjective reasons against ever doing anything at all other than voluntary paralysis. Why? Well, take some piece of active behaviour (my walking in the park, perhaps). My walking in the park won't impose *additional risks* upon any particular future person compared to a baseline in which I enter a state of voluntary paralysis. This is plausibly true because I do not have any relevant evidence, at the moment of choice, that any given future harm H_i is more likely to come about given walking rather than given immobility. Hence, I am not going to have a *pro tanto* reason against my walking in the park as opposed to voluntary immobility. I will have no such *pro tanto* reasons in virtue of any indirect distant future harms $H_1 - H_n$ that might follow my act.

3.4.3

This thesis about additional risk imposition, then, gives the intuitively correct answer when applied to those cases which motivated the DDA, but also escapes the central conclusion of the Paralysis Argument. Let's say, then, that we respond to the Paralysis Argument by replacing the DDA with just such a view about us having *pro tanto* reasons against additional risk imposition.

An immediate difficulty for a view along these lines stems from cases involving overdetermination. Mogensen & MacAskill provide a case which will help us to motivate the point. Consider:

> **Arms Trader**. You are approached with the opportunity to sell a large volume of weaponry to a brutal dictatorship, foreseeing that the weapons will be used to oppress and murder innocent civilians. You also know that if you do not make the sale, the dictator will just go to one of your less scrupulous competitors and purchase the arms they want from them.[17]

[17] Mogensen & MacAskill, 9.

This is a case in which you could do certain harms (let's say you would do them 'indirectly') by selling arms to the dictator. The alternative is to allow those same harms to come about by letting a *different* arms trader do business with the dictator. Intuitively, your reasons against doing harm in this case (by selling the weapons) are stronger than your reasons against merely allowing those same harms to come about (by letting others sell weapons). This is the sort of canonical judgement that the doing versus allowing distinction originally hoped to accommodate.

But it seems that the view at which we just arrived, about the special badness of additional risk imposition, cannot accommodate this intuitive verdict. Why? Well, it turns out that by selling weaponry in **Arms Trader** you don't impose an additional risk of harm upon anybody. This is because **Arms Trader** is a case of overdetermination. If you had entered a state of voluntary immobility instead of selling the guns, then *somebody else* would still have sold the guns anyway. Thus, by selling the guns you won't impose additional risks upon anyone. We can't use our rule about additional risk imposition to justify the claim that you have a special *pro tanto* reasons against selling (rather than letting others sell) the weapons.

Unfortunately, however, I have nothing particularly new to say as regards the overdetermination problem. I think we are simply best placed to follow elements of the framework provided in Christian Barry & Garrett Cullity's recent analysis of risk imposition.[18] That framework goes as follows. In figuring out whether I increase a population's prospect of harm in a risk-imposing way by my action A, where my action A is preemptive to someone else's action B, the relevant question is whether I increase the prospect of harm for that population in comparison with the *attributional baseline prospect of harm*. The attributional baseline prospect of harm is the prospect of harm that would have been faced by that population assuming the absence of the performance of either A or B. We can suppose that, as before, the attributional baseline prospect of harm assumes voluntary immobility on my part.

If we adopt the attributional baseline, then we can give the intuitively correct

[18] Barry & Cullity, "Offsetting," 367.

answer in **Arms Trader**. We can now say: you impose an additional risk of harm upon the civilian population by selling the weaponry to the dictator, relative to the attributional baseline prospect of harm for that civilian population. After all, the civilian population would have had much better prospects if neither you nor your competitors had done business with the dictator. Since you have a *pro tanto* reasons against imposing additional harm in this way, you have a *pro tanto* reason against selling those very weapons.

3.4.4

I've deliberately put aside, till now, a further and more troubling problem for the risk imposition view. It goes as follows. Sometimes, there are multiple ways in which you might impose the same additional risk of harm upon a specified population, where *one* way of imposing additional risk seems intuitively worse than the others. The view about additional risk imposition at which we have just arrived, it seems, will be unable to affirm as much.

It is easiest, here, to simply demonstrate with the following case:

> **Hospital IV**. You could kill a stranger in the waiting room and harvest their organs for medical research. Or you could allow your colleague to do the same deed, by you yourself going for an especially long walk on your lunch break. You could, of course, enter a state of voluntary paralysis right where you are. But if you enter a state of voluntary paralysis right where you are, your colleague will not proceed with the killing for fear of being watched.

In **Hospital IV**, it seems intuitively worse for *you yourself* to actively kill the stranger in the waiting room. It seems, intuitively, that you have greater reasons against doing harm in this way than you have against, say, merely allowing your violent colleague to do harm in this way. At least, I take it that this is the sort of canonical verdict that defenders of the traditional DDA would wish to accommodate.[19]

[19]Of course, this asymmetric verdict is still compatible with the claim that you have very strong reasons against allowing your colleague to act in this way in the first place.

But note that the view at which we've arrived (about you having special *pro tanto* reasons against imposing additional risk) can't accommodate this verdict. Why? Well, simply put, both of these harm-inducing options impose the same additional level of risk upon the stranger in the waiting room relative to the specified baseline of voluntary immobility on your part. You are going to have, then, an equally strong *pro tanto* reason against the performance of either act – against either doing harm to the patient yourself, or against allowing your colleague to do the harm.

We can further motivate this worry by drawing on more explicitly risky cases. Imagine that you are deciding whether or not to joyride in a gas-guzzler or to open the garage door such that your *friend* can go for that same joyride in that same gas-guzzler. There is also a third available option: a baseline of voluntary immobility on your part, under which the garage remains locked and under which neither you nor your friend go joyriding. By joyriding in the gas-guzzler, you impose a moderate additional risk of harm upon Betty, a passing pedestrian who might get hit. By opening the garage door and thus allowing your friend (an equally talented driver) to go for a joyride, you impose the *same* moderate additional risk of harm upon Betty, who might get hit.

What we might wish to say about such a case is that there is something especially bad about *you* going for a joyride and thus risking harm to Betty in this manner, whether or not that harm later obtains. You have greater reasons against going for a joyride in this manner, for instance, than you do against merely opening the garage door such that your friend can joyride. But as before, this is a verdict which the previous view about additional risk impositions cannot accommodate. Whether you yourself joyride, or whether you merely allow your friend to joyride, you impose the same additional risk of harm upon Betty relative to a specified baseline of voluntary immobility on your part. So you are going to have an equally strong *pro tanto* reason against the performance of either act.[20]

[20]Note that neither of these cases involve any overdetermination: it is not as if your friend will go on a joyride if you decide not to. And to be clear: your friend can only drive if you engage in the active behaviour of opening the garage door. It does not much matter for our purposes here whether we think of this door opening as an *allowing* or as a kind of *enabling*.

3.4.5

What we need, then, is some way of upgrading the previous risk imposition view at which we arrived such that it can give the right answers in **Hospital IV** and the joyriding case. At this point, it is tempting to simply say something along the following lines: "A case like **Hospital IV** shows that you must have greater reasons against *you yourself* imposing an additional risk of harm upon a particular individual than you have against merely allowing somebody else to impose that same additional risk of harm upon that same individual." But we will end up in confusion if we speak in this way. In **Hospital IV**, for instance, you would impose an additional risk of harm upon the stranger by cutting them open, but *you too* would impose that same additional risk of harm upon the stranger by choosing the option which allowed your colleague to cut them open. This simply follows from the way in which we've defined the baseline: whether or not your act imposes additional risk upon a particular individual, after all, is a matter of whether or not that act worsens their prospects in comparison with a baseline of voluntary immobility on your part.

What we're *really* trying to figure out is how we might justify an intuitive difference in reasons between your imposing the same additional risk of harm in two different ways – imposing that same additional risk of harm directly, if you like, versus indirectly.

