Morality Under Risk

CHAD LEE-STRONACH

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University

December 2018

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this thesis, and all sources used, have been duly acknowledged.

Signed: _____ 7 December 2018

Chad Lee-Stronach

School of Philosophy, RSSS, Australian National University

Acknowledgements

I'm humbled by how many people have given me their time, effort, and support in writing this thesis. In particular, I'm thankful to Seth Lazar and Alan Hájek for being such exceptionally dedicated, insightful, and generous supervisors over the years.

There are *many* people to whom I'm grateful for discussing or reading my ideas: Asbjørn Aagaard Schmidt, Christian Barry, Paul Bartha, Nic Di-Bella, Renée Bolinger, Chris Bottomley, Geoff Brennan, Ray Briggs, John Broome, Justin Bruner, Paula Casal, John Cusbert, Dale Dorsey, Edward Elliott, David Enoch, Helen Frowe, Adam Gastineu, Bob Goodin, Jesse Hambly, Caspar Hare, Iwao Hirose, Josef Holden, Jeff Howard, Frank Jackson, Alex King, David Kinney, Rob Kirby, Avery Kolers, Matt Kopec, Ten-Herng Lai, Holly Lawford-Smith, Christopher Lean, RJ Leland, Leon Leontyev, Hanti Lin, Matthew Lindauer, Christian List, Kirsten Mann, Daniel Nolan, David O'Brien, Philip Pettit, Theron Pummer, Wlodek Rabinowicz, Alex Sandgren, Wolfgang Schwarz, Shlomi Segall, Andrew Sepielli, Nic Southwood, Julia Staffel, Holly Smith, Matt Spike, Katie Steele, Kim Sterelny, Jeremy Strasser, Larry Temkin, Sergio Tenenbaum, Mike Titelbaum, Lachlan Umbers, Gabriel Watts, David Wiens, Andrew Williams, Timothy L. Williamson, Shang Long Yeo, Kevin Zollman, audiences of Philsoc Seminars, MSPT Graduate Workshops, ANU Graduate Reading Retreats, the Summer PhD Fellowship at the Hebrew University of Jerusalem, the Stockholm Graduate Reading Retreat, various journal editors and referees. I also want to thank our School's administrative staff: particularly, Stephanie Brinkworth, Sophie Napier, and Karen Smith. My study here at the ANU was supported by an Australian Postgraduate Award.

Finally, I want to thank Carina for her love, support, and intellectual company, Adam for trudging through my academese, and Angela, Jason, and Rhett, for believing in me all the way.

Abstract

Many argue that absolutist moral theories – those that prohibit particular kinds of actions or trade-offs under all circumstances – cannot adequately account for the permissibility of risky actions. In this dissertation, I defend various versions of absolutism against this critique, using overlooked resources from formal decision theory. Against the prevailing view, I argue that almost all absolutist moral theories can give systematic and plausible verdicts about what to do in risky cases. In doing so, I show that critics have overlooked: (1) the fact that absolutist theories – and moral theories, more generally – underdetermine their formal decision-theoretic representations; (2) that decision theories themselves can be generalised to better accommodate distinctively absolutist commitments. Overall, this dissertation demonstrates that we can navigate a risky world without compromising our moral commitments.

Contents

Introduction

0.1 Morality in a Risky World

The world is a risky place. Everyday actions like driving, making a promise, or purchasing a product can – for all we know – cause harm, be insincere, or sustain others' wrongful practices. In light of this fact, it is surprising that few moral theories provide us any systematic guidance about which kinds of risky actions are morally permissible, and which are not.[1] This is a problem: given that we are never completely certain about the nature or consequences of our actions, it seems that our moral theories are, for all practical purposes, silent.[2]

Recognising this problem, various philosophers have attempted to extend our existing moral theories to give verdicts in cases involving uncertainty. However, they have found that some kinds of moral theories – so-called *absolutist* moral theories – give unacceptable verdicts in such cases, and should therefore be rejected. I shall refer to this critique as:

The Problem of Risk: Absolutist moral theories cannot give adequate moral guidance in cases of uncertainty.

---

1. Here and throughout, I use the terms 'uncertain' and 'risky' interchangeably, assuming a broadly Bayesian approach that allows for assignments of probabilities (be precise or imprecise, objective or subjective). On these issues, see: Alan Hájek and Christopher Hitchcock, eds., *The Oxford Handbook of Probability and Philosophy* (Oxford: Oxford University Press, 2016).

2. Here and throughout, I will focus solely on empirical uncertainty: uncertainty about an action's kind or consequences, holding fixed a background moral theory. This is in contrast with moral uncertainty, which concerns uncertainty about which moral theory is true. See: Andrew Sepielli, "What to Do When You Don't Know What to Do," in *Oxford Studies in Metaethics* (2009).

In this dissertation, I argue that the Problem of Risk is false. Using overlooked resources from formal decision theory, I show how various absolutist theories can provide adequate guidance through an risky world.

The overall structure of the dissertation is straightforward. **Part 1** introduces the debate over Moral Absolutism and risk. To make the debate more tractable, I define a set of necessary and sufficient conditions for a theory to give adequate moral guidance in cases of uncertainty. I argue that most absolutist theories can satisfy these conditions. **Part 2** details ways they might do so.

0.2   What is 'Moral Absolutism'?

'Moral Absolutism' is something of a pejorative label. It is usually applied to moral theories on the basis of some perceived fanaticism or dogmatism. For instance, Michael Huemer (2010) takes the following theorists to be absolutists:

> Kant held that one must always keep one's promises, no matter how bad the consequences of doing so may be, or how much good might be brought about by breaking a promise. Elizabeth Anscombe held that it is always wrong to knowingly punish a person for a crime he did not commit. And Robert Nozick appears to have held that it is always wrong to violate an individual's negative rights against coercion. [3]

More formally, Frank Jackson and Michael Smith (2006) hold that Moral Absolutism is any theory that:

> Absolutely prohibits actions of kind K, where K ... is a property of an action as opposed to a relation between an action and available alternatives to that action:[4]

They, like others, have since expanded the category of Absolutism to include deontology (both moderate and strict forms), rights-theories, lexical priority theories, anti-aggregationist theories, and those that take some actions to be

---

3. Michael Huemer, "Lexical Priority and the Problem of Risk," *Pacific Philosophical Quarterly* 91 (2010): p. 332.

4. Frank Jackson and Michael Smith, "Absolutist Moral Theories and Uncertainty," *The Journal of Philosophy* 103, no. 6 (2006): p. 267.

unconditionally wrongful. These and other 'absolutist' theories are said to face
the Problem of Risk.

For instance, Robert Nozick (1974) discusses the Problem of Risk as it re-
lates to natural rights theories.[5] Dennis McKerlie (1986) reinforces Nozick's
worries, coming to the conclusion that rights-based theories cannot operate in
cases of risk.[6] J.E.J. Altham (1983) extends the Problem to other rights-based
theories, including contractualism.[7] Judith Jarvis Thomson (1986) argues that
theories that posit a right against risk lead to unacceptable verdicts about
whether an action is rights-violating.[8] Shelly Kagan (1991) explores how de-
ontological constraints – roughly, limitations on how we can make the world a
better place – operate in risky cases. He is unsatisfied with any approach to
dealing with risk, along similar lines to those given by McKerlie, arguing that
either deontology takes all, none, or some subset (defined by some arbitrary
threshold) of risks to be wrongful.[9] More recently, this trilemmatic approach
was finessed by Frank Jackson and Michael Smith (2006).[10] Similarly, Sven
Ove Hansson (2003, 2013) argues that almost all contemporary moral theories
face the Problem of Risk (which he calls the "Mixture Appraisal Problem").[11]
Michael Huemer (2010) argues that the Problem of Risk applies to all lexi-
cal priority theories: those that prioritise some moral considerations over any
number of particular other considerations.[12] Jackson and Smith (2016) argue
that all versions of deontology face the Problem.[13] Yoaav Isaacs (2014) agrees
and proposes, but ultimately does not endorse, a knowledge-first approach for

5. Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic, 1974).

6. Dennis McKerlie, "Rights and Risk," *Canadian Journal of Philosophy* 16, no. 2 (1986):
239–251.

7. J. E. J. Altham, "Ethics of Risk," *Proceedings of the Aristotelian Society, New Series*
83 (1983): 15–29.

8. Judith Jarvis Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory* (Cam-
bridge, Mass.: Harvard University Press, 1986), p. 177, 181.

9. Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1991), pp. 87-92.

10. Jackson and Smith, "Absolutist Moral Theories."

11. Sven Ove Hansson, "Ethical Criteria of Risk Acceptance," *Erkenntnis* 59, no. 3 (2003):
291–309; Sven Ove Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*
(Palgrave Macmillan, 2013).

12. Huemer, "Lexical Priority and the Problem of Risk."

13. Errol Lord and Barry Maguire, eds., "The Implementation Problem for Deontology,"
chap. 14 in *Weighing Reasons* (Oxford: Oxford University Press, 2016), 338–354.

deontology to avoid the Problem of Risk.[14] Holly Smith (2014) develops a specific version of the Problem of Risk, arguing that deontologists cannot, by their own lights, be required to gather more information before acting.[15] Collectively, these critiques lead us to the conclusion that a vast swathe of moral theorising cannot guide us through an uncertain world, and should therefore be rejected.

Why do these theories face the Problem of Risk? I shall argue that the root cause is structural, not substantive. By substantive, I mean *what* a moral theory considers to be important: life, love, liberty, what have you. By structural, I mean the logical relations that connect these substantive considerations. As we shall see, the Problem of Risk relies on attributing to a moral theory particular structural assumptions: specifically, assumptions that prevent the theory's substantive considerations from being sufficiently measurable or comparable. Over the course of this dissertation, I shall argue in detail that many allegedly absolutist theories are not committed to the structural assumptions that give rise to the Problem of Risk. In general, I will not enter a dispute about labels: from the vantage point of some other moral theory, each of these could well seem absolutist in some relevant sense. However, I will show that even if these moral theories are 'absolutist', that does not commit them to the Problem of Risk. In fact, I will attempt to show that practically all of the allegedly absolutist theories mentioned above can operate satisfactorily in a risky world.

0.3   Methodology

How can we determine whether absolutist theories can avoid the Problem of Risk? I believe that the history of cartography provides an answer.[16] For the longest time, maps were highly unreliable. This is because explorers often used intuition to mark the landmarks and the distances between them. Intuition is,

---

14. Yoaav Isaacs, "Duty and Knowledge," *Philosophical Perspectives* 28, no. 1 (2014): 95–110.

15. Holly M. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," *Ethics* 125, no. 1 (2014): 11–38.

16. John Noble Wilford, *The Mapmakers*, Revised Edition (New York: Vintage Books, 2000).

of course, a highly fallible guide to such matters, leading to mistaken judgements that differed from person to person, and allowing for *ad hoc* details to creep in to fill in the gaps left by intuition or observation. Jonathan Swift skewered such explorers:

> So Geographers, in Afric-maps,
> With savage-pictures fill their gaps;
> And o'er uninhabitable downs
> Place elephants for want of towns.[17]

The great advances in cartography came with the use of trigonometric methods and observational and measuring techniques. Using a set of axioms defining the relationships between distance and angles, alongside more advanced methods of measuring each, cartographers were reliably guided beyond their intuitions, providing a consistent and increasingly accurate representation of largely untouched terrain.

Today, moral philosophy still operates on the basis of intuitions about what is right and wrong under hypothetical, 'unexplored' scenarios. For the most part, however, this is a reliable approach. Unfortunately, in cases of uncertainty, this methodology is fraught: psychologists have persuasively shown that our intuitive judgements about risk are often naive or conflicting, or otherwise subject to one of many cognitive biases.[18] To advance moral theorising over these difficult cases, we must find an equivalent to the trigonometric methods used by cartographers, one that allows us to project our understanding from the firmer grounds of certainty to the cognitively distant terrains of uncertainty.

Not only do we need a method that does not rely heavily on intuition, we also need one that allows for a sufficient degree of measurability and comparability. Without a method for systematically distinguishing the *degrees* to which risky actions are right or wrong, moral philosophy will run afoul of a

---

17. ibid., p. 14.

18. We tend to overweight small probabilities, make mistakes when conjoining probabilities, mistake salience with probability, among many other errors. Amos Tversky and Daniel Kahneman, "Judgement Under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974): 1124–1131; Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus / Giroux, 2011).

basic cartographic principle, noted in the third century AD by Pei Xiu:

> If one draws a map without having graduated divisions, there is no
> means of distinguishing between what is near and what is far.[19]

I shall argue that formal theories of rational decision give us a useful way of systematically projecting our moral commitments beyond the *terra firma* of our strongly-held intuitions and into the domain of uncertainty, providing a more systematic and reliable way of charting the moral dimensions of risky cases. Specifically, I will argue that most absolutists can avoid the Problem of Risk by adopting some version of expected value theory. Expected value theory defines sets of axioms governing how we ought to treat the relationships between value, probability, and actions. Specifically, it holds that we should choose actions that maximise probability-weighted average value. In general, I will adopt broadly von Neumann-Morgenstern axioms, while also allowing for violations of its axioms of Continuity (allowing for a prohibition of particular kinds of trade-offs) and Completeness (allowing for the incommensurability of particular moral considerations).[20] At times, I will adopt a formally more general decision-theoretic approach, based on Dietrich and List (2017).[21] These formal frameworks are by no means the only ways of modelling absolutism – they simply suffice to frame the discussions and to provide positive proposals for how absolutists can avoid the Problem of Risk.

Why take this approach as opposed to some informal criterion, such as 'reasonableness' or 'foreseeability'? Firstly, expected value theory can be more *informative* in explaining why one risky action is permissible, yet another is not: this will be due to the values or probabilities (or some combination thereof) of

---

19. Wilford, *The Mapmakers*, p. 33.

20. John Von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, 3rd (1947; Princeton: Princeton University Press, 1990). Since these axioms are defined in terms of objective probabilities, I will implicitly supplement them with subjectivised axioms, as per: F. J. Anscombe and R. J. Aumann, "A Definition of Subjective Probability," *The Annals of Mathematical Statistics* 34, no. 1 (1963): 199–205. Alternative approaches that could have been used include: Leonard J. Savage, *The Foundations of Statistics*, 2nd rev. (New York: Dover, 1972); Richard C. Jeffrey, *The Logic of Decision*, Second (Chicago and London: The University of Chicago Press, 1983).

21. Franz Dietrich and Christian List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories," *The Philosophical Review* 126, no. 4 (2017): 421–479.

the various options available. Informal standards, by contrast, often obscure these variables or else 'pass the buck' by lapsing into disguised tautology: e.g. risks are permissible when they are 'reasonable' or 'acceptable'.[22]

Secondly, expected value theory is *comprehensive* in that provides verdicts for all combinations of probability and values.[23] Provided that we accept the axioms and value assignments, we can be confident in its verdicts across a range of cases which we haven't specifically considered – or indeed could not consider (with any degree of accuracy). Expected value theory is thus *ampliative* in the sense of providing verdicts beyond those already reached by existing theorising. Informal standards, by contrast, are rarely ampliative – we cannot rely on their verdicts about risky cases without having to directly rely on our questionable intuitions about them.

Thirdly, expected value theory is provably *consistent*: properly applied, it does not lead to a contradictory set of verdicts.[24] By contrast, those who avoid decision-theoretic approaches in favour of piecemeal, case-by-case principles, cannot be certain that these principles will yield a consistent set.[25]

Lastly, adopting this kind of formal approach allows us to better understand the commitments of moral theories and the sources of disputes between them. If there is disagreement about the acceptability of a model's verdicts, then the dispute lies in the assignment of values or in one or more of the axioms. In this way, the decision-theoretic approach helps to isolate the points of dispute. As we shall see, it also helps to reveal previously overlooked aspects of the moral

---

22. For example, Scanlon (2008) appeals to 'reasonable' belief to justify harm done in self-defence, but does not explain what a reasonable belief is. Hansson (2003) explictly rejects expected value approaches to risk, instead taking 'acceptable' risks to be those that are part of a 'persistently justice-seeking' social system. Unfortunately, he does not venture to say what it means for a risk to be 'part of' such a system. T.M. Scanlon, *Moral Dimensions* (Cambridge Massachusetts, and London, England: Belknap Press of Harvard University Press, 2008), p. 67; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*, Ch. 6.