I suspect we must here say something along the following lines. As a moral decision maker you have a *pro tanto* reason against imposing some additional level of risk upon some particular individual. But not all risk impositions of a given level are equally bad. There is something *especially* bad about you imposing an additional level of risk upon another, where this additional level of risk is imposed because of the particular way in which you *and only you* have chosen to exercise your agency. The corollary of this claim is that it is not as bad, all else being equal, for you to impose some additional level of risk upon an individual when that additional level of risk came about through the exercise of *somebody else's agency too*. Call this final view the *Doctrine of Directly versus Indirectly Doing Risk Impositions* – the DIDRI, for short.

I leave it open, here, exactly what it means for an additional risk imposition to come about *only* because of the exercise of my agency, versus to come about through the exercise of somebody else's agency too. This is partly because I take such a distinction to be clear enough in cases like **Hospital IV**. But it is also because there are, I suspect, different live options for spelling this view out. One option would be a counterfactual account: somebody else's agency is relevant if the additional risk imposition wouldn't have come about save the exercise of their agency. Another option would be an explanatory account: somebody else's agency is relevant if it is the best (or serves as part of the best) *explanation* of that risk imposition coming about.

I suspect that something like the DIDRI is our best shot for preserving the core intuitions behind the DDA, while also giving the appropriate verdicts in cases of risk and uncertainty. The DIDRI gives, we have seen, the intuitively correct verdicts in **Hospital I**, **Hospital II**. But it also gives the correct verdict in **Hospital IV**. Why? Well, there would be something *especially* bad about you imposing a certain level of risk on the stranger in the ward by killing them with your own hands, since this would be to impose an additional risk through the exercise of your own agency. It is worse for you to impose this risk via the sole exercise of your own agency, rather than also via the agency of somebody else such as your colleague.

Assuming that *something* like the DIDRI is true, then we still need not fear the central conclusion of the Paralysis Argument. Indeed, these further questions about whether or not I have especially strong reasons against being a *sole* imposer of an additional risk of harm upon another need not be answered in order to see why. For any of the indirect and distant future harms $H_1 - H_n$ that might follow my acts, I do not impose an additional risk of those harms befalling any particular person by choosing, say, to walk rather than than enter a state of voluntary paralysis. Hence, I do not have *pro tanto* reasons against walking given purely the possibility of those harms $H_1 - H_n$.

3.5 Conclusion

Here, in brief, is a roadmap of where we have been. The Paralysis Argument posed a challenge for the traditional DDA: if the DDA were true, and if that principle were coupled with a plausible account of deontological decision making under uncertainty, then we would have a preponderance of subjective reasons in favour of never doing anything at all. I began by exploring two intuitive responses to the Paralysis Argument: the response that I do not *do* a harm if that harm is best understood as the output of a chaotic or chancy system in which I merely play some contributory (but non-initiating) role, and the response that I might rationally put aside the possibility of freak harms at the moment of choice. Neither of these responses struck me as obviously successful. But in this final section, I've considered whether we might modify or upgrade the traditional DDA such that it gives more plausible verdicts in cases of risk and uncertainty and such that we can avoid the central conclusion of the Paralysis Argument. I've suggested that we can replace the DDA with a view along the following lines: one has a *pro tanto* reason against imposing an additional risk of harm upon another, relative to a specified baseline involving voluntary immobility on your part. Such a view preserves the same verdicts in **Hospital I** and **Hospital II** that the DDA originally hoped to preserve. One lingering issue is that sometimes there are different *ways* in which one could impose an additional risk of harm upon another, where it still seems intuitive that one has stronger reasons against imposing risk of harm in one way rather than in another. This was the case, for instance, in **Hospital IV**. I take this to be the crucial puzzle for those who might wish to replace the traditional DDA with a view about the special badness of risk impositions. One possibility, I have suggested, is to say that not all risk impositions of a given level are equally bad: one can have weaker reasons against imposing risk upon an individual if that risk imposition hinges, at least partly, on the exercise of others' agencies. Such a view gives plausible verdicts in particular cases, and in the end, it escapes the Paralysis Argument's central conclusion.

Chapter 4

A Guide to Action-Guidingness Objections

In this chapter I investigate whether we are ever justified in rejecting particular moral principles on the grounds that they are insufficiently *action-guiding*. I first present a fairly standard interpretation of action-guidingness objections: on the standard interpretation, a particular moral principle is said to be false insofar as it is not *usable*. I highlight two problems for the standard interpretation: first that the usability of a moral principle is an agent-relative matter, and second that the usability of a moral principle with respect to a particular agent comes in degrees. In light of these two facts, I suggest that it is problematic to simply reject a principle as false on the grounds that it not usable *simpliciter*.

Having presented this critique of the standard interpretation, I disarm three positive arguments historically given for thinking that moral principles must be widely or universally usable. These three historical proposals – concerning the inherent concept or function of a moral theory, concerning autonomy, and concerning justice – would seem to imply that moral principles can be rejected as false on the grounds that they are unusable for some or for all. I reject each historical proposal in turn.

I conclude by presenting, drawing on the literature of cluelessness arguments against consequentialism as a case study, an alternative interpretation of action- guid-

ingness arguments. Under an alternative 'epistemic' reading, action-guidingness arguments simply show that familiar and popular moral principles, when applied to particular decision scenarios, force us to deny certain deeply intuitive judgements of comparative betterness across the options that we take ourselves to already know. This result, the suggestion goes, gives us a reason for rejecting those moral principles as false.

4.1 Against Action-Guidingness Complaints

4.1.1

To begin with, let's say that an agent *uses* a moral principle or rule P to regulate their conduct as regards an act x when (i) that agent chooses x out of a desire to conform to P, and when (ii) that agent chooses x with the belief that, in doing so, they do indeed conform to P.[1] Under this definition of 'use', whether or not I can use a principle P in regulating my conduct is going to hinge on whether or not I can form certain beliefs and desires concerning my conduct as it *relates* to the principle P.

Consider the following example. Call *objective consequentialism* the thesis that one act is better than another so long as it brings about the better actual consequences. Under the analysis of 'use' just given, I use objective consequentialism *vis-à-vis* my donations to charity so long as I donate to charity out of a desire to bring about the better actual consequences, and so long as I donate to charity with the belief that, in acting as I do, I really do bring about the better consequences.

Importantly, the definition of 'use' just given does not reference the further issue of whether, in the end, I actually *do* conform to the principle to which I would like to conform – in the aforementioned case of objective consequentialism, whether or not my donation to charity actually *does* bring about the better actual consequences. In

[1] See Holly M. Smith, *Making Morality Work* (Oxford: Oxford University Press, 2018), 16, as well as "Making Moral Decisions," *Noûs*, 22, no.1 (1988):89–108 and "Two-Tier Moral Codes," *Social Philosophy and Policy*, 7, no.1 (1989):112–132. This is, roughly, the sense of usability employed by R.M Hare in *Freedom and Reason* (Oxford: Clarendon Press, 1963), 31–33.

response to this issue, say that I use a principle or rule P in the *external* sense in the special case that the belief featuring in (ii) is true: that is, in the special case that my act *actually does* conform to the principle to which I would like to conform. In the previous philanthropic example, this would be the special case in which my donation to charity really does bring about the better actual consequences.[2]

4.1.2

Plausibly, when one rejects a moral principle P on the grounds that it is insufficiently action-guiding, one is saying something like the following: the principle P must be false since it is not *usable*. For a moral principle P to be usable is just for agents to be able to use that principle in the sense just described: agents are able to act out of a desire to conform to P, and can act with the belief that they do indeed conform to P. (I'll assume, for now, that we aren't talking about usability in the external sense. It would be a bolder argument to insist that true moral principles must be externally action-guiding.)

Let us illustrate with a brief example. Imagine that by performing some acts rather than others, one brings about the intrinsically valuable property of *enchantedness*. Unfortunately, however, it is utterly impossible to tell at the moment of choice which acts are likely to promote the property of enchantedness and which are not. A moral principle, the *Rule of Enchantedness*, might say that you are always required to bring about the greatest possible quantity of enchantedness in any given choice scenario. But we might reject this principle as false given that it is not usable: agents cannot act out of a desire to conform to the *Rule of Enchantedness* with any sort of justified belief that, in choosing some courses of action rather than others, they *do* indeed conform to the *Rule of Enchantedness*.

Action-guidingness objections of this form appear fairly regularly throughout the normative ethics literature, especially in discussions concerning consequentialist accounts of right action.[3] Here, however, are two immediate complications for this style

[2] Smith, *Making Morality*, 13.

[3] In particular, see Shelly Kagan, *Normative Ethics* (Boulder: Westview Press, 1998), 64; James Lenman,

of argument.

The first point of complication is that the usability of a moral principle appears to be an agent-relative matter. It is perfectly coherent to claim, for instance, that some given moral principle *P* is usable for an ideally rational agent, extremely competent in their skills of moral and practical reasoning, while that very same principle *P* fails to be usable for another non-ideally rational agent who lacks the same level of moral and practical competence. It makes perfect sense, in other words, to say that objective consequentialism is usable for my neighbour Freya but not for *me*. In light of this fact, it is not immediately obvious what it means for a moral principle to be unusable *simpliciter*. And hence, it is not immediately obvious what it means to *reject* a moral principle on the grounds that it is unusable *simpliciter*.

One might suggest, by way of response, that there is still going to be a fairly straightforward way of figuring out a moral principle's usability *simpliciter*. Perhaps usability *simpliciter* depends somehow, for instance, on whether or not that principle is usable for every agent who has ever lived. I do not think, however, that any simple suggestion along these lines is going to work. Take the following example. One can easily imagine a principle *P* for which only 30% of persons find themselves capable of using *P* in regulating their conduct and navigating the moral world. It is extremely tempting to say that such a principle fails to qualify as usable *simpliciter*, since *P* is incapable of being used by the majority: for most of the agents who ever live, those agents will never be able to employ *P* in regulating their conduct and navigating the moral world. However, suppose it turns out that this 30% contains the majority of the individuals within the population whom we would identify as robustly having the full set of cognitive capacities and sensitivities typically required for careful moral reasoning. (Of the others, some find themselves incapable of using *P* because they are unusually bad at maths, some because they were mistaught at an early age and

"Consequentialism and Cluelessness," *Philosophy and Public Affairs* 29, no.4 (2000):342–370. Kagan, 64, takes this to be the textbook objection to consequentialist theories of moral action: "In fact lacking a crystal ball, how could you possibly tell what *all* the effects of your act will be? ... This seems to mean that consequentialism will be unusable as a moral guide to action." Similarly Lenman, 360, complains that consequentialism must surely "furnish us with a regulative ideal to guide our choices."

hence never grasped the relevant moral concepts, some because they happen to suffer from acute short-term memory loss, and so on.) If this is the case, then it is extremely tempting to say that *P does* qualify as usable *simpliciter*, insofar as it can be used by the majority of agents who realise the full set of cognitive capacities standardly required for careful moral reasoning. This seems like a another plausible way of determining usability *simpliciter*, but it contradicts the majoritarian rule used a moment ago. I am sceptical, then, whether there is going to be a simple means by which we might shift to discussion of the agent-relative usability of moral principles to the usability of moral principles *simpliciter*. It is not obvious, at the very least, how one could argue that a particular notion of usability *simpliciter* is the right one.

A second point of complication is that the usability of a moral principle, for any given agent, seems to come in degrees. Imagine, for instance, two principles *A* and *B*. Suppose that I can use *A* in regulating my conduct in every decision scenario I ever face. Suppose that I can use *B* in regulating my conduct in every decision scenario I ever face save one: the single choice coming up next Thursday of whether or not to order sashimi versus nigiri for dinner. The following seem to be correct things to say about both *A* and *B*. First, that both *A* and *B* are extremely usable principles for me as a decision maker. But second, that the principle *A* is *slightly more* usable for me as a decision maker than the principle *B*, insofar as there are more instances in which I can use *A* in regulating my conduct over the course of my moral life. Similarly, imagine two principles *C* and *D*. Suppose that I can never use *C* in regulating my conduct. Suppose that I can never use *D* in regulating my conduct save for one single choice scenario: the decision coming up next summer concerning whether or not to holiday in Spain. The following seem to be correct things to say about both *C* and *D*. First, that both *C* and *D* are extremely *non*-usable principles for me as a decision maker. But second, that the principle *C* is less usable for me as a decision maker than the principle *D*, insofar as *D* can still serve as a one-off moral tour-guide for my European vacation. All this makes the following thesis tempting: the usability of a moral principle is an agent-relative matter which permits of degrees, where the degree to which a moral

principle is *usable* for a particular agent will depend, at the very least, on the number of decisions in which that moral agent can use the principle across the course of their moral life.

4.1.3

These two complications make it much harder to see how we can simply reject a moral principle on the grounds that it is not *usable.* A quick-and-easy suggestion like this fails to address two crucial issues: first what is the relevant reference class, and second, what is the necessary degree or level of usability that must be obtained by a particular principle *with respect* to that reference class if the principle in question is to avoid being rejected as false. Let us see if these two issues can be resolved.

First, I have so far suggested that action-guidingness is an agent-relative concept. If, then, we reject a principle as false on the grounds that it is insufficiently action-guiding, we need to provide an answer to the prior question: *insufficiently action-guiding for whom*? We need to tell, in other words, some prior story as to why some particular reference class is salient, such that their inability to use a moral principle for guidance gives us grounds for rejecting that principle as false *simpliciter*.

I do not think any such story can be told. Take any moral principle P. If P guides some reference class x, but fails to guide some reference class y, this certainly tells us one thing – namely, that y will have a hard time attempting to use the principle P as they navigate the moral world. But this does not give y a reason for thinking that the principle is *false*. After all, it remains the case that by x's lights, the principle P is still a perfectly usable guide to life and action. And the same can be said, vice versa for any principle that turns out to be action-guiding for y but not for x.

What we can of course say, in either case, is that when x and y find themselves respectively unable to use a given principle P, they each have pragmatic reasons for looking elsewhere for guidance. After all, attempts to rely on the principle P itself for guidance may end in disaster. Things may instead go much better if x and y draw on auxiliary rules for the regulation of their conduct. But this is still a far cry from saying

that *x* and *y* have grounds for rejecting the principle in question as *false*: for saying that the general principle they are unable to use in their particular circumstances cannot be true.

Suppose, for instance, that *Mr Muddle* is attempting to evaluate subjective consequentialism as a moral principle. (Take *subjective consequentialism* to be the thesis that one act is better than another in virtue of having the greater consequential value in expectation.) Sadly, Muddle is terrible with probabilities. As a result, no matter how hard he tries, he can never select an act which he actively believes to have the greater value in expectation. Muddle's best friend, on the other hand, *Madam Expectation*, is brilliant when it comes to the calculation of expected value, and hence, regularly chooses the act she takes to have greater value in expectation. It should not follow from all this that Muddle has grounds for rejecting subjective consequentialism *simpliciter*, as literally *false*, simply because he himself cannot select acts with the belief that they maximise expected value. For Muddle to reject subjective consequentialism in this way would seem to neglect certain further facts that Muddle knows at the moment of choice: namely, that there exist other moral agents with an alternative set of cognitive facilities who *can* fruitfully use subjective consequentialism in forming useful beliefs about the status of their options. (In fact, in such a situation, Muddle might sensibly lament as follows: "If only I could use subjective consequentialism, the *true* moral principle, just as she does!")

Now what if, in the extreme case, some principle *P* couldn't ever be used by *anyone* to *any* degree in regulating their conduct? Would that give us a reason for rejecting the principle as false? Although this may seem an intuitively appealing suggestion, it turns out that is not easy to spell out what such a moral principle would look like. Take, for instance, the example of objective consequentialism. It is certainly true that *we* have a hard time in using objective consequentialism to regulate our conduct: we can't select some options rather than others with any kind of sufficiently justified belief that, in doing so, we bring about the better actual consequences. This is because we live in a complex causal web with countless other agents, a web in which even the most

seemingly trivial of our choices have uncertain and risky consequences which multiply over time. But it seems that there are some decision makers, at least in principle, who *could* use objective consequentialism in regulating their conduct: consider the last survivor, for instance, who has it on good authority that the consequences of their actions will affect none but themselves. Or imagine a certain future child who is jettisoned on a shuttle into space on a course they cannot control, with no hope of coming back, and whose decisions affect none but themselves in always very dull and predictable ways. It is plausible that objective consequentialism would be, at least sometimes, usable for such agents: such agents might act out of a desire to conform to the standards of objective consequentialism and with a well-justified belief that their behaviour did so conform. The question is then going to be whether the fact that *we* cannot use objective consequentialism in the way that *they* can gives us a reason for rejecting objective consequentialism as false *simpliciter*. It is not obvious to me that it does, just as Muddles didn't have a reason to reject subjective consequentialism as false *simpliciter* given that he (but not *Madam Expectation*) was unable to employ the principle in regulating his conduct.