23. Okay, maybe not *all* possible combinations: Harris Nover and Alan Hájek, "Vexing Expectations," *Mind* 113, no. 450 (2004): 237–249.

24. This has been proven by a range of representation theorems, which I need not discuss here. See: Von Neumann and Morgenstern, *Theory of Games and Economic Behavior*; Savage, *The Foundations of Statistics*; Jeffrey, *The Logic of Decision*.

25. On this point, see: Chapter 1, Section 3, Subsection 1; Chapter 4, Section 4.

theories we will consider.[26]

Based on these methodological considerations, I will adopt an approach known as *decision-theoretic ethics*.[27] This method involves first determining whether a given moral theory is logically compatible with some version of decision theory. If it is, then the challenge is to demonstrate that modelling the moral theory using decision theory does not lead it to give unacceptable verdicts about the permissibility of risky actions.[28] From the point of view of this method, the Problem of Risk can be understood as the claim that some types of moral theories are either incompatible with decision theory or else give palpably incorrect results when they are extended, via decision theory, to cover risky cases.

Before continuing, however, I should note that others have come to the defence of these various moral theories, many using broadly decision-theoretic approaches. For example, following arguments that rights-based theories have difficulty accommodating risk, David McCarthy (1997) argues that such theories are compatible with orthodox decision theory.[29] However, it seems that this response has been largely overlooked.[30]

Deontologists, more generally, have recently begun exploring these limits of expected value theory in modelling their distinctive moral commitments, such as agent-centred perogatives and options.[31] Some have argued that deontolo-

---

26. Thanks here to Christian List.

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

28. Some moral theorists appear to believe that it is open to them to reject decision theory. However, I will attempt to show that this involves rejecting very minimal and compelling axioms of moral decision-making. More generally, I hope to show that the conflict between some kinds of moral theories and expected value theory has been exaggerated, such that moral theorists do not need to give up expected value theory at all.

29. David McCarthy, "Rights, Explanation, and Risks," *Ethics* 107, no. 2 (1997): 205–225.

30. For instance, see the literature on the Paralysis Problem: Madeleine Hayenhjelm and Jonathan Wolff, "The Moral Problem of Risk Impositions: A Survey of the Literature," *European Journal of Philosophy* 20, no. 51 (2011): E26–E51; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*. For what seems to have been the only published response to McCarthy's proposal, see: Stephen Perry, "Risk, Harm, Interests, and Rights" [in eng], in *Risk: Philosophical Perspectives*, ed. Tim Lewens (Routledge: Oxford University Press, 2007).

31. Seth Lazar, "Deontological Decision Theory and Agent-Centered Options," *Ethics* 127, no. 3 (2017): 579–609; Seth Lazar, "In Dubious Battle: Uncertainty and the Ethics of Killing," *Philosophical Studies*, 2017, 1–27; Seth Lazar, "Anton's Game: Deontological

gists should reject expected value theory entirely.[32] I will not explore whether existing decision-theoretic models can accommodate agent-centred perogatives and options, but I will suggest that at least some deontological theories can benefit from the generalised decision-theoretic approaches I canvass here.

Others have attempted to model broadly absolutist theories without rejecting any axioms of expected value theory.[33] However, such approaches often require complex interpretations of the formal framework to prevent violations of the axioms.[34] For instance, deontological theories that try to hew too closely to orthodox decision theory (usually, by some acrobatic book-keeping) arguably trivialise the model, making it difficult to give principled determinations of whether individuals ever violate deontological morality.[35]

As we shall see, my approach often differs from the above by taking a middle-ground between all-out rejecting or accepting orthodox decision theory. I often adopt generalised models that allow for violations of some of the standard axioms of expected value theory or Bayesian epistemology. My contributions can thus be read as supplementing, rather than supplanting, those made elsewhere. In general, I will adopt a methodologically conservative approach of attempting to provide a decision-theoretic formal model that minimally departs from orthodox expected value theory, without trivialising the model. This approach optimises the explanatory power and systematicity of decision theory, providing moral theories an ampliative structure, without compromising their underlying substantive commitments.

Decision Theory for an Iterated Decision," *Utilitas* 29, no. 1 (2017): 88–109; Seth Lazar, "Risky Killing: How Risks Worsen Violations of Objective Rights," *Journal of Moral Philosophy*, 2017, 1–30; Seth Lazar, "Limited Aggregation and Risk" (2018), 1–28; Douglas W Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011), xx, 266 p.

32. Sergio Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model," *Ethics* 127 (2017): 1–36; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*.

33. Graham Oddie and Peter Milne, "Act and Value: Expectation and the Representability of Moral Theories," *Theoria*, 1991, Mark Colyvan, Damian Cox, and Katie Steele, "Modelling the Moral Dimension of Decisions," *Noûs* 44, no. 3 (September 2010): 503–529

34. Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford; New York: Oxford University Press, 2012), Ch. 13.

35. John Broome, *Weighing Goods* (Cambridge, Massachusetts: Blackwell, 1995), pp. 103-106; Lazar, "Deontological Decision Theory and Agent-Centered Options."

0.4   Outline of the Dissertation

This project is broadly a positive one in the sense that it *negates* negative results that have been pressed against absolutist moral theories. This involves presenting a series of 'proofs of possibility' in response to particular discussions where some or other theory is believed to be subject to the Problem of Risk. I take these results to show that accommodating risk in our moral theories does not require us to make any important substantive concessions about what matters, morally-speaking, or why. The dissertation has two main parts:

0.4.1   Part 1: The Problem of Risk

In Chapter 1: Prohibition and Probability, I give a formal representation of the structural features of a moral theory that make it vulnerable to the Problem of Risk. I establish that the Problem only applies to a very particular version of Moral Absolutism, namely: Option Absolutism, which holds that some actions are prohibited irrespective of their alternatives. I then argue that few, if any, existing absolutist theories are of this kind. Instead, such theories are better understood as subscribing to Relational Absolutism, which holds that an action's being prohibited depends on its relation to other available actions. I show that Relational Absolutist theories can potentially avoid the Problem of Risk. This clears the way for positive proposals in later chapters, which detail how various other absolutist theories might avoid the Problem of Risk.

0.4.2   Part 2: Solutions

In light of the possibility result given in Chapter 1, I set out to develop various 'proofs of possibility' that show how particular moral theories can potentially operate in risky cases, despite their being 'absolutist' in some broad sense.

Chapter 2: Moral Priorities Under Risk defends lexical priority theories using lexicographic expected value theory.[36] This chapter illustrates that lexical priority theories can be understood as being Relational Absolutist. It also

---

36. This chapter is very close to its published version: Chad Lee-Stronach, "Moral Priorities Under Risk," *Canadian Journal of Philosophy*, 2017, 1–19

identifies and responds to various challenges to lexical priority theories.

Chapter 3: Priorities and Uncertainties takes up a challenge identified in my co-authored article with Seth Lazar.[37] In that article, we argue that standard expected value theory can allow lexical priority theories to avoid the Problem of Risk, so long as there exists a principled account of whether individual or sequences of actions are the proper objects of moral evaluation. In this chapter, I attempt to give such an account.

Chapter 4: Duty and Ignorance considers a different version of the Problem of Risk, given by Holly Smith (2014), which I call the Problem of Ignorance.[38] It concerns cases where a moral theory makes incorrect verdicts about whether you ought to gather more information before acting. I respond to Smith's arguments that deontological theories face the Problem of Ignorance. I present a decision-theoretic model of deontological theories that avoids the Problem.

Lastly, Chapter 5: Authority, Obedience and Uncertainty considers the Problem of Risk in relation to the literature on legitimate authority: specifically, Joseph Raz's Service Conception of authority.[39] Raz's theory is relevant because it appears to exhort an absolutist-like obedience to legitimate authority. I argue that Raz's theory can be satisfactorily extended to cases of uncertainty using expected value theory. I then respond to objections to decision-theoretic approaches to legitimate authority.[40]

These chapters are designed to be separate interventions into particular debates over the Problem of Risk. Taken together, they illustrate that the Problem of Risk is a productive challenge to moral theories, but one that is not likely to be successful. Once we clarify a moral theory's substantive commitments and take a full view of the decision-theoretic resources available, we can see that most – if not all – 'absolutist' theories can guide us through a risky world.

Having charted the course forward into the terrain of uncertainty, we must

---

37. Seth Lazar and Chad Lee-Stronach, "Axiological Absolutism and Risk," *Noûs*, 2017, 1–17.

38. Smith, "The Subjective Moral Duty to Inform Oneself before Acting."

39. Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986).

40. Scott J. Shapiro, *Authority*, ed. Jules L. Coleman, Kenneth Einar Himma, and Scott J. Shapiro (Oxford University Press, 2004), 382–439.

now explore these relatively untouched grounds up close.

Onwards!

Part I

The Problem of Risk

Chapter 1

Prohibition and Probability

1.1 Introduction

Most contemporary moral theories hold that the nature or consequences of an action can render it morally permissible or impermissible.[1] Absolutist moral theories hold that some actions are *always* morally impermissible. For example, the following theorists are taken to be prototypical moral absolutists:

> Kant held that one must always keep one's promises, no matter how bad the consequences of doing so may be, or how much consideration might be brought about by breaking a promise. Elizabeth Anscombe held that it is always wrong to knowingly punish a person for a crime he did not commit. And Robert Nozick appears to have held that it is always wrong to violate an individual's negative rights against coercion.[2]

In recent years, many have argued that absolutist theories have very little to say about what we ought to do when we are, to some degree, *uncertain* about whether a given action is of the absolutely prohibited kind.[3] Moreover, they have argued that when such theories are extended to cover such cases, they seem to give incorrect verdicts. I shall call this critique:

The Problem of Risk: Absolutist moral theories cannot give adequate moral

---

1. For a recent theory that avoids such moral permissibility verdicts, see: Caspar Hare, *The Limits of Kindness* (Oxford: Oxford University Press, 2013).

2. Huemer, "Lexical Priority and the Problem of Risk," p. 332.

3. For example, see: Jackson and Smith, "Absolutist Moral Theories."

guidance in cases involving risk.

The Problem of Risk has been presented as a challenge for various, broadly absolutist, moral theories.[4] However, it has never been precisely explained why some moral theories seem to encounter the Problem, while others seem to avoid it. This paper identifies a set of structural assumptions that give rise to the Problem of Risk. It argues that the Problem of Risk – contrary to the prevailing view in moral theory – is actually only a problem for very few, if any, existing moral theories.

Part 1.2 presents the Problem of Risk. **Part 1.3** discusses two representative responses that have been made on behalf of Moral Absolutism. **Part 1.4** sets out a choice-theoretic framework that formalises the Problem of Risk as an impossibility result for moral theories that have a particular structure. It also sets out a possibility result that shows how moral theories can potentially avoid the Problem of Risk. **Part 1.5** discusses whether any existing moral theories necessarily face the Problem of Risk. **Conclusion** follows.

1.2   Absolutism and the Problem of Risk

To illustrate the Problem of Risk, consider the following scenario:

---

4. For instance, the Problem of Risk led Judith Jarvis Thomson (1986, 1990) and Dennis McKerlie (1986) to conclude that we do not have rights against having risk imposed on us. See: Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory*; Judith Jarvis Thomson, *The Realm of Rights* (Harvard University Press, 1990); McKerlie, "Rights and Risk." More recently, Sven Ove Hansson argues that a wide range of moral theories, from Consequentialism to Deontology, are unable to cope with risk and uncertainty. See: Hansson, "Ethical Criteria of Risk Acceptance"; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*. Michael Huemer argues that lexical priority theories – those that take some moral considerations to be more important than any number of 'lesser' considerations – also face the Problem of Risk. See: Huemer, "Lexical Priority and the Problem of Risk." Yoaav Isaacs (2014) presents the Problem for deontology, attempts solve it with a knowledge-first epistemology, then appears to disavow the solution. See: Isaacs, "Duty and Knowledge." Frank Jackson and Michael Smith (2006) raise the Problem for Moral Absolutism. See: Jackson and Smith, "Absolutist Moral Theories." They subsequently argue that the Problem extends to Moderate Deontology, which allows that, in extreme cases, deontological prohibitions (such as murder, or lying) can justifiably be violated so long as doing so would promote a sufficiently large amount of good. See: Lord and Maguire, "The Implementation Problem for Deontology."

Patient: You are a medical doctor in the emergency ward and you have a patient who needs urgent medical attention. Your options are: either give her Drug A or give her Drug B. The patient is either receptive to Drug A or receptive to Drug B. If the patient is A-Receptive, then giving Drug A will cure her, whereas giving Drug B will harm her. If, instead, the patient is B-Receptive, then Drug B will cure her and Drug A will harm her. You are uncertain whether the patient is A-Receptive or B-Receptive.[5]

Suppose that you are a Hippocratic absolutist: you believe that it is always morally impermissible to harm your patient. In cases of certainty, this moral norm gives you clear guidance about what to do: namely, don't harm your patient. In the present case, however, it is unclear which action is of the prohibited, harmful kind. As a Hippocratic absolutist, what are you to do, given your uncertainty?

To answer this, you need a *moral decision rule*: a rule that determines what you morally ought to do, given your options and your uncertainty. In their critique of absolutist moral theories, Jackson and Smith (2006) consider three candidate absolutist moral decision rules. However, they quickly show that each rule is highly problematic.

The first candidate rule, **Positive Probability**, prohibits any actions that have a *positive probability* of violating an absolutist moral norm. A problem with this rule is that, under our everyday conditions of uncertainty, it would prohibit all of our actions.[6]

The second candidate rule, **Certainty**, takes the opposite approach and encounters the opposite problem: it prohibits all and only those actions that are *certain* to violate a moral norm, but thereby permits practically all of our actions. The problem with this, of course, is that there seem to be many cases where an action is impermissible, even though it is not certain to violate a moral prohibition.

---

5. This is a simplified version of a case given in: Jackson, "Decision-Theoretic Consequentialism."

6. This is known as the 'Paralysis Problem'. See: Altham, "Ethics of Risk"; Hayenhjelm and Wolff, "The Moral Problem of Risk Impositions: A Survey of the Literature"; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*.

The third and final candidate rule that Jackson and Smith consider, **Threshold**, says: only perform actions that are less than t-probable to violate an absolutist moral norm. In cases like **Patient**, they consider a rule that says you ought to treat the patient if and only if doing so is less than t-probable to harm her. However, Jackson and Smith argue that this moral decision rule yields inconsistent verdicts. For example, consider the following case:

Second Patient: A second patient has just arrived, who presents the same symptoms and is just as likely as the first patient to be A-Receptive or B-Receptive. You must now decide whether to give the same drug to each patient, or to give them different drugs.

Suppose that for each patient the probability that your administering any particular drug will cause harm is less than t. As such, your moral decision rule holds that you ought to treat each patient, either with Drug A or Drug B.

However, suppose that if you treat both patients, the cumulative probability that at least one patient will be harmed is greater than t. According to your moral decision rule, you are therefore prohibited from administering the drug in both instances, even though each instance is required.

Thus, it seems that Threshold yields inconsistent verdicts about what you morally ought to do: you morally ought to treat each individual patient, but you ought not treat both. Jackson and Smith consider this an "especially implausible kind of moral dilemma" and, on that basis, reject Threshold.[7] For want of any obvious alternative moral decision rule, but also having not exhaustively investigated all of the possible moral decision rules that absolutists might accept, they make the following conjecture: Moral Absolutism cannot give adequate moral guidance in cases involving risk.