Now for the second point. I emphasised previously that usability is not only an agent-relative concept, but further, a gradated concept. The usability of a principle is measured in degrees, and the degree to which a principle is action-guiding may depend on the number or percentage of decisions within one's moral life for which that principle is usable. Given this fact, we need a plausible story about how exactly rejection works for principles which exhibit degrees of usability. Just as before, I am not convinced that any such plausible story can be told.

First, consider the following. A principle P might possess a higher degree of usability for some reference class x than does an alternative moral principle Q, but the question remains whether this gives us a *pro tanto* reason for the rejection of Q. And it seems to me that, clearly, it does not. For a start, the mere fact that P guides x to a greater degree than Q does tells us nothing about the action-guidingness of P or Q in *absolute* terms. It could still be the case, for instance, that for almost every decision x

ever makes, x can select options with a desire to conform to either P or Q, and with a justified belief that, in choosing those select courses of action they do, they conform to the standards of either P or Q. In that case, the fact that P has a greater degree of usability for x than Q does ought play no role in our comparative evaluation of either principle. It should not, for instance, give us a *pro tanto* reason in favour of endorsing P over Q.

What we would need to endorse, then, is a kind of *non-comparative usability threshold*: some specified degree or level of usability which moral principles must reach (for a specified reference class) in order to avoid being rejected as false. On this view, it would be a necessary condition on any true moral principles that they reach the usability threshold (for a specified reference class).

Here, though, we face some of the familiar problems that stem from drawing a sharp threshold across a gradated concept. Suppose that subjective consequentialism guides me for 80% of those decisions I face in the course of my moral life – for 80% of those moral decisions I make, I can use subjective consequentialism to discriminate between the deontic status of the options, form the desire to perform some rather than others, and subsequently, perform the act which I take to maximise expected value. Suppose, further, that 80% is just on the cusp of the action-guidingness threshold. (Suppose too, for simplicity, that I am the relevant reference class.)

Unless more is said, however, this view is going to have strange consequences. It is going to imply that, counterfactually, if subjective consequentialism had been usable for only one less decision in the course of my moral life, then subjective consequentialism would have been insufficiently usable, and hence, no longer a viable account of right action. This would have been the case in the world in which I was only a tad worse at maths – the sad world in which, on just one occasion, I was too slow to wheel-and-deal in the requisite probabilities in calculating expected value. This would be the case even if the *contents* of the subjective consequentialist principle – that one act is better than another so long as it has the greater expected value – were to remain unchanged. A conclusion like this is strange: it is strange to admit that the viability of a moral

principle is so sensitive to the precise number of cases in which certain agents can use that principle in regulating their conduct.

Normally, one can ameliorate the difficulties stemming from precise thresholds by falling back on a range: admitting that there may be a range of points at which, for instance, there is no fact of the matter about whether or not a principle is sufficiently action-guiding. (80% is in the range, as is 79% and 81%, and so on.) But in this particular case, when one is attempting to use action-guidingness as a criterion for the *evaluation* of our moral principles, this only makes matters worse. By falling back on ranges we are in effect forced to concede that, for any number of *prima facie* plausible moral principles, those principles cannot be true normative generalisations insofar as there is no fact of the matter about whether those principles are sufficiently usable. This does not constitute progress on the problem of usability thresholds.

4.2 Against the Pragmatic Tradition

Those, then, are two initial problems with what I have coined the standard reading of action-guidingness objections. Under the standard reading, recall, we said that a moral principle is false insofar as it cannot be used. I have suggested that this standard reading is difficult to sustain given both the agent-relative and gradated nature of usability.

In arguing against this standard reading, I have been cutting against a long-established tradition of moral thought which says that the correct moral principles, if any, must be usable. More specifically, this tradition says that the correct moral principles, if any, must be capable of fulfilling two roles: a theoretical role in specifying the general features which account for the normative status of our options, but also a practical role in leading suitably-motivated agents towards some forms of conduct rather than others.

Holly M. Smith calls this the *pragmatic* tradition, since its proponents maintain that moral principles serve an inherently pragmatic role.[4] If the pragmatic tradition is right,

[4]Smith, *Making Morality*, 47, provides an extended survey of those whom she identifies as pragmatists,

then it seems like a moral principle *could* indeed be rejected on the grounds that it was not usable – bracketing off, at least for now, the problems I have sketched concerning the gradated and agent-relative nature of usability.

In order to advance the discussion, then, it will be most helpful to consider some of the standard justifications given by pragmatic theorists in defence of the claim that the correct moral principles must be usable. In this section I consider three such justifications, before rejecting each in turn. Some of these justifications also attempt to provide support for the stronger claim that the correct moral principles must be universally usable, that is, usable by all morally competent agents in regulating their conduct.

The inherent function or concept of a moral theory

One common suggestion in the pragmatic tradition is that the correct moral theory must be usable because this is somehow tied up in *what it is* for a theory to be a moral one.[5] As Elinor Mason puts it, "the most important function of a moral theory *is* to guide action."[6] On such a view, a moral theory couldn't be a *moral* theory if it indeed turned out that it was incapable of being used.

As Smith emphasises elsewhere, it would be a much broader meta-ethical project to try and figure out exactly why moral theories might be inherently regulative in nature.[7] That broader project is not the topic of this paper. We can, however, here emphasise a few points of caution.

The first point of caution is that this claim – that moral theories must be usable

including Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge MA: Harvard University Press, 1990); Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics*, 101, no.3 (1991):461–482; John L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977); John Rawls, *A Theory of Justice* (Cambridge MA: Harvard University Press, 1971); and Thomas Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge MA: Harvard University Press, 2008). One could add to this list, plausibly, those featuring in the next footnote.

[5]This suggestion appears variously in Robert Goodin, "Demandingness as a Virtue," *The Journal of Ethics*, 13, no.1 (2009):1–13, 3; GE Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1993), 151-152; Jan Narveson, *Morality and Utility* (Baltimore: John Hopkins University Press, 1967), 112; Peter Singer, *Practical Ethics* (Cambridge: Cambridge University Press, 1979), 2; and George Sher, *Who Knew?* (Oxford: Oxford University Press, 2009), 139.

[6]Elinor Mason, "Consequentialism and the 'Ought Implies Can' Principle," *American Philosophical Quarterly*, 40, no.4 (2003):319–331, 327 with emphasis added.

[7]Smith, *Making Morality*, 55.

because this is what it *is* for a theory to be a moral one – does not furnish us with an answer to the two worries I have outlined in the previous section. Namely, it does not solve the problems I have outlined concerning the reference-class relative and gradated nature of usability. Without a solution to these two problems, we should be hesitant to simply assert that a moral principle ought to be rejected as false given that it is not usable. A more detailed story than this would need to be told.

The second point of caution is that you might well think that moral *theories* have a necessarily regulative function – indeed, that this regulative function is required in order for a theory to be a moral one – but that it is still a further matter whether or not *the correct moral principles themselves* must be usable.

Here is why this point of nuance is so important. One common thought is that moral theories contain moral principles *and other things too*, where those 'other things too' are capable of serving a regulative function for individual agents as they navigate their moral lives. This is the case, for instance, in the so-called hybrid conception of moral theory advocated historically by Smith and by a range of others.[8] Under the hybrid conception of moral theory, a moral theory contains some given *evaluation principle*, a theoretical generalisation which specifies the general features in virtue of which our acts have the normative status they do. But a moral theory also contains some set of *selection procedures* or *decision guides*, the sort of things that agents can draw upon, in the course of their moral lives, in coming to form the intention to perform some acts rather than others. In following the dictates of the decision guides, agents hope to (indirectly) follow the dictates of their given evaluation principle.

What matters for our purposes here is that under the hybrid conception of moral theory, a moral evaluation principle like objective consequentialism might not be usable, but the multi-tiered moral *theory* in which it is embedded might still be usable. So

[8]Plausibly, John Stuart Mill endorses something like the hybrid view in *Utilitarianism* (Indianapolis: Bobbs-Merrill, 1957). See also Roger Crisp, "Utilitarianism and the Life of Virtue," *The Philosophical Quarterly*, 42 (1992):139–160; 43-44 of R.M. Hare, *Freedom and Reason* (Oxford: Clarendon Press, 1963); Philip Pettit & Geoffrey Brennan, "Restrictive Consequentialism," *Australasian Journal of Philosophy*, 64, no.4 (1986):438–455; also see 50-53 of J.J.C. Smart, *An Outline of a System of Utilitarian Ethics* in *Utilitarianism: For and Against*, J.J.C. Smart & Bernard Williams (Cambridge: Cambridge University Press, 1973).

it is not obviously true that acknowledging usability as a necessary function of moral theories demands that we reject particular moral principles (objective consequentialism, say) in the case that those principles fail to be usable.

The autonomy argument

Here is a second argument for thinking that the correct moral principles, if any, must be widely or universally usable. The correct moral principles, if any, must be usable since this is an important means by which agents' *autonomy* as moral decision makers is secured. As Smith puts it:

> "The usability of a principle is important because it ensures that motivated agents can achieve a certain form of autonomy in making their choices. An agent who cannot find any way to translate his moral values into his *choice* of what to do is an agent who cannot find a way to govern his decision by the considerations he deems most relevant."[9]

Call the sense of autonomy at stake here *value autonomy*. One who endorsed the above line of thought might reason as follows. If a moral principle is usable for some agent x, then that principle, when used, secures value autonomy for x. Since value autonomy must be widely secured, moral principles must be usable.

Note, however, that as it stands this argument risks affirming the consequent. It might certainly be true that, if a moral principle can be used by agents in regulating their conduct, then those agents achieve value autonomy insofar as they use it. But it does not follow from this that in order to secure value autonomy, agents must regulate their conduct using that moral principle. It might still be the case that agents are able to secure value autonomy via *other* means. What we would need for this argument to hold weight, in other words, is some kind of reason for thinking that the deliberative application of usable moral principles is the *only* stable means by which individual agents might secure value autonomy.

[9]Smith *Making Morality*, 194–195. Also see Holly M. Smith, "Making Moral Decisions," *Noûs*, 22, no.1 (1988):89–108.

It seems to me, however, that it is not: there appear to be other means by which agents might still secure value autonomy in the relevant sense. Smith's suggestion, recall, was that value autonomy consists of having a stable way in which to translate ones values into one's choices. But it seems that agents can secure value autonomy in this sense, for instance, by unreflectively choosing to join certain communities where particular values they share are privileged, or by unreflectively tending to cultivate certain psychological dispositions that regularly select in favour of some of their held values over others. These would be means by which agents could inform their choices in light of their values without at any point deliberatively attempting to apply moral principles in the regulation of their conduct. There may be means other than these, too, but the key point is that it seems there exist mechanisms for expressing one's values in one's choices other than the deliberative application of moral principles to one's particular circumstances.[10]

Here, then, is what we should say about the autonomy argument. Smith draws our attention to an interesting fact: if moral principles or theories were usable, then that would seem to offer an important means by which individual agents might secure value autonomy, where value autonomy consists of having a stable means by which to translate one's values into one's choices. Even taking value autonomy to be of the utmost importance, however, this alone does not seem to imply that the correct moral principles must be widely usable. A further relevant question would remain – namely, whether there could exist certain *other* means by which agents might translate their values into choices in the relevant sense. Until we have a reason for thinking that *only* used moral principles can secure value autonomy in the sense described by Smith, we should hesitate to accept the conclusion of the autonomy argument.

[10]This is, I take it, a key theme in Philip Pettit's moral psychology of consequentialism. Agents need not pursue the values that move them in a "rationalistic, calculative manner" – see Philip Pettit, "Consequentialism and Moral Psychology," *International Journal of Philosophical Studies*, 2, no.1 (1994):1–17, esp. 11.

The Morally Successful Life

One line of thought attributable at least in part to Bernard Williams is that the correct moral principles would have to be universally usable in order to secure a certain kind of justice.[11] The reasoning, here, goes as follows. It would be unjust if the morally successful life was open to some but not all, where this discrepancy stemmed from purely arbitrary factors beyond an agent's control. If this were the case, then morality itself would exhibit a kind of in-built unfairness: some agents would be, through no fault of their own, 'locked out' of moral success.

We then note the following. If moral theories *weren't* universally usable, then there would exist some agents unable to regulate their conduct by drawing upon the correct moral theory, even if they so desired. Some agents would thus be unable to access the morally successful life. And morality itself would exhibit, as I phrased it a moment ago, a kind of in-built unfairness.

Here is one possible response to this suggestion: as I have just argued, and as is perhaps becoming a recurring theme, agents might still be able to act in accordance with their values via means *other* than the calculative and deliberative application of moral theory to their particular circumstances. If this is the case, then agents who find themselves unable to use the correct moral theory in regulating their conduct are not *necessarily* locked out of moral success. They may be 'swept up' in moral success via other means.

But as a more general point, we ought to say the following. It is difficult to make sense of William's suggestion until we spell out what we mean by the phrase *morally successful life.* Following Smith, there are two senses of the phrase between which we ought to distinguish.[12] A strongly morally successful life would involve never doing wrong by the lights of the true moral principle. Under objective consequentialism, for instance, a strongly morally successful life would involve always choosing the act which happens to maximise total consequences. An agent capable of living a

[11]Bernard Williams, *Moral Luck: Philosophical Papers 1973–1980* (Cambridge: Cambridge University Press, 1981), 20, 36.

[12]Smith, *Making Morality*, 197-199.

strongly morally successful life would be capable of living blamelessly for this very reason, namely, for the reason that they would never deviate from the courses of action recommended by the true moral principle.

It would be a wonderful world if we all, in virtue of the nature of morality, had guaranteed to us the possibility of living a strongly morally successful life. But the simple fact is that we live in a world of uncertainty and a world of mistaken beliefs. In light of our uncertainty and of our mistaken beliefs, no single agent has guaranteed to them the possibility of living a strongly morally successful life. There is always the possibility that even the best-intentioned agents might act wrongly by the lights of the evaluation principle they endorse.[13]

A *modestly* morally successful life, however, would simply be realised whenever an agent lived blamelessly; such a modestly successful moral life would be available to an agent whenever that agent had the capacity to live blamelessly. Note that an agent might achieve this modestly morally successful life and still regularly act wrongly by the lights of the true evaluation principle(s). This would be the case so long as there existed adequate excusing conditions for those scenarios of uncertainty or non-culpable mistaken belief in which well-intentioned agents made moral errors unwillingly. In sum: a modestly successful moral life would be realised *even if* some agent violated the recommendations of the true evaluation principle(s), so long as that agent was not blameworthy for doing so.

For our purposes, all that matters is that the modestly morally successful life might still be available to all even if the correct moral theory or principles were not universally usable. In fact these two issues, of the modestly successful moral life and of the widespread usability of moral theory or principles, appear to be largely orthogonal. The condition of universal usability might fail: certain agents might not be able to use the correct moral theory in regulating their conduct. But whether or not those agents are still capable of achieving the modestly successful moral life is then going to be a question of whether those agents are deserving of blame. And *this* further question will

[13]Similar problems emerge for 'subjectivised' moral codes, insofar as agents can also be mistaken or uncertain about their own beliefs, motivations, and attitudes. See Smith, *Making Morality*, 80.

depend on the more general issue of whether there exist relevant excusing conditions for the agent in question.[14] In sum, it is *not* the case that, for the morally successful life to be open to all, moral principles must exhibit widespread usability. The widespread usability of moral principles is not required in order to achieve this distinctive kind of justice.