---

7. Jackson and Smith, "Absolutist Moral Theories," p. 276. For responses, see Ron Aboodi, Adi Borer, and David Enoch, "Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith," *Journal of Philosophy* 105, no. 2 (2008): 259–272; Patrick Hawley, "Moral Absolutism Defended," *Journal of Philosophy* 105, no. 5 (2008): 273–275; Lazar, "In Dubious Battle: Uncertainty and the Ethics of Killing"; Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model." At least some of these responses fail due to the impossibility result presented in Part 3.

If this conjecture is correct, then there is a strong case for rejecting Moral Absolutism. If we reject Moral Absolutism, then it turns out that no kinds of actions are always morally impermissible.

1.3 Responses to the Problem of Risk

Since Jackson and Smith did not prove the Problem of Risk, it has been open to defenders of Absolutism to offer proposals for avoiding it. I will review two such proposals below. As I shall argue, each of these proposals is problematic in its own right. However, to forestall ongoing debate, Section 1.4.4 proves that Jackson and Smith (2006) are correct: there is a particular type of absolutist moral theory that incorrectly determines the permissibility of risky actions. To the extent that defenders of Moral Absolutism are defending this type of theory, their responses fail.

1.3.1 A Rights-Based Approach

Ron Aboodi, Adi Borer, and David Enoch (2008) argue that a particular class of moral absolutist theories can adopt Threshold without generating inconsistent permissibility verdicts.[8] This class of theories is *patient-centred, rights-based, individualistic deontology.* A moral theory is 'patient-centred' when it grounds the permissibility of actions on their effects on other people; it is 'rights-based' when it holds that our moral duties are determined by the rights of others; it is 'individualistic' when it holds that only individuals – not groups – possess such rights. Many moral theorists appear to subscribe to this kind of view. As such, if it does avoid the Problem of Risk, then Jackson and Smith's critique is less important than previously thought.

In **Second Patient**, Aboodi et al.'s account holds that you ought to treat each patient, since each of your actions is not sufficiently probable to violate each patient's right against being harmed. Since it is permissible to treat each patient, it is permissible to treat both, even though the cumulative risk is

---

8. Aboodi, Borer, and Enoch, "Deontology, Individualism"

above *t*. This is because there is no group agent consisting of 'Patient A-and-Patient B' that possesses the right against incurring a sufficiently high risk of harm. Thus, it seems that Jackson and Smith's critique is mistaken: there are absolutist moral theories that can give adequate advice in cases of uncertainty.

However, there are various problems with this particular approach.[9] More generally, there are problems with the general strategy of refining a theory's substantive moral commitments in order to avoid what is, as I shall argue, a structural problem. Simply put, the Problem of Risk does not concern *what* a moral theory believes to be morally important or *why*, but rather *how* it models the relations between these substantive elements.

In terms of their specific proposal, it is clear that – to the extent that they are defending Moral Absolutism – Aboodi et al. must deny the existence of group rights in *all* cases, since Threshold will give inconsistent verdicts when there are both group rights and individual rights at play. To avoid this inconsistency, they would need to abandon the absolutist idea that rights violations are always prohibited, or else explain why group rights can never come into conflict with individual rights.

Leaving this aside, there are still other commitments they must make, because it turns out that denying the existence of group rights is not sufficient to avoid inconsistencies. To see why, note that there are two ways of understanding patient-centred duties: on one reading, such duties are owed to particular, specified people (*de re*); on another reading, they are duties to people in general, whoever they happen to be (*de dicto*). On a *de re* reading, Aboodi et al.'s solution works.[10] If, however, Aboodi et al. opt for a *de dicto* reading, then their account gives the opposite verdict to the one they presented: your treating both patients would now be impermissible, because it has a sufficiently high probability of harming *someone* (whichever patient it happens to be).

---

9. For other critical discussions of Aboodi et al.'s proposal, see: Lazar, "In Dubious Battle: Uncertainty and the Ethics of Killing"; Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model"

10. However, this opens them to other theoretical problems, such as the Non-Identity Problem: cases where it seems that we are entitled to choose avoidably harmful options, since the particular people we harm would not otherwise exist. See: Caspar Hare, "Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?," *Ethics* 117, no. 3 (2007)

This means that to retain their initial verdict, Aboodi et al. must not only deny group rights, but also adopt a strictly *de re*, patient-centred theory. However, this leads to other cases of inconsistency, calling for further substantive refinements. Consider the following case, based on Jackson and Smith (2016):[11]

Multiple Treatments: A third patient has arrived. She has three life-threatening conditions, each of which can be treated as follows:

Condition	Treatment
1	Drug A (if A-Receptive) or Drug B (if B-Receptive)
2	Drug C (if C-Receptive) or Drug D (if D-Receptive)
3	Drug E (if E-Receptive) or Drug F (if F-Receptive)

Suppose that there are no interaction effects between the drugs, but you are uncertain about the receptivity of the patient to particular drugs. If the probability that any particular drug will harm your patient is less than t, then Threshold permits you to administer any particular combination of drugs. At the same time, however, suppose that over the course of the three treatments, the cumulative probability of harm is greater than t. Assessing your actions in terms of individual treatments renders your actions permissible, whereas assessing them collectively renders them impermissible.

This case shows that denying the existence of group rights and committing ourselves to understanding duties *de re* as opposed to *de dicto* does not help us to avoid the inconsistency. What to do? Following Aboodi et al.'s approach, we could make further substantive commitments to ensure that we always get the right verdicts. For instance, given a particular metaphysical theory of action-individuation, we may determine whether your treating the patient is a single action (as in the coarse-grained description: 'Giving the A/D/E regime of drugs') or multiple actions (as in the fine-grained description: 'Give Drug A; Give Drug D; Give Drug E'). Adopting a coarse-grained or fine-grained approach to action-individuation may avoid this version of the

---

11. Lord and Maguire, "The Implementation Problem for Deontology." This type of case is also credited to David Lewis, in McKerlie, "Rights and Risk," n. 3

dilemma. However, to some, making this kind of metaphysical commitment without any independent explanation would be objectionably *ad hoc*. Also, as it stands, it seems that we can expect still other cases of inconsistencies to arise, where aggregated and disaggregated moral considerations come into conflict, requiring further 'just-so' commitments.[12]

At best, the strategy of adopting further substantive refinements seems to be an overcommittal, though potentially successful, approach to avoiding the Problem of Risk. At worst, this strategy risks having to play a never-ending game of philosophical 'Whack-A-Mole', with a structurally-identical problem re-emerging under increasingly complex permutations of substantive commitments. While it is important for a moral theory to be clear and consistent about its substantive commitments, it is worth considering whether the Problem of Risk can be solved by clarifying its structural commitments. To see this, we will next consider a different defence of Absolutism, one which also adopts the crucial, problematic structural assumption that generates the Problem of Risk.

1.3.2 A Sequential Decision Rule

Patrick Hawley (2008) argues that absolutists can avoid generating inconsistent verdicts by adopting the following sequential moral decision rule:

Eliminate: Do not perform actions that are greater than t probable to violate a prohibition.

---

12. For example, a proponent of this strategy must also take a stance on metaphysical questions about, for instance, personal identity: consider the metaphysical thesis called the Memory Criterion: a person is the same across time if and only if she can remember her previous experiences. Now, given a choice between imposing risk on a person within a short (memorable) time-frame or a long (not-memorable) time-frame, unless the deontologist decides on the validity of the Memory Criterion, it will turn out that imposing a series of risks actions across a long-time will be both permissible (because, accepting the criterion, the risks do not accumulate for any single person) and impermissible (because, denying the criterion, the risks do potentially accumulate for a single person). Likewise, Fission Cases – where psychological continuity of a person is 'branched' to multiple bodies, and some small quantum of risk is imposed on each – will generate further inconsistencies, unless the deontologist comes down on an answer, once and for all, about these metaphysical problems of personal identity. Needless to say, this seems like a laborious approach to avoiding the Problem of Risk.

Decide: Choose from the remaining options (perhaps by using some other decision rule).

Applying this approach to **Second Patient**, you should eliminate any option that has a higher than t probability of harming an individual (Hawley seems to assume that the moral duty is interpreted *de dicto*, not *de re*).[13] Since treating both patients is greater than t probable to lead to you harming a patient, you do not have the option of treating both. This means that you must treat only one of the patients. Hawley suggests performing an action that maximises expected utility.

Unfortunately, this proposal does not prevent implausible moral dilemmas arising. After all, in **Second Patient**, you do not have the option of treating only one patient. You must treat both, and by stipulation, whatever treatment you choose will be sufficiently low-risk for the particular recipient, but whatever combination you choose will be too risky and therefore impermissible.

It is important to note that this problem generalises: for any probability threshold t, if the number of states is greater than $1/(1-t)$, it will be possible that the threshold-based decision rule will consider each action permissible, but the aggregate of actions impermissible (and *vice versa*).[14] For instance, suppose that we continue using the given threshold of $t = 0.95$. This means that any action that is greater than 0.95 probable of violating an absolutist moral norm is impermissible. Suppose that the relevant absolutist norms are: do not harm; do not use contraindicated drugs (these are drugs that are more likely than other drugs to harm the patient – using them would impose unnecessary risk on your patient).

Many Patients: Now there are thirty patients presenting the same symptoms. You must decide whether to treat them with the same drug (in which case, it would be Drug A or Drug B), or Drug A for some and Drug B for others. According to the blood tests, every patient is much more likely to be A-Receptive (0.9) than B-Receptive (0.1).

---

13. Hawley, "Moral Absolutism Defended," p. 274

14. This is because Threshold, by its structure, commits Absolutism to moralised versions of the Lottery Paradox, as noted in Jackson and Smith, "Absolutist Moral Theories," n. 10.

Given the information above, the probability that *all* of the patients are A-Receptive is only 0.04.[15] This means that treating all of the patients with Drug A is more than 0.95 probable to harm at least one of the patients. According to the Eliminate step of the sequential decision rule, you are prohibited from treating all patients with Drug A – indeed, it is "unavailable for rational deliberation".[16] The same applies, however, to the option of treating all patients with Drug B, since it is even less probable that all thirty patients are B-Receptive. However, you are also morally prohibited from treating some patients Drug A and some patients Drug B, since you are sure that Drug B is strongly contraindicated for each patient. This means that you face a moral dilemma. But this seems implausible: after all, if you have the best interests of each patient at heart, then you would give each patient Drug A.[17] Thus, it seems that Hawley's sequential decision rule does not address the underlying source of the Problem of Risk.

As we shall soon see, these responses to the Problem of Risk for Moral Absolutism fail because, in fact, there exists no adequate moral decision rule for the kinds of absolutist theories that Aboodi et al. and Hawley are purporting to defend. To show this, we must first formalise the Problem of Risk.

1.4   Formalising the Problem of Risk

So far, discussions of the Problem of Risk for Moral Absolutism have only been described it in informal terms. We have only been given a partial explanation of what is 'Moral Absolutism', what it means to 'give adequate moral guidance', and, indeed, what exactly are 'cases involving risk'. In this section, I will draw on the work of Dietrich and List (2017) to formalise the Problem of Risk.[18] This will allow us to determine, in more precise terms, which moral theories face the Problem of Risk and which ones avoid it.

---

15. Calculated as: $0.9^{30} \approx 0.04$

16. Hawley, "Moral Absolutism Defended," n. 2

17. See also Caspar Hare, "Should We Wish Well to All?," *The Philosophical Review* 125, no. 4 (2016): 451–472

18. Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

1.4.1   A Choice-Theoretic Representation

The Problem of Risk applies to cases of moral decision-making in the face of uncertainty. To better understand such cases, we will define the following general formal framework.[19]

   We will be concerned with what a moral theory says about the permissibility of an action in a given choice context. A *choice context k* consists of a set of actions, denoted $[k]$.[20] An action x is an element of $[k]$. In situations of moral decision-making under risk, actions are formally defined as functions from a set of *states S* to a subset of possible *outcomes O*.[21] In cases of risk, an action will have various potential outcomes, depending on which state of the world happens to obtain.[22] For instance, giving the patient Drug A in the state of the world in which she is A-Receptive will cure her, while doing so in the state of the world in which she is B-Receptive will not.[23] Note that outcomes need not be interpreted as *causal consequences*; they can equally be interpreted as fixing the *nature* of the action performed.[24] We will assume that we can measure the probability that the world is one way or another.[25] This probability could be interpreted subjectively (as representing our degrees of belief) or objectively (as evidential probabilities or chances). To simplify the analysis, we will assume that the possible states of the world are probabilistically independent of the action chosen: for instance, the fact that you give Drug A does not change the probability that the patient is A-Receptive.[26] Given a probability assignment to the possible states of the world, each of

---

19. For clarity and consistency, in my explanations of much of the formal framework below, I closely follow: Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

20. Let $\mathscr{K}$ be the set of all possible choice contexts. A choice context, k, is an element of $\mathscr{K}$. Within a choice context, you have a set of available actions that you can perform, denoted by $[k]$. $[k]$ is a subset of the universal set X of possible actions.

21. Savage, *The Foundations of Statistics*.

22. In a given choice context, there is a finite set of possible states of the world, $S = \{1, ..., n\}$, which are mutually exclusive and jointly exhaustive.

23. Formally, outcomes are action-state pairs, $O_{xn} = \langle x, S_n \rangle$, which represent an action having been performed in a particular state of the world.

24. Colyvan, Cox, and Steele, "Modelling the Moral Dimension of Decisions," p. 511.

25. We will assume that in a well-defined choice context, the probability of the states sums to 1.

26. For discussion, see: Jeffrey, *The Logic of Decision*, pp. 8-9.

your actions has a *prospect*, which identifies the set of outcomes an action could bring about, as well as their respective probabilities.[27]

In moral decisions, we are concerned with the morally relevant *properties* P of an action in a particular context.[28] There are different kinds of properties that a moral theory might take to be relevant to the evaluation of an action's permissibility:

- A property P is an *option property* if its possession by the action in a particular context depends only on the action, not on the context.[29]

- A property P is a *context property* if its possession by an action in a particular context depends only on the context, not on the action.[30]

- A property P is a *relational property* if its possession by an action in a particular context depends on both the action and the context.[31]

Aside from identifying the set of morally relevant properties in a given choice context, a moral theory also tells us how an action's properties determine its permissibility. This will involve some kind of ranking of properties and sets of properties. As Dietrich and List note, different moral theories will adopt different types of rankings, depending on how they understand the relations between different types of moral properties.[32] For example, a ranking that is based on a *weighing relation* will presume that all moral properties are comparable – being either more, less, or equally weighty as, each other – leading

---

27. Formally, an action's prospect is the set of its possible outcome-probability pairs in a context $\lambda(x, K) = \{\langle O_{x1}, \pi_1 \rangle, ..., \langle O_{xn}, \pi_n \rangle\}$. The universal set of choice contexts $\mathscr{K}$ includes the set of all possible combinations of actions and their prospects.

28. Formally, a property P is a primitive object that picks out a set of action-context pairs $\langle x, k \rangle$, called the *extension* of P and denoted $[P]$. An action x in a particular context k *possesses* P when $\langle x, k \rangle \in [P]$. Note that the use of properties is a particularly important aspect of Dietrich and List's framework, allowing them to define a more general choice-theoretic framework than that of expected utility theory or other decision theories. See: Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories," p. 431.

29. Formally, for all x in X and all k, k' in $\mathscr{K}$, $\langle x, k \rangle \in [P]$ if and only if $\langle x, k' \rangle \in [P]$.

30. Formally, for all K in $\mathscr{K}$ and all x, x' in X, $\langle x, k \rangle \in [P]$ if and only if $\langle x', k \rangle \in [P]$.

31. Formally, such properties are those that are neither option properties nor context properties.

32. Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

to a transitive and complete ranking. This approach suits most consequentialist theories, but it is not fitting for some deontological theories.[33] A less committal ranking is one that is based on a *defeat relation*, which represents which sets of moral properties are more choice-worthy than others, while also allowing that the overall ordering of sets of moral properties may be incomplete or intransitive. In what follows, to ensure that we are not loading the deck against any particular moral theory, we will use a defeat relation rather than a weighing relation.[34]

The ranking of morally relevant properties entails a *moral decision rule*, which classifies actions as permissible or impermissible based on their respective bundles of morally relevant properties.[35] Using a ranking that is based on a defeat relation, a moral decision rule will hold that an action is permissible if and only if its set of properties is undefeated; an action is impermissible if and only if its set of properties is defeated by the set of properties of some other action.[36]

It is important to note that a moral decision *rule* is not a moral decision *procedure*. A *decision procedure* is a method for determining what ought to be done, consisting of a cognitive procedure, checklist or set of 'rules of thumb'. Decision procedures are adequate if they are reliable, accurate, efficient ways of determining what ought to be done. However, it is a *decision rule* that defines

---

33. For instance, deontological theories often allow that we have moral options to serve our own interests, even if doing so does not maximise the Good. In doing so, they deny that our interests are equally weighty as or more weighty than the Good we forego. See: Theron Pummer, "Whether and Where to Give," *Philosophy and Public Affairs* 44, no. 1 (2016): 77–95; Lazar, "Deontological Decision Theory and Agent-Centered Options."

34. A weighing relation or defeat relation, denoted $\unrhd$, over sets of properties S is a binary relation whose relata are subsets of the universal set of properties, $\mathscr{P}$. When one set of properties S stands in this relation to another set S', formally $S \unrhd S'$, then S weakly outweighs S', or S is ranked weakly above S', or S weakly defeats S'. Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories," p. 432.

35. In Dietrich and List's terminology, a Moral Decision Rule corresponds to a Rightness Function.

36. Formally, where R is a function from k to $R(k)$ and $R(k)$ is a subset of $[k]$, a defeat relation is defined as: $R(k) = \{x \in [k] : \text{it is not the case that the morally relevant properties}$ of any action y in context k strictly defeat the properties of x in k. As Dietrich and List note (p. 468, n. 70), this could be further weakened to allow a moral decision rule to select the action that defeats the most alternatives, even though it is defeated by some.

what ought to be done. Decision procedures thus aim to 'track' the verdicts of a decision rule. While both elements are important to moral decision-making, the Problem of Risk must be understood as the problem of determining what kinds of moral decision rules are logically compatible with a given background moral theory.[37]

1.4.2 Modelling Moral Absolutism

Moral Absolutism holds that there are some properties of actions that make that an action always morally impermissible. More precisely:

Moral Absolutism: Some kinds of actions have a particular wrong-making property $\bar{P}$, such that they are always impermissible.[38]

Although this definition helps to clarify the structure of Moral Absolutism, it needs to be further refined. This is because, as Jackson and Smith (2006) note, there is a trivial sense in which almost all moral theories are absolutist:

> Here we need to understand $\bar{P}$ as a property of an action as opposed to a relation between an action and available alternatives to that action. Classical utilitarianism absolutely prohibits doing actions that fail to maximize utility. But an action's failing to maximize utility is a relation the action has to available alternatives. The distinctive feature of the kind of absolutism that we find, for example, in Kant and the Catholic tradition is that the absolutely prohibited kind is independent of the nature of any available alternatives.[39]

---

37. On this distinction, see: R Eugene Bales, "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?," *American Philosophical Quarterly* 8, no. 3 (1971): 257–265; Holly M. Smith, "Making Moral Decisions," *Noûs* 22, no. 1 (1988): 89–108.

38. Formally, if $\langle x, K \rangle \in [\bar{P}]$, then $\langle x, K \rangle \notin R(K)$. In terms of the defeat relation, actions that have the prohibited property are ranked lower than all actions that do not have that property. Also, to capture idea that such actions are always impermissible, the defeat relation must be nonreflexive, such that it does not rank $\bar{P}$ at least as highly as itself in cases where all of your available actions contain $\bar{P}$. Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories," p. 458.

39. For notational consistency, here I have replaced Jackson and Smith's 'K' for $\bar{P}$. Jackson and Smith, "Absolutist Moral Theories," pp. 267-268. See also: Alan Gewirth, "Are There Any Absolute Rights?," *The Philosophical Quarterly* 31, no. 122 (1981): p. 4.

Following the earlier distinction between types of properties, we can capture the distinction between these kinds of Absolutism within the choice-theoretic framework as follows:

Option Absolutism: Whether an action in a particular context possesses the absolutely prohibited property $\bar{P}$ depends only on the action, not on the context.

Contrast Option Absolutism with:

Relational Absolutism: Whether an action in a particular context possesses the absolutely prohibited property $\bar{P}$ depends on both the action and the context.

Relational Absolutism can include comparative considerations in the definition of the $\bar{P}$. For instance, whether or not an action possesses $\bar{P}$ may depend on it having the least number of possible violations of a moral norm; or the lowest probability-weighted average value of norm violations; and so on.

1.4.3   Adequacy Conditions

As we saw earlier, some moral decision rules seem to be inadequate guides for risky situations. However, discussions of the adequacy conditions for a moral decision rule have tended to be elliptical and informal, leaving it unclear whether there could be some other moral decision rule that fits the bill. I will now identify and formalise a minimal set of adequacy conditions for moral decision rules that together generate an impossibility result for Option Absolutism.

To begin, the fact that a moral theory possesses a moral decision rule at all entails that it satisfies the following adequacy condition:

Action Guidance: For all choice contexts k and all actions x, x is either permissible or impermissible.[40]

---

40. Formally, for all k in $\mathscr{K}$ and for all x in $[k]$, x is either in $R(k)$ or not in $R(k)$.

This condition falls out of the fact that a moral decision rule is a function that maps an available action to either the subset of permissible actions or to the subset of impermissible ones. This may seem like an overly strong condition, since it entails that a moral theory is never 'silent' about the permissibility of some action. After all, there are many cases where it seems that moral judgements simply do not apply (we do typically not expect moral theories to cast judgement on every trivial decision we make). We can alleviate this worry by noting that moral theories can satisfy Action Guidance by relying on a presumption of permissibility, such as: if a choice context has no morally relevant properties, then all of its actions are morally permissible.

A moral theory must do more than just satisfy Action Guidance. After all, moral decision rules like Certainty and Positive Probability satisfy Action Guidance: the former consistently regards almost all actions to be permissible, while the latter is consistent in prohibiting practically everything. These decision rules are inadequate because they conflict with other conditions. For example:

Impermissible Risk: There are contexts in which actions are impermissible even though they are not certain to violate a moral norm.

Permissible Risk: There are contexts in which actions are permissible even though they are not certain to conform to a moral norm.

There are various other adequacy conditions that we might put forward. To generate the impossibility result in Part 1.4.4, we will add just one other condition:

Dominance: An action x dominates another action y, if in a state-by-state comparison y does not defeat x in any state, but in some state, x defeats y. Given a choice between either a dominating action or a dominated one, a moral decision rule holds that you should perform the dominating action.

Dominance is a well-accepted condition for rational choice.[41] In cases where our actions affect others, it is also a compelling condition for moral choice. To

---

41. Here, as before, we are assuming Act-State Independence.

illustrate using a slightly altered version of **Patient**, if under every contingency about the patient's receptivity, Drug A will lead to at least as good and potentially better outcomes for the patient than Drug B, then you ought to give Drug A and you ought not to give Drug B. Indeed, giving Drug B would needlessly deprive the patient of a potential cure – and in fact would do so in exchange for a heightened possibility of harm. In this way, Dominance is akin to a necessity constraint for risky situations, exhorting us to avoid choosing an action that imposes unnecessary risk by being guaranteed to be no better than – and potentially worse than – some alternative.

1.4.4   An Impossibility Result

In their discussion of Option Absolutism, Jackson and Smith were correct in holding that if such a moral theory rejects Certainty and Positive Probability, then it must adopt some kind of threshold approach. They illustrated this with the rule, Threshold. However, there is a wider class of threshold rules that could be adopted, depending on other statistical features of an action's prospect. For instance, a normative relevance function could instead take the *number* of possible moral norm violations to be relevant, or the *probability-weighted number*, or the *risk-weighted number*, or *mean-variance* value of possible outcomes, and so on.[42] These statistical features will not necessarily be functions of some fixed probability value; the same overall probability of violating a moral norm may not capture, for example, the number of norm violations or the range of such violations. Generalising from Threshold, any Option Absolutist theory that accepts Permissible Risk and Impermissible Risk must posit:

Prohibited Prospects The set of actions that have $\bar{P}$ are those that have a particular statistical feature P_s, where P_s applies to at least some actions that are not certain to violate an absolutist moral norm.

Option Absolutism must posit some kind of statistical feature that renders an action permissible or impermissible. A moral decision rule that satisfies Action

---

42. For a discussion of various ways of measuring and evaluating a prospect, see: Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013), chap. 2.

Guidance maps actions to either the subset of permissible or impermissible actions (there being no 'in-between' deontic status). As such, this statistical feature will have a sharp cutoff, amounting to a statistical threshold of some kind. Option Absolutism must posit Prohibited Prospects or else it will be silent about the permissibility of risky actions, and will thereby fail all of the adequacy conditions set out earlier. This preliminary observation helps us to establish the following:

Proposition 1. *No Option Absolutist theory satisfies Impermissible Risk, Action Guidance, and Dominance.*

Proof. Suppose that an Option Absolutist theory satisfies Impermissible Risk. This implies that there are contexts k in which an action x possesses a prohibited prospect P_s (and, therefore $\bar{P}$). Now suppose that x is also in k', such that all other available actions y in k' are dominated by x. Since $\bar{P}$ is an option property, if x possesses that property in context k, then x also possesses it in context k'. That is, Option Absolutism holds that the presence of any action y does not change the impermissibility of x. By Action Guidance, this moral theory determines a moral decision rule that entails that either: x is permissible in context k' (adhering to Dominance, but violating Option Absolutism) or x is not permissible in context k' (violating Dominance, but adhering to Option Absolutism).

$\square$

To illustrate, suppose that in **Patient**, your Hippocratic moral theory holds that some treatments are 'off the table', being too dangerous to ever be permissibly used regardless of the alternatives. For instance, suppose that you should never give Drug C to your patient because it is too dangerous, even though there is some chance that it will harmlessly cure your patient. We can imagine cases where you could face a tragic choice between Giving Drug C or Giving Drug D, where Drug D *guarantees* death.[43] In such cases, Option

---

43. One might argue that for all such choice contexts, Do Nothing is always available and permissible, even if Do Nothing is also guaranteed to cause (by omission) your patient's death. The worry with this approach is that in cases of uncertainty, where 'doing' actions have some risk of being wrongful, we will always be required to 'do nothing'. See: Seth Lazar, "Deontological Decision Theory and the Grounds of Subjective Permissibility" (2018).

Absolutism must maintain that Drug C is prohibited, even though it offers your patient the only means of survival.

How can Option Absolutists avoid this result? They may reject Dominance and argue that you face a prohibition dilemma, such that whatever you do, you are doing something you ought not to do. However, this seems highly implausible. Firstly, this is not the rather plausible kind of moral dilemma whereby there is a conflict of different, incommensurable considerations.[44] Here, the very same consideration is at play: namely, the well-being of your patient. This is also not the kind of moral dilemma where your two available actions (Drug C and Drug D) are equally bad or 'wrongful'. Indeed, Giving Drug D amounts to needlessly denying your patient the possibility of survival, whereas Giving Drug C grants your patient some prospect of survival.[45] Option Absolutism blindly rules out the one action that could save your patient.

Overall, it is difficult to see what could be said in defence of Option Absolutism in such cases. However, we may not have to say anything at all: in Part 1.5 we will see that the standard targets of the Problem of Risk can plausibly avail themselves of the following:

1.4.5 A Possibility Result

Unlike Option Absolutism, Relational Absolutism avoids the Problem of Risk.

Proposition 2. *There are moral theories that satisfy Relational Absolutism, Permissible Risk, Impermissible Risk, Action Guidance, and Dominance.*

Proof. Given Permissible Risk and Impermissible Risk, there are contexts in which an action x is impermissible in k despite not being guaranteed to violate a moral norm. Suppose that the relevant relationally prohibited property is: $\bar{P}$

---

44. See Lisa Tessman, *Moral Failure: On the Impossible Demands of Morality* (Oxford: Oxford University Press, 2015).

45. Note that in this case, the commonly-held intuition that small differences in risk do not make a difference to the relative permissibility of actions is less compelling than normal. Even if the difference between Drug C and Drug D is only a small risk, it is nevertheless the case that there is nothing to say in favour of Drug D, but something to say in favour of Drug C, such that it would be wrong to not give Drug C instead of Drug D. *Pace* Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model," pp. 9-11.

= "Having the lowest expected deontic value".[46] Given that $\bar{P}$ is a relational property, if x is in a context k' where there is only one other available action, y, which has a lower expected deontic value, then x does not have the relational property $\bar{P}$. Given Action Guidance, the moral decision rule determines the permissibility of actions x and y in k'.

Suppose for *reductio*, however, that this moral theory violates Dominance. This leads to an immediate contradiction: it is impossible that y dominates x, but y has lesser expected deontic value than x.[47]   □

The decision rule given above is just one of the many available decision rules that satisfy the adequacy conditions we have set out. The availability of particular decision rules will depend on the particularities of the reasons structure, such as whether its defeat relation can be represented by a cardinally measurable value function (allowing for the moral decision rule: maximise expected deontic value, or a risk-weighted value function).[48] If the weighing relation can only be represented by an ordinal value function, then moral theories can, for example, attempt to adopt an Expected Borda Count.[49] If there is no probablistic information available at all, a moral theory could adopt a rule such as Maximin: choose the action that has the least worst possible outcome.[50] In

---

46. An action's 'expected deontic value' is the probability-weighted average of the deontic values of each of its possible outcomes, where the deontic value of an outcome represents (roughly) the importance of the moral norms at play in that outcome. This property is well-defined providing that a moral theory satisfies a set of axioms of expected value theory, suitably interpreted for moral theories rather than rational choice. For instance, see: Colyvan, Cox, and Steele, "Modelling the Moral Dimension of Decisions."

47. Given our background assumption of Act-State Independence.

48. Cardinal value functions represent whether one bundle of properties is more, less, or equally weighty than some other bundle, and also gives a meaningful measure of the difference in the weights of bundles. Formally, cardinal value functions are unique up to positive linear transformation (where the value function is multiplied by a positive number, plus a constant). On risk-weighted approaches to decision under risk, see: Buchak, *Risk and Rationality*.

49. An ordinal value function simply determines whether two bundles of properties are equally weighty or whether one is more weighty than another. Formally, such functions are unique up to positive transformation (that is, they maintain the same ordering when their values are multiplied by a positive constant). On the use of the Borda Count in a different theoretical context, see: William MacAskill, "Normative uncertainty as a voting problem," *Mind* 125, no. 500 (2016): 967–1004.

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

any case, once a moral theory accepts Relational Absolutism, it can potentially avoid the Problem of Risk.

1.5 Which Moral Theories Face the Problem of Risk?

The above results show that Option Absolutist theories cannot avoid the Problem of Risk, but Relational Absolutists can. How do these results bear on presently-held moral theories? I argue, firstly, that the standard targets of the Problem of Risk can actually be read as Relational Absolutists. I then show how Relational Absolutism can accommodate the kinds of considerations that would motivate one to accept Option Absolutism, namely: a concern for the inherent wrongfulness of particular kinds of actions, and a rejection of moral aggregation.