I have considered, then, three arguments from the pragmatic tradition – three attempts to justify the claim that the correct moral principles must be usable to some or even to all. Such arguments, if successful, would seem to imply that non-usable moral principles can be rejected as false or defective. These three were the argument from the very concept or function of a moral theory, the argument from value autonomy, and the argument from the morally successful life. I have rejected each. Without these positive arguments, we lack further justification for rejecting moral principles as false or incorrect on the grounds of their unusability.

4.3 An Epistemic Gloss on Action-Guidingness Objections

4.3.1

I have argued, so far, that the standard reading of action-guidingness objections is unsuccessful: it does not make much sense reject a moral principle as incorrect on the grounds that it is insufficiently usable *simpliciter*. I first gave an argument against this standard reading: such a standard reading is implausible given the gradated and agent-relative nature of usability. I then dismissed some positive proposals, historically given, in favour of thinking that the correct moral principles are necessarily usable for some or for all. Such arguments, if successful, would seem to imply that moral principles *can* be rejected as false or incorrect on the grounds that they are not usable.

All this is not to say, though, that there is *no* sensible way of reading action-guidingness objections. In this section, using the literature of so-called cluelessness arguments against objective consequentialism as a case study, I show that there is an

[14] Smith, *Making Morality*, endorses this line of argument on 199.

alternative way in which one might interpret action-guidingness objections.

First, then, a brief sketch of the way in which cluelessness arguments against objective consequentialism usually go.[15] Proponents of cluelessness arguments against objective consequentialism note that we are *clueless* as to the total consequences of our acts, given the chaotic and chancy world in which we live. (We have enough trouble predicting the immediate consequences of our acts, let alone the total consequences of our acts as causal history proliferates hundreds and thousands of years into the future.) The proponents of such cluelessness arguments then go on to note that these total consequences, of which we are clueless, are the sole tool used by objective consequentialists in evaluating comparative betterness across their options. Since we are clueless as to total consequences, the conclusion reads, the objective consequentialist is ultimately clueless as to comparative status of their options.

Hilary Greaves has aptly termed this the *cluelessness worry* since, although the cluelessness worry is certainly unsettling, it is often left unclear how, exactly, the cluelessness worry serves to undermine the truth of objective consequentialism.[16] It is not immediately obvious, for instance, the best way in which one could spell out the cluelessness worry into a deductive argument.

If the cluelessness argument is read as a standard action-guidingness objection, the sort of which I have been critical in the previous sections, it might bear the following deductive form:

The Action-Guidingness Cluelessness Argument

1. Given the empirical facts, agents who subscribe to objective consequentialism cannot use objective consequentialism in regulating their conduct.

2. The correct moral principle can be used by agents who subscribe to it in regulating

[15]See, for instance, "Cluelessness Redux." The most widely cited text on cluelessness for consequentialists is Lenman, "Consequentialism and Cluelessness." But also see Frances Howard-Snyder, "The Rejection of Objective Consequentialism," *Utilitas*, 9, no.2 (1997):241–248; Elinor Mason, "Consequentialism and the Principle of Indifference," *Utilitas*, 16, no.3 (2004):316–321; Gerald Lang, "Consequentialism, Cluelessness, and Indifference," *The Journal of Value Inquiry*, 42 (2008):477–485; and Hilary Greaves, "Cluelessness," *Proceedings of the Aristotelian Society*, 116 (2016):311–339.

[16]Greaves, 312.

their conduct.

C. Objective consequentialism is not the correct moral principle.

I have already suggested, however, that such standard readings of action-guidingness arguments are unsuccessful. It is not plausible to suggest that a moral principle can be rejected as false given only that it is not usable.

There is, however, a fairly different way of spelling out the cluelessness worry. That way would go as follows. Note that if objective consequentialism were true, then we would be clueless as to the comparative status of our acts. Hence, we could not know certain moral facts that we *do* regularly take ourselves to know. In particular we could not know, when making a pairwise comparison between the available options in the course of our daily moral lives, whether some available option was better than another: whether it was better to donate money to an effective charity or to burn it, to run over the pedestrian or swerve, to save the drowning man or leave him be, or to torture the puppy rather than nurture and care for it.

This result, one might worry, cannot be right. Surely, when making pairwise comparisons in the course of our daily moral lives, we at least *sometimes* know about the comparative status of our acts. We at least *sometimes* have such pieces of moral knowledge at our disposal when engaging in moral theory building. If any moral principle would force us to deny as much, then we have a *reductio* against that moral principle.

On this reading, interestingly enough, it turns out that (so-called) action-guidingness objections are *not* about whether certain moral principles can play a pragmatic role in leading us to some forms of conduct rather than others. What really ends up mattering is that the truth of certain moral principles (objective consequentialism, say) would preclude us from affirming the commonsense judgements about the comparative status of our options that we ought to be able to affirm. This underlying problem *just so happens* to come with the further result – a kind of unlucky byproduct, as it were – that such principles are tough to use in the regulation of one's conduct.

This reading of cluelessness arguments is not too strained. James Lenman, for

instance, alludes to this reading of the cluelessness worry for consequentialists in the following passage:

> "So we have only the feeblest of grounds, from an objective consequentialist perspective, to suppose that the crimes of Hitler were wrong. Here, if anywhere, surely, there is a considered moral judgment at stake that is well-enough entrenched not to be up for grabs in the cut and thrust of reflective equilibrium, a judgment far enough from the periphery of the web of our moral beliefs to furnish a compelling reductio of any theory that might undermine it."[17]

The idea, here, is much as I have just described: objective consequentialism cannot be right since the principle does not allow us to affirm the commonsense judgements about the comparative status of our options that we ought to take as our starting points. This, too, is the broad way in which I have developed the cluelessness argument against consequentialism elsewhere.[18] I have argued previously, for instance, that since we regularly have conflicting evidence as to the long-term effects of our actions, where the way in which this conflict ought to be resolved remains unclear, subjective consequentialists cannot affirm the commonsense judgements of comparative status of the options that we can surely take ourselves to know when engaging in the project of moral philosophy. It is for *this* reason that we might be sceptical of the subjective consequentialist thesis; not for the further reason that the subjective consequentialist thesis is difficult to employ in regulating our conduct.

4.3.2

If we read action-guidingness arguments in this way, are they any more successful? It is worth concluding, I think, by sketching some tentative issues for this style of argument.

[17] Lenman, 349. Also see Lang, 477.
[18] See "Cluelessness Redux."

The first issue is merely one of framing. I have framed this epistemic interpretation of action-guidingness objections in terms of our ability to *know* about the comparative status of acts. But there is no reason, I suspect, why the argument could not also be framed in terms of our credential states. That sort of argument, I suspect, would operate in the following way: we have high credence in the truth of some comparative judgements across the options, but certain moral theses would force us to revise (radically downwards) our credence in the truth of those particular judgements. And we may think that the fact that a moral principle entails such radical and widespread credential shifts might sometimes render that moral principle implausible. In any case, I am happy to talk about our moral knowledge, and so I am happy to frame the argument in terms of our moral knowledge too.

The second and more crucial issue is a possible line of response from the objective consequentialist. Let's begin by noting the following. While objective consequentialists (say) may not be able to affirm certain commonsense judgements about the comparative status of our options out there in the world *as it is*, objective consequentialists might still be able to affirm certain counterfactual or hypothetical judgements about the comparative status of options. The objective consequentialist can still affirm that *all things being equal*, *holding fixed future histories,* it would be better to donate money to charity rather than to burn it, to swerve rather than hit the pedestrian, to save the stranger rather than leave them to drown, to pat the puppy rather than torture it, and so on. The objective consequentialist can still maintain that, in such-and-such hypothetical and counterfactual cases, with such-and-such specified consequences at stake, some acts would clearly be better than others in virtue of having the better total consequences. The objective consequentialist might claim that our ability to affirm these judgements about such hypothetical and counterfactual cases is sufficient: that, insofar as they can uphold such counterfactual or hypothetical verdicts, they cannot be criticised as somehow lacking moral knowledge.