1.5.1 Will the Real Option Absolutists Please Stand Up?

It has often been suggested that simply subscribing to an *uncompromising* moral theory makes you subject to the Problem of Risk. As we saw earlier in this paper, moral theorists such as Kant, Anscombe, and Nozick have been tarred with the Absolutist brush, and subjected to the Problem of Risk.

However, despite their absolutist reputations, on closer inspection it actually appears that none of them are committed to Option Absolutism. I shall briefly argue that the above prototypical victims of the Problem of Risk can faithfully be interpreted as Relational Absolutists, not Option Absolutists.

There are tell-tale signs that a moral theory subscribes to Option Absolutism. For instance, a moral theory is Option Absolutist only if it accepts the possibility of prohibition dilemmas: cases where all of your actions are prohibited.[51] Also, if a moral theory is Option Absolutist, then it will ignore

---

51. Although the reverse is not true: accepting the existence of prohibition dilemmas does not commit you to Option Absolutism.

comparative information about actions. By contraposition, if a moral theory rejects the possibility of prohibition dilemmas or pays attention to comparative information, it is not an Option Absolutist theory.

As it turns out, Kant seems to hold that prohibition dilemmas are impossible.[52] Moreover, he appears to accept the relevance of comparative information about our actions, holding our obligating reasons (as Onora O'Neill calls them) can conflict and that, in such cases, we are required to choose the stronger reasons over the weaker ones:

> [A] subject may have, in a rule which he prescribes to himself, two obligating reasons (*rationes obligandi*), one or other of which is not sufficient to put him under obligation (*rationes obligandi non obligantes*), so that one of them is not a duty... When two such reasons conflict with each other, practical philosophy says, not that the stronger obligation takes precedence (*fortior obligatio vincit*) but that the stronger obligating reason prevails (*fortior obligandi ratio vincit*).[53]

Kant's position regarding obligating reasons can be modelled as a Relational Absolutist theory, whereby the prohibited actions are those that are not supported by the 'stronger' (or 'undefeated') obligating reasons:

- for every context k, the set of relevant properties is the set of obligating reasons P_R.

- the defeat relation ranks P_R over $P_{R'}$ if and only if P_R is a stronger obligating reason than $P_{R'}$.

- the class of absolutely prohibited actions $\bar{P}$ are those that have weaker obligating reasons than that of some other available alternative in that context.

The key suggestion here is that even though Kant is apparently committed to the idea that some kinds of actions are always morally impermissible, this

---

52. For a defence of this reading of his view, see: Onora O'Neill, "Instituting Principles: Between Duty and Action," *Southern Journal of Philosophy* 36, no. S1 (1998): 79–96.

53. Immanuel Kant, *The Metaphysics of Morals*, ed. Mary J Gregor (trans.) (Cambridge University Press, 1996), 6:224; O'Neill, "Instituting Principles: Between Duty and Action," p. 88.

alone does not commit him to Option Absolutism. It all depends on whether the relevant kinds are individuated with reference to one's other available actions. As it is, his discussions of conflicting obligating reasons and denial of prohibition dilemmas suggests that he is better understood as a Relational Absolutist. While this is by no means a definitive analysis, it should be enough to establish that purveyors of the Problem of Risk must do more than issue a summary judgement of whether a particular moral theory can deal with risky situations; they must show why such moral theories are specifically committed to Option Absolutism.

Let us now turn to another putative moral absolutist, G.E.M. Anscombe. At first glance, it is puzzling to label her a moral absolutist, since a headline message of her ethics that "the concepts of ... moral obligation and moral duty ... ought to be jettisoned."[54] It is hard to be a Moral Absolutist if you reject the importance of deontic verdicts.

However, in her discussions of the ethics of war, she does make such verdicts. Indeed, when she does so, she seems to commit herself to some version of Moral Absolutism:

> We may not commit any sin, however small, for the sake of any consideration, however great, and if the choice lies between our total destruction and the commission of sin, then we must choose to be destroyed.[55]

Although such statements certainly fit the caricature of Moral Absolutism, it is important to note that Anscombe's position regarding the relative importance of avoiding sinful actions versus ensuring our survival does not commit her to Option Absolutism. To determine whether Anscombe is an Option Absolutist, we must determine whether she takes the relative merits of our other available actions to determine whether an action is impermissible (or 'sinful').

In fact, even Anscombe appears to adopt a Relational Absolutist approach to permissibility. For example, in her discussion of President Truman's decision to bomb Nagasaki, she argues that the decision was 'murderous' because

---

54. G. E. M. Anscombe, "Modern Moral Philosophy," in *Virtue ethics*, ed. Roger Crisp and Michael A Slote (Oxford: Oxford University Press, 1997), 26–44.

55. G. E. M. Anscombe, *Ethics, Religion and Politics* (Oxford: Basil Blackwell, 1981), p. 79.

it was marred by "the fixation on unconditional surrender [and] the disregard of the fact that the Japanese were desirous of negotiating peace."[56] Bombing Nagasaki was murder, according to Anscombe, because there were better options available, not because killing the innocent is a categorically prohibited action. Indeed, she makes this last point explicit:

> I intend my formulation to be taken strictly; each term in it is necessary. For killing the innocent, even when you know as a matter of statistical certainty that the things you do involve it, is not necessarily murder. I mean that if you attack a lot of military targets, such as munitions factories and naval dockyards, as carefully as you can, you will be certain to kill a number of innocent people; but that is not murder. On the other hand, *unscrupulousness in considering the possibilities turns it into murder.*[57]

Thus, in so far as Anscombe can be read as making absolutist 'moral ought' claims at all, she seems to accept that whether an action possesses a particular absolutely prohibited property (say, that of being 'murderous') at least partly depends on how that action compares to its alternatives. This suggests that Anscombe's ethical theory cannot be easily categorised as a version of Option Absolutism. Rather, a charitable reading of her view will be that of Relational Absolutism, allowing it to avoid the Problem of Risk.

Finally, let us consider whether Robert Nozick's Libertarianism, and rights-based theories in general, necessarily face the Problem of Risk. In fact, it seems that even Nozick believes that rights-based theories have problems dealing with risk:[58]

> Actions that risk crossing another's boundary pose serious problems for a natural-rights position ... Imposing how slight a probability of a harm that violates someone's rights also violates his rights? ... [Solutions relying on expected value calculations] cannot be utilized by a tradition which holds that stealing a penny or a pin or anything from someone violates his rights. That tradition does not select a threshold measure of harm as a lower limit, in the case of harms certain to occur. It is difficult

---

56. ibid., p. 65.
57. ibid., p. 66 Emphasis mine.
58. Nozick, *Anarchy, State, and Utopia*. See also: McKerlie, "Rights and Risk"; Altham, "Ethics of Risk."

to imagine a principled way in which the natural-rights tradition can draw the line to fix which probabilities impose unacceptably great risks upon others.[59]

It is arguable that Nozick is overstating the incompatibility of rights-based theories with expected utility calculations. Notice that in the often-quoted passage above, Nozick is concerned with finding a principled means of determining which kinds of risks count as rights-violating. Under Dietrich and List's framework, this amounts to determining which kinds of risks are included in the set of morally relevant properties. Nozick seems to be worried that if we have rights against risk being imposed on us, then all of our actions will be impermissible because all of our actions tend to impose some risk on others.[60] However, this worry is misplaced.

Simply put, a rights-based theory can accept that all risk impositions matter, while holding that they do not all matter equally.[61] The key question will be whether and, if so, under what conditions, our rights against risk may be outweighed by other moral considerations. Short-cutting some lengthy Nozickian exegesis, his theory corresponds to roughly the following reasons-structure:

- for every context k, the set of morally relevant properties consists of rights-based considerations P_R and goodness-based considerations P_G.

- the defeat relation ranks P_R over P_G for all instances of P_R and P_G.[62]

- within the class of P_R actions, those that involve *you* violating a right P_{R_i} are ranked lower than those that involve *others* violating a right $P_{R_{\neg i}}$.

This reasons structure is consistent with Relational Absolutism. For instance, Nozick's theory can adopt the following moral decision rule:

---

59. Nozick, *Anarchy, State, and Utopia*, pp. 74-5

60. Many have echoed this worry. See: Hayenhjelm and Wolff, "The Moral Problem of Risk Impositions: A Survey of the Literature"; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*.

61. On this approach, see: McCarthy, "Rights, Explanation, and Risks."

62. This may not be quite right: in a well-known footnote, Nozick leaves open that there may be extreme cases where a right may be permissibly violated for the sake of the good. Nozick, *Anarchy, State, and Utopia*, pp. 29-30.

- the class of absolutely prohibited actions $\bar{P}$ are those in which you do not maximise your expected degree of rights conformity.[63]

Overall, it is both uncharitable and inaccurate to assume that uncompromising theories (like those of Nozick, Anscombe and Kant) are Option Absolutist theories. Rather, they are instances of a class of moral theories called *lexical priority theories*: those that hold that some kinds of moral considerations cannot be defeated by any number of particular other considerations. In general, lexical priority theories can accept Relational Absolutism by, for example, holding that we should choose the action that best upholds the higher consideration, using only the lower considerations to break ties between actions.[64] Hence, these theorists can avoid the Problem of Risk.

1.5.2 Objections to Relational Absolutism

Notwithstanding the fact that Option Absolutism can lead to violations of Dominance, some moral theorists may nevertheless baulk at accepting Relational Absolutism. And understandably so: it seems, at least at first glance, that Relational Absolutism will misrepresent moral theories that take some actions to always be wrongful or those that prohibit the weighing and aggregation of people's lives. However, I will suggest that such theories can accept Relational Absolutism without giving up these important substantive commitments.[65]

One motivating concern is based on the idea that there is a close connection between an action's permissibility and whether it is rightful or wrongful. From this perspective, accepting Relational Absolutism seems to ignore the idea that some actions are always morally *wrongful*, and therefore cannot be morally *permissible*. For instance, it seems bizarre to say that it is *morally*

---

63. One might object that any such rule will be 'goal-directed', and hence ruled out by Nozick: ibid., pp. 30-33. Such a charge would need to show why the moral decision rule given here necessarily commits us to treating others as mere means to our ends, which is Nozick's chief concern. However, so long as we are careful about the reasons structure, however, it doesn't seem like there will be any such commitment.

64. See: Chapter 2 and Chapter 3.

65. Thanks to Seth Lazar for pressing these points.

permissible to play five-bullet Russian Roulette on an unconsenting individual just because the only available alternative is to play six-bullet Russian Roulette. Both actions are, in a very immediate sense, morally *impermissible*. If that means giving up Dominance, so be it! Many moral theorists – deontologists, in particular – seem to share this kind of sentiment. However, I will briefly sketch how this position can be captured by Relational Absolutism, allowing one to maintain Dominance while still recognising the wrongfulness of particular kinds of actions.

Consider the following view: some kinds of actions, such as playing Russian Roulette on someone – are always wrongful. Performing such an action provides grounds for shame, compensation, or perhaps even punishment. However, if you find yourself forced to play a round of Russian Roulette on someone with five bullets or six bullets, although both options are wrongful, you ought to choose to play with five bullets, since it is the *least* wrongful action available (it at least gives your victim *some* chance of survival).

I submit that this would be a Relational Absolutist view, whereby the relevant properties of an action include wrongfulness properties, and the absolutely prohibited actions are those that do not minimise wrongfulness. Of course, formally speaking, your moral decision rule will deem your five-bullet option 'permissible', but this is a very thin sense of permissibility (after all, your action is still wrongful). Moreover, this permissibility verdict is extremely *modally fragile* – under almost any other choice context, your action would be clearly impermissible. Note also that accepting Relational Absolutism does not rule out the possibility of strong moral dilemmas: it only rules out such dilemmas when one action dominates another. In other cases where Dominance does not apply, there may be a prohibition dilemma at hand. For instance, if a moral theory holds that the impermissible actions are those that are defeated by some other action, then they may encounter prohibition dilemmas in cases where all actions are defeated by some other action. Overall, accepting that some actions can be inherently wrongful and that prohibition dilemmas exist does not confine a moral theorist to Option Absolutism. Such theorists can accept an appropriately interpreted Relational Absolutist approach, and thereby avoid the Problem of Risk.

A second reason for resisting Relational Absolutism is based on a rejection of particular kinds of moral aggregation. On this kind of view, individual claims should not be weighed against each other – the numbers, as Taurek put it, don't count.[66] To weigh the aggregate interests of a collection of people against those of others would be to overlook the separateness of persons.[67] The worry is that accepting Relational Absolutism and Dominance commits us to violating this tenet.

However, this does not follow. Relational Absolutism can deem the numbers of claims to be morally irrelevant, and instead rank actions according to whether or not they give individuals the appropriate chance to have their claim satisfied. In terms of the defeat relation, the highest-ranked properties of actions are those that give the most appropriate probabilities of satisfying individuals' respective claims. This is compatible with Dominance, since it only requires that you do not choose an action that has an unnecessary risk of satisfying no-one's claim. More generally, accepting Relational Absolutism does not entail accepting any particular stance towards moral aggregation. Thus, even ardent anti-aggregationists can accept this version of Relational Absolutism.

Overall, it seems that few moral theories, if any, are necessarily Option Absolutist. It also seems that there are few reasons, if any, to be an Option Absolutist rather than a Relational Absolutist. For critics of a moral theory to successfully 'use structure as a weapon', their best approach may instead be to show that their target moral theories face the Problem of Risk under stronger adequacy conditions for moral decision rules.[68] However, this remains to be seen.

---

66. John M. Taurek, "Should the Numbers Count?," *Philosophy and Public Affairs* 6, no. 4 (1977): 293–316.

67. Thomas Nagel, *Mortal Questions* (Cambridge University Press, 1979); John Rawls, *Justice as Fairness: A Restatement*, ed. Erin Kelly (Cambridge, Massachusetts and London, England: Harvard University Press, 2001); T.M. Scanlon, *What We Owe to Each Other* (Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press, 1998); Alex Voorhoeve, "How Should We Aggregate Competing Claims?," *Ethics* 125, no. 1 (2014): 64–87.

68. To use Nozick's phrase: Robert Nozick, "Moral Complications and Moral Structures," *Natural Law Forum* 13 (1968): p. 44.

1.6 Conclusion

The Problem of Risk has been levelled against a number of moral theories over the years. It has led to valuable improvements in our understanding of moral theories. This paper has demonstrated, however, that much of the discussion of the Problem of Risk has been mistaken. Jackson and Smith (2006) are correct that Option Absolutism faces the Problem of Risk. However, subsequent attempts to extend the Problem of Risk to other moral theories fail, since such theories can be interpreted as accepting Relational Absolutism. In fact, it is arguable that even prototypical Option Absolutists are actually Relational Absolutists. The upshot is that since few, if any, moral theories are committed to Option Absolutism, it turns out that a much wider range of moral theories can potentially guide us through uncertain situations than previously thought possible.

Part II

Solutions

Chapter 2

Moral Priorities Under Risk

2.1 Introduction

A public official is deliberating about whether to approve a marginal increase in
the speed limit for autonomous vehicles. Surveys show that this increase would
improve passenger satisfaction, but would not lead to any other substantial
improvements. Against this move, however, is evidence that increasing the
speed limit poses some risk of increasing the current incidence of pedestrian
deaths: some people crossing the road may underestimate the speed of the
self-driven vehicles as they quietly shuttle through the streets. There are no
other relevant considerations. The decision is hers alone.

The official takes a moment to reflect on the moral considerations at hand.
One such consideration is public safety: she ought to choose the option that
minimises pedestrian deaths. The other consideration is people's pleasure:
increasing passenger satisfaction will increase the amount of pleasure in the
world. Although both considerations are morally significant, to her mind they
are not equally important. In fact, she believes that no amount of passenger
satisfaction could ever morally justify a pedestrian's death. As far as she is
concerned, considerations of public safety are the primary moral consideration
at hand.

The primacy of public safety in this case can be spelled out more precisely

using the notion of *lexical priority*.[1] A moral consideration has lexical priority over another just in case, given the choice between upholding a higher-ranked consideration versus upholding any number of lower-ranked ones, we ought to uphold the higher-ranked consideration. Lower-ranked considerations are only relevant insofar as they help to break ties between options that uphold higher-ranked considerations equally well. Lexical priority theories thus prohibit trade-offs between different kinds of moral considerations. In the official's case, considerations of passenger satisfaction ought not to be traded-off against pedestrians' lives.

To her disappointment, however, the official realises that her awareness of these moral considerations is not sufficient to guide her decision. The problem, of course, is that she does not know whether increasing the speed limit will in fact lead to more deaths. Since she is to some degree uncertain about whether her choices will lead her to violate or uphold particular moral considerations, she faces a *moral decision under risk*. What she requires is a *moral decision rule*: a rule that identifies which options are morally permissible, given her degrees of uncertainty. The question is: is there a moral decision rule that both maintains the lexical priority of public safety over additional passenger satisfaction, while also giving acceptable guidance in moral decisions under risk?

If the recent decision-theoretic critique of lexical priorities is correct, then it turns out that there is no acceptable moral decision rule available to the official.[2] This is because lexical priority theories appear to commit a decision-maker to at least one of the following problems: the Permissiveness Problem, the Low Risk Problem, and the Agglomeration Problem. The Permissiveness Problem consists of cases where all of one's options are equally permissible, simply because they have some positive probability of violating a lexical priority. The Low Risk Problem arises when arbitrarily small probabilities make

1. On the history and applicability of this concept to various domains in moral and political philosophy, see: John Rawls, *A Theory of Justice: Revised Edition* (Cambridge, MA: Belknap Press, 1999). Rawls himself takes lexical priorities to only be a 'useful approximation' for the purposes of his theory.

2. Huemer, "Lexical Priority and the Problem of Risk"; Jackson and Smith, "Absolutist Moral Theories"; Lord and Maguire, "The Implementation Problem for Deontology."

particular actions impermissible. The Agglomeration Problem involves cases where lexical priority theories give contradictory advice depending on whether the relevant moral considerations are viewed separately or together. Lexical priority theories must avoid each of these problems in order to have an acceptable moral decision rule for risky situations.

This chapter uses lexicographic expected value theory to rebut the decision-theoretic critique. As I shall argue, many of the problems that have been raised for lexical priority theories are actually the result of inappropriate decision-theoretic modelling. Once a more appropriate model is applied, the problems do not arise. Nevertheless, although the decision-theoretic critique fails in this respect, it succeeds in revealing important but under-theorised aspects of lexical priority theories.

Part 2.2 briefly sets out the preliminaries for modelling a moral theory using decision theory. **Part 2.3** responds to the Permissiveness Problem. **Part 2.4** responds the Low Risk Problem. **Part 2.5** responds to the Agglomeration Problem. **Part 2.6** discusses the value and limits of decision-theoretic critiques of moral theories, and identifies further questions that lexical priority theories must answer to give a complete account of moral decision-making under risk. **Conclusion** follows.

2.2   Preliminaries

The decision-theoretic critique of lexical priority theories takes place on the following theoretical terrain: firstly, it proceeds on the assumption that *expected value theory* is an appropriate framework for representing moral theories; secondly, the objections to lexical priority views are premised on a *particular way of modelling* lexical priorities using expected value theory. I will explain these theoretical points in turn.

In its normative application, expected value theory determines what you ought to do, given your uncertainty. It identifies the best option relative to a description of a decision problem.[3] A decision problem consists of: the *op-*

---

3. For an accessible introduction to the normative applications of expected utility theory, see: R.A. Briggs, "Normative Theories of Rational Choice: Expected Utility Theory," *The*

tions available to you; the possible *states* of the world; the *probabilities* of those states (represented by some value in the [0, 1] interval), given that you perform some option;[4] and the value of the *outcomes* that result from choosing a particular option in a given state of the world. According to expected value theory, the best option is the one that maximises *expected value*, where this is the option whose possible outcomes together have the greatest sum of probability-weighted value.

To apply expected value theory to moral theory, we must assume that the relative importance of moral considerations can be numerically represented by a value function.[5] This 'moral' value function must assign equal value to equally important considerations, greater value to more important considerations, and lesser value to less important considerations.[6] However, to operate in a context of risk, a moral value function must also accurately represent the *differences* in the relative importance of moral considerations. That is, we must assume that the importance of moral considerations can be represented by a cardinal moral value function.

As we shall see, the precise specification of the cardinal value function is important to how a moral theory operates in an expected value framework. A common modelling assumption in discussions of lexical priority theories is that higher-ranked considerations are *infinitely* more important than lower-ranked considerations, and so should be represented by an infinite value difference.[7]

4. Here I wish to remain neutral about whether evidential decision theory or causal decision theory is correct. Readers who prefer evidential decision theory can read the probabilistic dependence of states as conditional probabilities (the probability of a particular state, given that a particular option is chosen), whereas those who prefer causal decision theory can read the dependence in terms of the probabilities of subjunctive conditionals, imaging functions, dependency hypotheses, etc. See: James M. Joyce, *The Foundations of Causal Decision Theory* (Cambridge; New York: Cambridge University Press, 1999), chap. 5

5. More precisely, in an expected value framework, the values are assigned to the *outcomes* in which the moral considerations are upheld or violated. For ease of discussion, however, I will simply say that the considerations have value.

6. Slightly more formally, for all considerations C_1 and C_2, a moral importance relation w (where $\succ_w$ stands for 'more important than' and $\sim_w$ stands for 'equally as important as'), and a moral value function v: $C_1 \succ_w C_2$ if and only if $v(C_1) > v(C_2)$; and $C_1 \sim_w C_2$ if and only if $v(C_1) \sim v(C_2)$.

7. See, for example: Jackson and Smith, "Absolutist Moral Theories"; Lord and Maguire, "The Implementation Problem for Deontology"; Colyvan, Cox, and Steele, "Modelling the

As we shall see, this is a problematic and unnecessary assumption. Strictly speaking, lexical priority theories only subscribe to an ordinal ranking of moral considerations: given a choice between upholding either a higher-ranked consideration or any number of lower-ranked considerations, it is *more* important to uphold the higher-ranked consideration. Lexical priority theories need not make any commitment about *how much more* important it is to uphold a higher-ranked consideration than lower-ranked ones. Infinite values simply offer one way of representing the fact that lexical priority theories stratify some kinds of moral considerations over others.

To test whether lexical priority theories can give adequate guidance in cases of risk, we will model the official's case as follows. We will assume that she has only two available options: increase the speed limit (by some fixed amount) or maintain the current speed limit. There will be only two possible and mutually exclusive states of the world: either the pedestrians will be careful when crossing the road (*Careful*) or they will not (*Careless*). We will allow that the probabilities of these states may be affected by the option chosen. Following the critics of lexical priority theories, we will hold that outcomes in which a lexical priority is violated – for instance, cases where there is an increase in pedestrian deaths – have infinite disvalue ($-\infty$), whereas outcomes in which a lexical priority is upheld are assigned some finite value (f). The above framework sets out a particular way of determining what lexical priority theories require of us in cases of uncertainty. One might object to this framework on a number of fronts.[8] For dialectical purposes, however, we will assume that it is appropriate. We will see that by making amendments within the expected value framework, lexical priority theories can satisfactorily deal with the decision-theoretic critique.

Moral Dimension of Decisions"; Huemer, "Lexical Priority and the Problem of Risk"; Hayenhjelm and Wolff, "The Moral Problem of Risk Impositions: A Survey of the Literature"; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*; Adam Bjorndahl, Alex John London, and Kevin J. S. Zollman, "Kantian Decision Making Under Uncertainty: Dignity, Price, and Consistency," *Philosopher's Imprint* 17, no. 7 (2017): 1–22. Huemer's model is an informal rendition of the infinite values model introduced in Jackson and Smith, "Absolutist Moral Theories."

8. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, chap. 8

2.3   The Permissiveness Problem

Lexical priority theories hold that we are morally prohibited from trading-off higher considerations for the sake of lesser ones. Framing lexical priority theories in this way, an important question is: what are we to do when all of our options have some positive probability of being of this prohibited kind? Some worry that lexical priority theories may commit us to a life of implausibly strong moral dilemmas, whereby all of our options are prohibited, merely by virtue of having a positive probability of violating a lexical priority. They call this the Paralysis Problem.[9]

However, when the problem is cast in an expected value framework, the Paralysis Problem is in fact better understood as the Permissiveness Problem. Expected value theory, after all, exhorts us to choose the option that maximises expected value, even if that option has low (or, indeed, negatively infinite!) expected value. As we shall see, when lexical priority theories are modelled using infinite values, the problem is that all of our options turn out to be *equally permissible* simply by virtue of having some positive probability of violating a lexical priority.

To illustrate, suppose that the official must decide on the basis of the following evidence: one report suggests that increasing the speed limit may greatly increase pedestrian deaths, whereas another report suggests that maintaining the current speed limit may also increase pedestrian deaths, albeit to a lesser extent. Given that both options are risky, what ought she to do? As shown in Table 2.1, since both options have some positive probability of violating a lexical priority, both have negatively infinite expected value. According to the expected value model, both options are therefore permissible because they have the same expected value.

Clearly, however, this can lead to absurd results. For example, even if one option is far riskier than the other, it will have the same expected value and

---

9. For versions of this problem, see: Huemer, "Lexical Priority and the Problem of Risk," p. 337; Bjorndahl, London, and Zollman, "Kantian Decision Making Under Uncertainty: Dignity, Price, and Consistency," p. 8; Hansson, *The Ethics of Risk: Ethical Analysis in an Uncertain World*, chap. 2; Hayenhjelm and Wolff, "The Moral Problem of Risk Impositions: A Survey of the Literature," E26-51.

Action	Careless p	Careful $(1-p)$	Expected Value
Increase speed limit	$-\infty$	f_1	$-\infty$
Maintain speed limit	$-\infty$	f_2	$-\infty$

Table 2.1: The Permissiveness Problem.

therefore will be deemed equally choice-worthy. The Permissiveness Problem seems to show that lexical priority theories, blinded by their own fanaticism, lack the ability to distinguish impermissible risky options from permissible ones.

As it turns out, however, the Permissiveness Problem does not pose a serious challenge to lexical priority theories. Rather, the problem is due to a flawed decision-theoretic model of lexical priority theories. Specifically, the Permissiveness Problem arises because infinite values swamp expected value calculations, rendering all options equally permissible regardless of their riskiness. A more appropriate model will allow lexical priority theories to use expected value calculations to distinguish between permissible and impermissible options. This gives us good reason to look for non-infinitistic, expected value representations of lexical priority theories under risk.

The rich literature on Pascal's wager offers many candidate approaches.[10] One example is lexicographic expected value theory. Unlike most other approaches, the lexicographic approach constitutes a minimal departure from orthodox expected value theory.[11] Indeed, it is simply a generalisation of standard expected value theory. Where standard expected value theory operates with a one-dimensional value function, the lexicographic theory operates with a multi-dimensional function. It also has the benefit of being well known and comprehensively theorised in contemporary economic and decision theory.[12] Famously, Rawls defended a version of lexicographic decision theory in

---

10. For a review of such approaches, see: Alan Hájek, "Waging War on Pascal's Wager," *The Philosophical Review* 112, no. 1 (2003): 27–56

11. The Permissiveness Problem can also be avoided without abandoning infinite values, using Relative Utility Theory – see: Paul Bartha, "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities," *Synthese* 154, no. 1 (January 2007): 5–52. For simplicity of exposition, I have chosen an alternative, lexicographic approach.

12. See, for example: Nicolas Houy and Koichi Tadenuma, "Lexicographic compositions

his Theory of Justice, although he did not apply it to decision-making under risk (due to his denial that probabilities are available behind the veil of ignorance).[13] As we shall see, the lexicographic model allows lexical priority theories to systematically avoid the Permissiveness Problem.

Using the lexicographic model, rather than using infinite values, we will represent lexical priorities using a ranking of finite-valued value functions, $<v_1, v_2, \ldots, v_n>$. In the official's case, we have assumed for simplicity that there are only two kinds of moral considerations at hand: public safety and additional passenger satisfaction. As such, we will require only two value functions, $<v_1, v_2>$, where v_1 represents the moral significance of the level of public safety in that outcome and v_2 represents the moral significance of the level of passenger satisfaction in that outcome. We represent the priority of public safety over additional passenger satisfaction using the following moral decision rule: the official ought to choose the act that has the highest expected value for v_1 and – just in case there is a tie among the options – then she ought to choose the act that has the highest expected value for v_2. In the unlikely event that there is yet another tie, she is permitted to choose either option (since there are no other considerations in this example).[14]

Without loss of generality, we will assume that the official has a confidence of 0.95 that the public will be careful and we will assign more-or-less arbitrary lexicographic values to the various outcomes. This will give us a clear demonstration that our chosen moral decision rule can guide decision-making in cases where all options are risky.

of multiple criteria for decision making," *Journal of Economic Theory* 144, no. 4 (2009): 1770–1782.

13. Rawls's theory adopts a lexicographic framework. However, his approach rejects expected utility maximisation altogether, in favour of a maximin decision rule for decision-making under ignorance (John Rawls, "Some Reasons for the Maximin Criterion," *The American Economic Review* 64, no. 2 (1974): 141–146. An early exposition of the idea of lexicographic utility functions is found in Von Neumann and Morgenstern, *Theory of Games and Economic Behavior*. However, it wasn't until later that lexicographic utility functions received their first systematic study in Melvin Hausner, "Multidimensional Utilities," chap. 12 in *Decision Processes*, ed. R. M. Thrall, C. H. Coombs, and R. L. Davis (New York: John Wiley / Sons, Inc., 1954).

14. Slightly more precisely, for two options $A = (a_1, a_2)$ and $B = (b_1, b_2)$: $A \sim B$ if and only if $[(a_1 = b_1)\&(a_2 = b_2)]$; $A \succ B$ if and only if $(a_1 > b_1)$ or $[(a_1 = b_1)\&(a_2 > b_2)]$. This generalises to n-components.

Action	Careless	Careful	Expected Value
Increase speed limit	$-1000, 1$ $p = 0.05$	$0, 1$ $(1 - p) = 0.95$	$-50, 1$
Maintain speed limit	$-2, -1$ $q = 0.05$	$0, -1$ $(1 - q) = 0.95$	$-0.1, -1$

Table 2.2: Lexicographic Expected Value.

As shown in Table 2.2, the option that uniquely maximises expected moral value in this case is to maintain the speed limit. However, that is not the important point. The important point is that there is an option that uniquely maximises expected moral value, even though all options have some probability of violating a lexical priority. As the official is not permitted to pursue just any of her available options, the Permissiveness Problem has been avoided. This shows that infinite values were indeed at the root of the problem. By eschewing infinite values, the expectations of the official's options are now sensitive to, among other things, the probabilities of the states. The lexicographic model is therefore a more appropriate decision-theoretic representation of lexical priority theories.

2.4   The Low Risk Problem

Although lexicographic decision theory solves the Permissiveness Problem, it seems to expose lexical priority theories to the Low Risk Problem. This involves cases where an intuitively acceptable option is ruled out simply because it has a positive – albeit negligible – probability of violating a lexical priority. To many, such cases give us grounds for rejecting lexical priority theories altogether.[15]

To illustrate the Low Risk Problem, let us suppose that the official has reason to believe that increasing the speed limit will almost certainly cause the public to become more careful. By contrast, maintaining the current speed limit will encourage complacency, leading to an almost equal probability of

---

15. For example, see: Huemer, "Lexical Priority and the Problem of Risk."

carelessness or carefulness. According to the lexicographic model, what ought she to do?