"But surely," the defender of cluelessness might press, "a moral principle must vindicate the ordinary and commonsense judgments about comparative status that

we make *out there in the real world*, in the course of our actual moral lives. A moral principle that could never correctly discriminate between the options, in the course of our actual moral lives, would be too far removed from reality. It would make us, of a sort, moral skeptics."

I am not sure, however, whether this further insistence is entirely persuasive. We are, in the end, limited creatures. It may ultimately turn out, as an empirical matter of fact, that it is incredibly difficult to tell which of our acts do in-fact have the better consequences in the extremely long run. If objective consequentialism really were a correct moral principle, then as a result, it would be incredibly difficult for even the best of us to tell which of our acts (out there in the real world) were in fact better than which, and how each of us ought to act. This may inspire a certain melancholy, but ought it inspire us to abandon objective consequentialism? This melancholy result, for instance, would not preclude the possibility of the objective consequentialist still doing moral theorising: the objective consequentialist might still have considered intuitions about particular hypothetical or counterfactual moral decision scenarios, hypothetical or counterfactual moral decision scenarios in which the world is simplified, and in which the distant future consequences of our acts are held fixed.

A sharper response from the defender of cluelessness might here involve pointing to our collective moral practices – for instance, our collective practices of praise and blame – and suggesting that these practices would be undermined by this last objective consequentialist manoeuvre. The thought, here, goes as follows. We just saw that the objective consequentialist can only affirm certain hypothetical judgements about the comparative status of our acts – affirm, for instance, that *all things being equal* it would have been better to swerve away from the pedestrian or to spare the puppy. But our collective practices of praise and blame, it seems, typically presuppose more than an affirmation of such mere hypotheticals. When I criticise you for acting wrongly, this seems to presuppose that I have at least some level of confidence that they *really did* act wrongly, in the actual world, in behaving as they did.

If the objective consequentialist cannot be at all confident that you acted wrongly in

swerving to hit the pedestrian, or in torturing the puppy, then it seems unclear whether the objective consequentialist could blame you for doing as much. And *this* result is indeed implausible. I suspect the defender of objective consequentialism would at this point need to offer an account of praise and blame which can account for our lack of confidence concerning whether or not those around us whom we criticise do indeed act wrongly, in behaving as they actually do.

4.4 Conclusion

In this chapter I have argued against a fairly standard interpretation of action- guidingness objections: an interpretation under which we say that a principle is false insofar as it is not usable. I have argued against this common interpretation because it neglects the agent-relative and gradated nature of usability as a property of moral principles. Although there already exist positive arguments for thinking that usability is a necessary property of moral principles, I have, in this paper, attempted to cast doubt on these common positive arguments. I have suggested that neither the argument from the concept or function of moral theory, nor the argument from value autonomy, nor the argument from morally successful lives, establishes the following point: that we can reject a proposed moral principle as incorrect on the grounds that it is not usable.

This does not mean that we must disregard action-guidingness arguments altogether. I have, in this paper, offered an alternative way of interpreting action-guidingness arguments, drawing on the cluelessness literature against consequentialism as a case study. We might accuse a principle of being insufficiently action-guiding because, more fundamentally, the truth of that principle would seem to imply that we cannot know about the comparative status of our options in the way that we usually *do* take ourselves to know about the comparative status of our options. This result may lead us to think that the proposed moral principle is false. We can think of this as an epistemic reading of action-guidingness objections.

In the particular case of cluelessness arguments against objective consequentialism,

the objective consequentialist could respond to the epistemic interpretation of the argument by insisting that they can still affirm or claim to know certain important hypothetical or counterfactual judgements of comparative betterness across options. This response may have merit. I have suggested, however, that this response does not fit neatly with our collective practices of praise and blame. How could we blame others for acting wrongly, as we regularly do, if we do not have a clue as to whether, acting as they do in the actual world, they have acted wrongly? That is the puzzle with which objective consequentialists would be left under this line of response.[19]

[19] And for that matter, subjective consequentialists too, if what I say in *Cluelessness Redux* is right.

Concluding Remarks

This thesis, it may seem, has contained a lot of negativity. I've spent a lot of time arguing *against* certain ideas: in the first chapter against consequentialist accounts of right action, in the second chapter against recently popular relevance approaches to claims aggregation, in the third chapter against a traditional interpretation of the DDA, and in the last chapter against the merits of 'action-guidingness' objections. By this point, you might think I've had just about everybody in my sights.

My central concern in the preceding chapters, however, has been to show that we must take the stunning facts of our cluelessness seriously. When we apply established moral principles and doctrines to decision scenarios in which the empirical facts of our cluelessness are properly foregrounded, surprising and absurd results quickly follow. This proved true for consequentialists, but it proved just as true for those who preferred to think in terms of relevant claim satisfaction, and it proved just as true for those sympathetic to a non-consequentialist distinction like the DDA. All this has, in the previous chapters, tended to leave us with choice points in the following vein: either abandon those established moral principles and doctrines we hold dear, or refine them such that they can better handle the severe risks and uncertainties inherent in moral decision making.

The refinements in question, as we have seen, typically involve a subjective shift. The consequentialist starts evaluating right action by asking whether some act maximises *expected moral value*; the relevance theorist starts grounding individuals' claims in their *ex ante prospects*; the defender of the doing versus allowing distinction starts talking instead about whether or not our acts impose *additional risks of harm* upon

individuals, relative to some specified baseline.

At various points in the previous chapters, nonetheless, I have tried to emphasise that subjective shifts like these are not so simple. Take the example of subjective consequentialism: given our evidential situation as regards the extremely long-run future, it is not obvious whether the subjective consequentialist can still affirm certain commonsense judgements about the comparative status of the options that we ought to be able to take as our starting points. And take the example of *ex ante* relevance views: if the relevance theorist wishes to ground individuals' claims in their *ex ante* prospects, then we are going to need a story about which designators are relevant for determining relevance. Depending on the story we tell, relevance views may still end up with the implausible result that minor claims never count. Things were more optimistic in the case of the DDA: it seems that we can indeed preserve many of the canonical judgements associated with the DDA by shifting to a view concerning additional risk imposition. The lingering question for such a view, I suspect, is whether there are sometimes different *ways* in which we might bring about a particular risk imposition, where we have greater reasons against imposing risks in some ways rather than in others.

I do not claim, here, to have completely settled these further questions. But until much more is said, it strikes me that the woes of cluelessness are yet to be resolved – even assuming that consequentialists, relevance theorists, and defenders of the doing versus allowing distinction all modify their views via subjective shifts of the sorts I have considered in the previous pages. Of course, we might *mitigate* the woes of cluelessness by subjectivising those moral views we hold dear. But there is a difference between mitigating our woes, and between resolving them entirely.

Bibliography

Adler, Matthew. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. New York: Oxford University Press, 2012.

Adler, Matthew, and Nils Holtug. "Prioritarianism: A Response to Critics." *Politics, Philosophy & Economics,* 18, no.2 (2019):101–144.

Barry, Christian, and Garrett Cullity. "Offsetting and Risk Imposition." *Ethics*, 132, no.3 (2022):352–381.

Barry, Christian, and Garrett Cullity. "Do We Impose Undue Risks When We Emit and Offset? A Reply to Stefansson." *Ethics, Policy, & Environment* (forthcoming).

Bennett, Jonathan. *The Act Itself*. Oxford: Oxford University Press, 1995.

Bradley, Richard. *Decision Theory With a Human Face*. Cambridge: Cambridge University Press, 2017.

Broome, John. "Should We Value Population?" *The Journal of Political Philosophy*, 13, no.4 (2005):399–413.

Broome, John. "Against Denialism." *The Monist*, 102, no.1 (2019):110–129.

Burch-Brown, Joanna. "Clues for Consequentialists." *Utilitas* 26, no.1 (2014):105–119.