Action	Careless	Careful	Expected Value
Increase speed limit	$-1000, 1$ $p = 0.001$	$0, 1$ $(1-p) = 0.999$	$-1, 1$
Maintain speed limit	$-2, -1$ $q = 0.45$	$0, -1$ $(1-q) = 0.55$	$-0.9, -1$

Table 2.3: The Low Risk Problem.

As shown in Table 2.3, the official maximises expected moral value by maintaining the speed limit, despite the fact that it is very improbable that the public would be careless, were she to increase the speed limit. Critics of lexical priority theories hold that in these types of cases, there is a point at which the probabilities of violating a lexical priority are so low as to be negligible. It cannot be that *any* prospect of upholding a lexically-prior consideration, no matter how slim, is more important than a certainty of upholding *any* number of lesser considerations.

To appease proponents of the Low Risk Problem, lexical priority theories must give a principled explanation for why decision-makers should ignore very unlikely prospects of upholding a lexically-prior consideration. One approach is to posit a probabilistic threshold, t, that governs which possibilities decision-makers should ignore and which they should attend to. Note that this option is available because expected value theory merely identifies the best option relative to a specification of the decision problem. It is silent with respect to what information is or is not included in the decision problem. It is therefore open to lexical priority theories to supplement the lexicographic decision model with rules for determining which possibilities are relevant to a decision and which should be ignored.

Critics of lexical priority theories doubt that there is any principled way of fixing a value for t.[16] However, they seem to assume that the threshold value must be context invariant: that is, the same under all circumstances.

---

16. Jackson and Smith, "Absolutist Moral Theories," p. 276; Isaacs, "Duty and Knowledge," p. 97.

Admittedly, it is difficult to see what could justify a universal choice of t. However, since there is no need to assume context invariance, I will offer a context-variant approach to assigning a value to t. On this approach, the chosen threshold should be that which, when followed as a rule, maximises expected moral value.[17]

For example, a policy of always carefully attending to all manner of highly dubious conspiracy theories is unlikely to prove the most effective way to make accurate and timely decisions about public safety. Rather, public policy decisions should only be based on 'live' possibilities, where these are possibilities that are sufficiently probable. The correct threshold of probability will be that which, relative to the stakes and other factors, has the greatest expected lexicographic value when employed in decision-making. This, in turn, will be a matter of balancing contingent factors about the stakes at play in such decision contexts, the quality of the information available, the abilities of the decision-maker to accurately weigh evidence and moral considerations, and the time and resource pressures surrounding the decision. Once this threshold is set, the official will adopt a probabilistic threshold that, when followed as a rule, leads her to maximise the expected moral value of public safety, with the expected moral value of passenger satisfaction acting as a tie-breaker between equally choice-worthy threshold values.[18]

Note that lexical priority theories can adopt this approach because they

---

17. A similar justification is given in non-moral cases by Hannes Leitgeb, "The Stability Theory of Belief," *Philosophical Review* 123, no. 2 (2014): pp. 150-151. Note that the justification put forward here for ignoring low probabilities differs from that proposed in Smith, "The Subjective Moral Duty to Inform Oneself before Acting," in two main respects. Firstly where Smith posits that practical norms are 'tolerant' of slight deviations from infinite precision, the approach put forward here is silent on this point, and instead offers a pragmatic justification based on our bounded cognitive capacities and the importance of efficient versus accurate decision-making. See also: Jacob Ross and Mark Schroeder, "Belief, Credence, and Pragmatic Encroachment," *Philosophy and Phenomenological* ... 88, no. 2 (2012): 259–288. Secondly, Smith's account merely permits, but does not require, ignoring probabilities below the threshold. The justification offered in this chapter is that we are morally required to ignore sufficiently improbable possibilities, since doing so best ensures that we conform to our moral requirements.

18. One might worry that this approach to fixing the threshold leads to an infinite regress of decisions about how to decide. There are resources available to avoid this problem, notably: Hanti Lin, "On the Regress Problem of Deciding How to Decide," *Synthese* 191, no. 4 (2014): 661–670.

are not, strictly speaking, Option Absolutist moral theories. Such theories categorically hold that some kinds of actions are always morally prohibited, irrespective of what other options are available to the decision-maker.[19] Since prohibition does not seem to come in degrees or to be amenable to aggregation, Option Absolutists cannot optimise in the face of risk. By contrast, lexical priority theories allow moral considerations of the same ranking to be traded-off against each other, making room for decision-makers to adopt a probabilistic threshold to optimise their conformity to lexically prior considerations. Under all but the most rarefied circumstances (those involving theoretically ideal agents in ideal situations), a context-variant, probabilistic threshold will best allow decision-makers to navigate moral decision-making under risk.

In what follows, we will equip lexical priority theories with a context-variant, odds-based threshold for determining which states to include in deliberations and which to exclude.[20] Where standard probability threshold approaches assess whether or not a given state is sufficiently probable, the odds-based threshold evaluates whether a state is sufficiently more probable than its most probable alternative. This latter approach ensures not only that a consistent set of states is included in the decision problem, but also that the probabilities of these states are updated in accordance with the Bayesian rules of belief revision. Using this threshold, the official will include in her deliberations the states that are not sufficiently less probable than any other states. She will exclude from her deliberations any state that is sufficiently less probable than some other state. She will then update her beliefs about the included states, normalising the probabilities according to the rules of belief revision, and then maximise expected moral value.

To illustrate, let the relevant odds-based threshold, t, equal $1 : 99$. In the case of the Low Risk Problem, the probability that the public will be careless if she raises the speed limit is less than 0.01. As such, it is insufficiently probable. Given this information, the official should exclude the possibility

---

19. Jackson and Smith, "Absolutist Moral Theories," p. 268. See Chapter 1.

20. Hanti Lin and Kevin T. Kelly, "Propositional Reasoning that Tracks Probabilistic Reasoning," *Journal of Philosophical Logic* 41, no. 6 (2012): 957–981; Hanti Lin and Kevin T. Kelly, "A geo-logical solution to the lottery paradox, with applications to conditional logic," *Synthese* 186 (2012): 531–575.

that the public will be careless if she increases the speed limit, update her credences, and frame her decision problem as follows (Table 2.4).

Action	*Careless*	*Careful*	*Expected Value*
Increase speed limit	$-1000, 1$ $p = 0$	$0, 1$ $(1 - p) = 1$	$0, 1$
Maintain speed limit	$-2, -1$ $q = 0.45$	$0, -1$ $(1 - q) = 0.55$	$-0.9, -1$

Table 2.4: The Low Risk Problem Avoided

Even though increasing the speed limit has a positive probability of violating a lexical priority, this fact is excluded from the official's deliberations. Updating her beliefs on the fact that the public is sufficiently unlikely to be careless (given that she increases the speed limit), the official maximises expected moral value within the revised decision problem by increasing the speed limit.[21] In this way, the official upholds her commitment to lexical priorities while also avoiding the Low Risk Problem. However, as we shall see, positing any kind of probabilistic threshold seems to raise a further problem for lexical priority theories.

2.5 The Agglomeration Problem

The Agglomeration Problem consists of cases where a lexical priority theory offers inconsistent verdicts about what ought to be done, depending on whether moral considerations are responded to separately or together. There are two versions of the Agglomeration Problem: one conjunctive and the other disjunctive. As critics of lexical priority theories note, both versions of the problem result from positing a probabilistic threshold. They conclude that lexical priority theories therefore cannot appeal to probabilistic thresholds to avoid the Low Risk Problem. Below, I will explain how lexical priority theories

---

21. This approach is similar to that proposed in Hawley, "Moral Absolutism Defended." One main difference is that Hawley's threshold approach, being based on a fixed probability value rather than an odds-based threshold, does not obey plausible principles of belief revision and also encounters Lottery Paradox-style Agglomeration Problems. See: Section 2.5.

can avoid both versions of the Agglomeration Problem, without abandoning a probabilistic threshold.[22]

The Conjunctive Agglomeration Problem involves cases where each of an agent's options has some moral property (say, that of respecting a lexical priority), but the conjunction of those options has a different property (say, that of violating a lexical priority). To illustrate, suppose that the official has just received additional information about the carefulness of the public, indexed to weekday behaviour and weekend behaviour. To simplify the following discussion, we will assume that the official should raise the speed limit if and only if there is a live possibility that the public is careful all week; otherwise, maintaining the speed limit is the most appropriate option available.

Given this simplifying assumption, if the official were to apply a probabilistic threshold, t, she might face the following dilemma. Suppose that the probability that the public will be careful on weekdays is greater than or equal to t (hence, it is a live possibility) and the probability that the public will be careful on weekends is greater than or equal to t (hence, a live possibility). It therefore seems that the official should increase the speed limit. However, it also seems possible that these states are incompatible (or, at least, anti-correlated), such that the probability that the public will be careful on both weekdays and weekends could be less than t (and, hence, not a live possibility), in which case she should not be open to increasing the speed limit. What ought she to do? Lexical priority theories appear to offer no guidance about what should be done. As it stands, the official seems to face an especially implausible kind of moral dilemma: if she decides on the basis of the time periods taken separately, she should perform an action that would be ruled out if she were to decide on the basis of the time periods taken together (Table 2.5).[23]

In response, it is first worth noting that nothing about this problem relies on the notion of lexical priority. The problem is more generic: when a threshold is

---

22. This discussion addresses synchronic versions of the Agglomeration Problem. For a discussion of strategies for solving the diachronic versions of the Agglomeration Problem, see: Chapter 3.

23. As noted in: Jackson and Smith, "Absolutist Moral Theories," pp. 276-278; Huemer, "Lexical Priority and the Problem of Risk," pp. 336-339.

State	Probability $(p, q, r,$ given **Increase Speed Limit**$)$
Public is careful on weekdays	$p \geq t$
Public is careful on weekends	$q \geq t$
Public is always careful	$r < t$

Table 2.5: The Conjunctive Agglomeration Problem.

posited, there can be cases where the relevant conjuncts fall on one side of the threshold, but the conjunction falls on the other.[24] This should be encouraging for lexical priority theorists: the Agglomeration Problem is, strictly speaking, orthogonal to the concept of lexical priority; the problem's source – as well as its solution – lies elsewhere.

Note, also, that unlike the previous problems, the Conjunctive Agglomeration Problem is based on an informal description of the decision situation. This is important because, as it turns out, the problem disappears when it is placed in a more formal decision-theoretic framework. The informality of the Conjunctive Agglomeration Problem exploits an ambiguity in how to describe the relevant states in a decision problem. The decision-theoretic model forces a resolution of this ambiguity, thereby preventing the dilemma from arising.

In this case, the ambiguity driving the Conjunctive Agglomeration Problem is that there are two ways for the public to be careful: by being careful during a particular time period only (say, weekdays) or by being careful during that period and the remaining time period (say, weekdays and weekends). Since different degrees and types of risk may be associated with these different possibilities, they should be distinguished. A more accurate specification of the states of the decision problem would be as shown in Table 2.6.

Applying the relevant probabilistic threshold truncates the decision problem to rule out states whose probability is less than t. This rules out the possibility that the public is careful all week. For simplicity, we will also assume that the probability that the public is careless all week is also less than t. Having ruled out these possibilities from her deliberations, the official updates her beliefs

---

24. See, for example, Henry Kyburg's epistemological puzzle, the Lottery Paradox, which is structurally almost identical to the Agglomeration Problem. See: H.E. Kyburg, *Probability and the Logic of Rational Belief* (Middletown, CT: Wesleyan University Press, 1961).

State	Probability (p, q, r, s, given **Increase Speed Limit**)
Public is always careful	$p < t$
Public is careful on weekdays only	$q \geq t$
Public is careful on weekends only	$r \geq t$
Public is always careless	$s = 1 - p - q - r$

Table 2.6: A more accurate specification of the states.

and maximises expected moral value (Table 2.7).[25]

Action	*Public is careful on weekdays only*	*Public is careful on weekends only*	*Expected Value*
Increase speed limit	$-10, 1$ p	$0, 1$ $(1-p)$	$-10p, 1$
Maintain speed limit	$0, -1$ q	$0, -1$ $(1-q)$	$-0, -1$

Table 2.7: The Conjunctive Agglomeration Problem Avoided.

Given the arbitrary values above, the official should, in this case, maintain the speed limit. The more important point, however, is that there is no dilemma.

Moreover, there will be no dilemma, whatever values are plugged into the decision problem. By disambiguating the informal description of the case, we created a well-defined decision problem relative to which the official can maximise expected moral value. The Conjunctive Agglomeration Problem has been solved.

The Disjunctive Agglomeration Problem relies on a similar ambiguity about the relevant state-space of a decision problem. To illustrate, let us again assume that the official should raise the speed limit if and only if there is a live possibility that the public is careful all week; otherwise, maintaining the speed limit is the most appropriate option available.

25. As mentioned in the previous section, using an odds-based threshold rule, the official's belief revision will obey Bayesian Conditionalisation (meaning that the probabilities are updated in proportion to their prior relative probabilities, renormalised so as to sum to 1. See: Lin and Kelly, "A geo-logical solution to the lottery paradox, with applications to conditional logic"; Lin and Kelly, "Propositional Reasoning that Tracks Probabilistic Reasoning"

Suppose now that the official's evidence indicates that it is sufficiently improbable that the public is careless on weekdays, such that she should exclude that possibility and be provisionally open to raising the speed limit. If it is also sufficiently improbable that the public is careless on weekends, then it seems that she should therefore hold that there is a live possibility that the public is always careful and so should raise the speed limit.

Given this evidence, however, the following situation is also possible: it may also be sufficiently probable that the public is careless during weekdays or the weekend. Even though the probability of each disjunct falls below the threshold, the sum of their probabilities may be above it. In this case, the official cannot exclude the possibility that the public is careless during at least one of the time periods, whichever it happens to be. Given these assumptions, it appears that the official faces a strange kind of moral dilemma: if she decides on a 'time-period-by-time-period' basis, she ought to raise the speed limit; if she decides 'all together', then she ought to maintain the speed limit. It seems that whatever she chooses, she will be acting wrongfully.

As we shall see, the Disjunctive Agglomeration Problem exploits the fact that the concept of lexical priority underdetermines the appropriate decision-theoretic representation of a moral decision problem. The Disjunctive Agglomeration Problem consists of cases where there are conflicting but seemingly equally eligible ways of specifying the relevant states of a decision problem. For example, taken on a 'time-period-by-time-period' basis, the relevant states (S) of the official's decision problem are:

S1: The public is always careful.

S2: The public is careful on weekdays but careless on weekends.

S3: The public is careless on weekdays but careful on weekends.

S4: The public is always careless.

This specification of the decision problem suggests that considerations of public safety pertain to particular, identified, risks. For instance, it may be that different subgroups or individuals in the population are exposed to heightened

risk at different times periods (such as children on weekends), and the official may be particularly concerned to ensure that disproportionate risk is not imposed on those subgroups or individuals. That the public is careless at some time or other is not relevant, given this specification of the decision problem.

However, having specified the decision problem as above, standard rehearsals of the Disjunctive Agglomeration Problem then make salient an alternative framing of the decision problem, whereby unidentified risk is morally relevant. This leads to a different specification of the decision problem's relevant possibilities, such as:

S1*: The public is always careful.

S2*: The public is at least sometimes careless.

On this approach, evidence about whether the public is careful is only relevant insofar as it indicates whether S1* or S2* is true. The official need not bother about which particular time period the public is careless during, so long as she is sufficiently certain that there is some time during which the public faces an increased risk of death.