Crisp, Roger. "Utilitarianism and the Life of Virtue." *The Philosophical Quarterly*, 42 (1992):139–160.

Donagan, Alan. *The Theory of Morality* Chicago: The University of Chicago Press, 1977.

Foot, Philippa. "The Problem of Abortion and the Doctrine of Double-Effect." In *Virtues and vices*. Oxford: Basil Blackwell, 1978.

Foot, Philippa. "Killing and Letting Die." In *Killing and Letting Die 2nd Edition*, edited by Bonnie Steinbock and Alastair Norcross, 280–289. New York: Fordham University Press, 1994.

Frick, Johann. "Contractualism and Social Risk." *Philosophy & Public Affairs*, 43, no.3 (2015):175–223.

Gardner, John. "Complicity and Causality." *Criminal Law and Philosophy*, 1 (2007):127–141.

Gibbard, Allan. *Wise Choices, Apt Feelings*. Cambridge MA: Harvard University Press, 1990.

Godfrey-Smith, Peter. *Explaining Chaos*. Cambridge: Cambridge University Press, 1998.

Goodin, Robert. "Demandingness as a Virtue." *The Journal of Ethics*, 13, no.1 (2009):1–13.

Greaves, Hilary."Cluelessness." *Proceedings of the Aristotelian Society* 116 (2016):311–339.

Greaves, Hilary, and William MacAskill. "The Case for Strong Longtermism." *Global Priorities Institute Working Paper Series*, 5 (2021):1-43.

Hare, R.M. *Freedom and Reason*. Oxford: Clarendon Press, 1963.

Horton, Joe. "Aggregation, Complaints, and Risk." *Philosophy & Public Affairs*, 45, no.1 (2017):54–81.

Howard-Snyder, Frances. "The Rejection of Objective Consequentialism." *Utilitas,* 9, no.2 (1997):241–248.

Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics,* 101, no.3 (1991):461–482.

John, S.D. "Risk, Contractualism, and Rose's 'Prevention Paradox'." *Social Theory and Practice,* 40, no.1 (2014):28–50.

Joyce, James M."A Defense of Imprecise Credences in Inference and Decision Making." *Philosophical Perspectives*, 24 (2010):281–323.

Kagan, Shelly. *Normative Ethics*. Boulder: Westview Press, 1998.

Kamm, Frances. *Morality, Mortality, I: Death and Whom to Save from it*. New York: Oxford University Press, 1993.

Kamm, Frances. *Morality, Mortality, II: Rights, Duties, and Status.* Oxford: Oxford University Press, 1996.

Lang, Gerald. "Consequentialism, Cluelessness, and Indifference." *The Journal of Value Inquiry,* 42 (2008):477–485.

Lazar, Seth. "In Dubious Battle: Uncertainty and the Ethics of Killing." *Philosophical Studies*, 175 (2018):859-883.

Lazar, Seth. "Limited Aggregation and Risk." *Philosophy & Public Affairs*, 46, no.2 (2018):117–159.

Lenman, James. "Consequentialism and Cluelessness." *Philosophy and Public Affairs,* 29, no.4 (2000):342–370.

Lernpaß, Christoph. "A Diachronic Argument Against the Ex Ante Complaint Model." (unpublished manuscript).

Lighthill, MJ, and G Whitham. "On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences,* 229, no.1178 (1955):317–345.

Lorenz, Edward. "Deterministic Nonperiodic Flow." *Journal of the Atmospheric Sciences,* 20 (1963):130–141.

Lorenz, Edward. *The Essence of Chaos*. Seattle: University of Washington Press, 1995.

Mackie, John. *Ethics: Inventing Right and Wrong*. New York: Penguin, 1977.

Mahtani, Anna. "The Ex Ante Pareto Principle." *The Journal of Philosophy,* 114, no.6 (2017):303–323.

Mann, Kirsten. "Relevance and Nonbinary Choices." *Ethics,* 132, no.2 (2022):382–413.

Mann, Kirsten. "The Relevance View: Defended and Extended." *Utilitas,* 33, no.1 (2021):101–110.

Mason, Elinor. "Consequentialism and the 'Ought Implies Can' Principle." *American Philosophical Quarterly*, 40, no.4 (2003):319–331.

Mason, Elinor. "Consequentialism and the Principle of Indifference." *Utilitas,* 16, no.3 (2004):316–321.

Mill, John Stuart. *Utilitarianism*. Indianapolis: Bobbs-Merrill, 1957.

Mogensen, Andreas. "Maximal Cluelessness." *The Philosophical Quarterly,* 71, no.1 (2021):141–162.

Mogensen, Andreas. & William MacAskill, "The Paralysis Argument." *Philosophers' Imprint*, 21, no.15 (2021):1–17.

Moore, G.E. *Principia Ethica*. Cambridge: Cambridge University Press, 1993.

Nagel, Thomas. *Equality and Partiality*. New York: Oxford University Press, 1995.

Narveson, Jan. *Morality and Utility*. Baltimore: John Hopkins University Press, 1967.

Newell, G.F. "A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory." *Transportation Research Part B Methodological,* 27, no.4, (1993): 281–287.

Newell, G.F. "A Simplified Car-Following Theory: A Lower Order Model." *Transportation Research Part B Methodological,* 36, no.3 (2002):195–205.

Norcross, Alastair. "Comparing Harms: Headaches and Human Lives." *Philosophy & Public Affairs*, 26, no.2 (1997):135-167.

Otsuka, Michael. "Risking Life and Limb: How to Discount Harms by their Improbability." In *Identified versus statistical lives: an interdisciplinary perspective*, edited by Glenn Cohen, Norman Daniels, and Nir Eyal, 77–93. Oxford: Oxford University Press, 2015.

Palmer, T.N., A. Döring, and G. Seregin, "The Real Butterfly Effect." *Nonlinearity*, 27, no.9 (2014):R123.

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

Pettit, Philip. "Consequentialism and Moral Psychology." *International Journal of Philosophical Studies,* 2, no.1 (1994):1-17.

Pettit, Philip, and Geoffrey Brennan. "Restrictive Consequentialism." *Australasian Journal of Philosophy*, 64, no.4 (1986):438–455.

Quinn, Warren. "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing." *The Philosophical Review*, 98, no.3 (1989):287–312.

Rawls, John. *A Theory of Justice*. Cambridge MA: Harvard University Press, 1971.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Scanlon, Thomas. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge MA: Harvard University Press, 2008.

Singer, Peter. *Practical Ethics*. Cambridge: Cambridge University Press, 1979.

Sher, George. *Who Knew?* Oxford: Oxford University Press, 2009.

Smart, J.J.C. "An Outline of a System of Utilitarian Ethics." In J.J.C. Smart and Bernard Williams, *Utilitarianism: For and Against*, 1–76. Cambridge: Cambridge University Press, 1973.

Smith, Holly M. "Making Moral Decisions." *Noûs*, 22, no.1 (1988):89–108.

Smith, Holly M. "Two-Tier Moral Codes." *Social Philosophy and Policy*, 7, no.1 (1989):112–132.

Smith, Holly M. *Making Morality Work*. Oxford: Oxford University Press, 2018.

Steuwer, Bastian. "Contractualism, Complaints, and Risk." *Journal of Ethics and Social Philosophy*, 19, no.2 (2021):111–147.

Steuwer, Bastian. "Aggregation, Balancing, and Respect for the Claims of Individuals." *Utilitas*, 33, no.1 (2021):17–34.

Unruh, Charlotte Franziska. "Doing and Allowing Good." *Analysis* (2022):1–9.

Voorhoeve, Alex. "How Should We Aggregate Competing Claims?" *Ethics*, 125, no.1 (2014):64–87.

Wilkinson, Hayden. "Chaos, Add Infinitum." (unpublished manuscript).

Wilkinson, Hayden. "In Defense of Fanaticism." *Ethics*, 132, no.2 (2022):445–477.

Williams, Bernard. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press, 1981.

Woollard, Fiona. *Doing and Allowing Harm*. Oxford: Oxford University Press, 2015.