The Disjunctive Agglomeration Problem thus raises a valuable question about what lexical priority theories, and moral theories in general, really care about. In the official's case, the question is: should she care about identified or unidentified risk imposition? Strictly speaking, the concept of lexical priority is silent about this. The Disjunctive Agglomeration Problem thus shows that lexical priority theories are under-theorised in this important respect. In order to provide determinate advice in risky situations, such theories must explain how we should frame decision problems. To do this, they must spell out their substantive commitments in more detail.

As it happens, when it comes to the ethics of distributing risk, it is usually not enough to know that some quantum of risk is being imposed, whoever it might befall; it is often important to know whether particular groups or individuals are bearing an unfair burden of the risk. Such fairness considerations would support a framing of the decision problem in terms of identified risks. On the other hand, in cases where there is a diffuse risk that affects the population equally, an official may be concerned solely with the probability that

the risk will eventuate somehow or other (this might apply, for example, when evaluating the risk of a catastrophic nuclear disaster). The framing of the decision problem is determined, in the end, by the details of the background moral theory.

Once the framing is settled, however, the Disjunctive Agglomeration Problem does not arise: the odds-based threshold ensures that only a consistent set of possibilities is included in the decision problem.[26] Once the decision problem is specified, the official should maximise expected lexicographic moral value.

2.6   Further Uncertainties

We have seen that lexical priority theories have the resources to avoid the various decision-theoretic objections that have been raised against them. The Permissiveness Problem relied on the mathematical oddities of infinite values, but lexical priorities can be modelled without infinite values. The Low Risk Problem was driven by the assumption that decision-makers should always attend to all possibilities, no matter how improbable. Lexical priority theories can explain why some possibilities should not be factored into a decision problem. The Agglomeration Problem exploited the fact that the concept of lexical priority underdetermines the appropriate framing of decision problems. Once lexical priority theories decide on how a given decision problem should be framed, they avoid the Agglomeration Problem. Each of these problems was rectified without having to retreat from the idea that some kinds of moral considerations cannot be defeated by any number of particular other considerations. In each case, the problem was actually a symptom of the chosen decision-theoretic representation. There are a few ways of interpreting this result.

One interpretation takes the failure of any such decision-theoretic critique as a foregone conclusion. After all, it seems that we know (indeed, *a priori*)

---

26. This is one of the key results in Lin and Kelly, "Propositional Reasoning that Tracks Probabilistic Reasoning"; Lin and Kelly, "Propositional Reasoning that Tracks Probabilistic Reasoning"

that the truth of moral theories depends on the adequacy of their substantive justifications, not the adequacy of their decision-theoretic representations. Faced with a problematic decision-theoretic representation of a moral theory, it is open to the moral theorist to say: 'So much the worse for your model!' On this view, the decision-theoretic critique of lexical priority theories was bound to fail because it does not engage with the substantive justifications of the lexical priorities in question.

However, even if the above is true, this does not mean that decision-theoretic critiques of moral theories are without value. In the process of responding to these objections, lexical priority theories have been forced to clarify their substantive commitments: do they really consider some considerations to be infinitely more valuable than others? Do they require that we always attend to arbitrarily small probabilities in our decision-making? Do they care about identified risks (such as protecting the lives of particular people) or unidentified risks (protecting the lives of people in general)? Even if the decision-theoretic critique fails, as argued, it has nevertheless succeeded in revealing these under-theorised aspects of lexical priority theories.

Indeed, once we scratch below the surface of lexical priority theories, we discover further puzzling features, depending on the particular kind of substantive justification at play. For example, consider the view that lexical priorities outweigh other considerations. How is this to be modelled? We have seen that positing infinite values is highly problematic. Instead, lexical priority theories could perhaps adopt a different kind of value function. One idea is to hold that the marginal moral value of upholding a lesser consideration diminishes asymptotically towards a limit, such that upholding any number of lesser considerations never has as much moral value as upholding a higher consideration.[27] This kind of moral value function could ensure that, in the official's case, no matter how many millions of passengers may benefit, pleasing an additional passenger will never outweigh the moral importance of a person's life.[28]

---

27. For discussion of this idea, see: Erik Carlson, "Organic Unities, Non-Trade-Off, and the Additivity of Intrinsic Value," *Journal of Ethics* 5 (2001): 335–360. As applied to ethics and risk, see: Chapter 3.

28. This idea has been explored with respect to 'the good' by: Erik Carlson, "Aggregating Harms — Should We Kill to Avoid Headaches?," *Theoria* 66, no. 3 (2000): 246–255; John

The question is: why should the moral importance of an additional passenger's pleasure diminish due to external factors, like the number of other passengers who happen to also be benefiting from slightly more exhilarating rides? If moral value is conditional in this way, what exactly are the conditions?

Or consider, instead, the cancellation approach to justifying lexical priorities. It holds that there is no moral value to acting upon some considerations unless we also conform to particular other considerations. For example, we might say that there is no moral value in the official deviating from her role as a protector of public safety to satisfy the passengers' need for speed: optimising passenger satisfaction has no objective moral value when it involves disregarding the lives of those who may suffer the consequences.[29] This approach has its own mysteries: for instance, how do we weigh considerations when we are uncertain if they have been cancelled? Is there such thing as partial cancellation of moral value?[30]

Finally, consider an exclusionary approach to justifying lexical priorities. This approach is silent with respect to whether lexical priorities outweigh or cancel other considerations; it instead argues that lexical priorities deliberatively *exclude* them.[31] That is, there are cases where we should act first and foremost on the balance of only some types of considerations (such as public safety), irrespective of how many other considerations of a particular kind (such as additional passenger satisfaction) are at stake. As with the cancellation approach, it is not at all obvious how exclusionary reasons operate in cases of uncertainty. Are they nothing more than useful guides for decision-

Broome, "No Argument against the Continuity of Value: Reply to Dorsey," *Utilitas* 22, no. 04 (November 2010): 494–496. For an approach that applies to multiple kinds of values, see: Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, chap. 10.

29. A related notion of conditionality of moral worth can be found in, for example: Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. Allen W. Wood (New Haven; London: Yale University Press, 2002), sec. 1.

30. As far as I am aware, the closest work that systematically addresses this question is: John Horty, *Reasons as Defaults* (Oxford University Press, 2012). However, Horty's approach does not seem to be sufficiently general for the purposes of most lexical priority theories, since it avoids both probabilistic uncertainty and the idea that moral considerations can be meaningfully weighed against each other.

31. See: Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press, 1999). Thanks to Andrew Williams for discussion.

making (similar to the threshold approach presented earlier)? Or, are they better understood as epistemic considerations that give us reason to change our credences that the world is one way or another? In any case, does an exclusionary reason's importance diminish with probability and, if so, can such 'diminished' reasons be weighed against the reasons they purport to exclude?

These are almost completely unexplored issues in moral theory. The reason is that they only become salient once we adopt a decision-theoretic perspective of lexical priority views. Although the recent decision-theoretic critique of lexical priorities fails, it has succeeded in revealing new, potentially important lines of inquiry in our moral theorising.

2.7   Conclusion

The decision-theoretic critique of lexical priority theories suggests that they have no acceptable moral decision rule for cases of uncertainty, and should therefore be abandoned. However, as shown, this critique relies on contentious modelling assumptions that lexical priority theories need not – and, indeed, should not – accept. By using additional resources in decision theory and spelling out their substantive moral commitments in more detail, lexical priority theories can guide us through risky situations.

Chapter 3

Priorities and Uncertainties

3.1 Introduction

Suppose that you are an *effective altruist*. You believe that, other things equal, your donations to charity should do as much good as possible.[1] You are in a position to give to just one of the following charities:

Save Life: This charity saves lives by providing mosquito nets to poor individuals in malaria-prone areas.

Cure Headaches: This charity cures mild headaches by distributing aspirin.

You have \$1000 to give – enough, we will suppose, to save a life or to cure ten thousand mild headaches. So, which charity should you give it to? That all depends on how morally important saving a life is compared to curing headaches. Some argue that saving life has priority over any number of more trivial considerations, such as curing mild headaches.[2] On their view, if donating to **Save Life** will lead you to save at least one life, then you should do

---

1. On effective altruism, see: William MacAskill, *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference* (Guardian Faber Publishing, 2015); Peter Singer, *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically* (Text Publishing Company, 2015); Pummer, "Whether and Where to Give."

2. See, for example: Dale Dorsey, "Headaches, Lives and Value," *Utilitas* 21, no. 01 (2009): 36–58; Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*; Voorhoeve, "How Should We Aggregate Competing Claims?"

so, no matter how many headaches you might otherwise cure by donating to **Cure Headaches**.

In Chapter 2, we saw that those who subscribe to this kind of priority can use lexicographic expected value theory to navigate risky cases. In this chapter, we will adopt an alternative approach that uses orthodox expected value theory. The key difference is this: the lexicographic model effectively posits an infinite value difference between higher- and lower-considerations; the model presented here posits only a finite value difference.[3] As we shall see, this difference helps to side-step a number of problems related to decision-making with infinite values.

Specifically, this alternative model will hold that the cumulative value of lower considerations is *bounded* in the following way: additional instances of satisfying a lower consideration contribute less than previous ones, such that they asympototically approach a 'ceiling' of value that is below the value of higher considerations. This model offers a general approach to modelling lexical priorities with standard decision theory. To illustrate, consider Figure 3.1, where the dotted line *Headache Limit* denotes the maximal limit of the value of curing headaches in that choice context.

As we can see, as the number of headaches cured approaches infinity, its value approaches but never reaches ten units of value. At the same time, the value of saving one life is ten units of moral value, such that it is *always* more valuable than *any* number of headaches. Thus, lexical priority theorists can maintain their position without having to posit infinite value differences between higher and lower considerations. I will call this the *Bounded Approach* to modelling lexical priority theories.[4]

---

3. In more technical terms, the difference is that the lexicographic model violates an axiom of von Neumann-Morgenstern decision theory called Continuity, which holds that for any three prospects A, B, and C, where A is strictly preferable to B, and B is strictly preferable to C, there is always a probability value p such that a prospect of A (with probability p) or C (with probability $(1-p)$) is equally preferable to a certainty of B. Infinite values represent a violation of Continuity because if there is an infinite value difference between A and B or B and C, then there is no (non-infinitesimal) p such that a prospect of A (with probability p) or else C (with probability $(1-p)$) is equally preferable to a certainty of B. For discussion, see: Bartha, "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities."

4. For details of the formal features of this kind of value function as a model of lexical
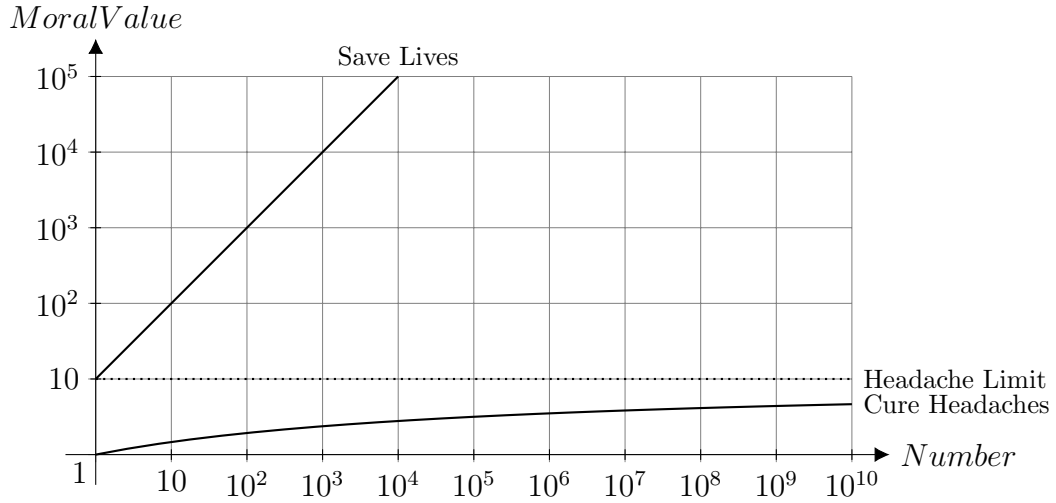
Figure 3.1: The Bounded Approach to modelling lexical priorities

Part 3.2 shows how the Bounded Approach easily avoids the **Low Risk Problem** and the **Permissiveness Problem**. However, we will see that it is particularly vulnerable to a version of the **Agglomeration Problem**. **Part 3.3** considers various solutions to the Agglomeration Problem for the Bounded Approach, eventually sketching a solution that draws on Alex Voorhoeve's *Aggregate Relevant Claims* principle.[5] **Part 3.4** turns the tables, showing how the Bounded Approach helps to defend Voorhoeve's theory against an independent objection, called the **Spectrum Problem**. In **Conclusion**, it seems that lexical priority theories have an additional viable way of guiding us through uncertainty.

3.2   Testing the Bounded Approach

The approach sketched above has a clear advantage over the lexicographic approach of the previous chapter: it deals more easily with the Low Risk Problem,

---

priority and (more generally) value superiority, see: Gustaf Arrhenius and Wlodek Rabinowicz, "Value Superiority," in *The Oxford Handbook of Value Theory* (Oxford University Press, 2015), 225–243.

5. Voorhoeve, "How Should We Aggregate Competing Claims?"

which – you will recall – involves cases where negligible prospects of upholding or violating lexical priorities outweigh any prospect of upholding or violating lower considerations. To avoid this problem, I argued that the lexicographic approach must hold that it maximises expected (lexicographic) moral value to ignore sufficiently small probabilities. This justification was based on contingent, though realistic, shortcomings of non-ideal agents in non-ideal situations. The limitation to this approach is that more capable agents than us may be able to handle extremely small probabilities in their deliberations, thereby reviving the Low Risk Problem. As we shall see, the Bounded Approach can avoid the Low Risk Problem without holding that some probabilities should be ignored.

To illustrate, let us suppose that **Save Life** is not a very effective charity. In fact, it is almost certain to have no effect at all, since its mosquito nets are generally defective and provide only false security to the individuals who receive them. Nevertheless, there is a positive probability that your donation to that charity will save a life. By contrast, **Cure Headaches** is highly effective at what it does: it reliably uses donations to cure many temporary, mild headaches. However, it is certain to save no lives. The question is: even though there is a positive probability that donating to **Save Life** will save a life, should this extremely unlikely prospect determine what you ought to do? Drawing on our evaluation of the considerations at play in Figure 3.1, you face the following decision problem (Table 3.1).

| Possibilities | NOT DEFECTIVE | DEFECTIVE | |
Probabilities	0.01	0.99	EMV
Give to **Save Life**	10	0	0.1
Give to **Cure Headaches**	≈ 5	≈ 5	≈ 5

Table 3.1: Avoiding the Low Risk Problem

In this case, you ought to give to **Cure Headaches**, even though it is possible that your donation to **Save Lives** might be successful in saving a life. Thus, the Bounded Approach clearly avoids the Low Risk Problem without having to posit that risks must meet some threshold of sufficient probability. The reason is that since it does not adopt infinite values, it does not treat

any positive (real-valued) probability of upholding a higher consideration as more valuable than any prospect of upholding a lower one. This is an important result: it shows that even though lexical priority theories do not permit trade-offs between higher and lower considerations, they can allow trade-offs between prospects of them. As such, the Bounded Approach constitutes a more moderate – and perhaps to many a more plausible – approach to dealing with risk.

It is easy to see that the Bounded Approach also easily avoids the Permissiveness Problem. Even if we were to suppose that there were also risks associated with giving to **Cure Headaches**, you would still be able to distinguish between the expectations of your choices, since the values are represented by finite, real-valued numbers.

But what of the Agglomeration Problems? Does the Bounded Approach face cases where framing effects lead you to contradictory courses of action? Let's see.

Windfall: Faced with the Low Risk Problem, you chose to give to **Cure Headaches** because doing so maximised expected moral value. But guess what: here's another $1000 to donate! Your options are the same as before. The question is: who should you donate it to?

To answer this question, we need to know whether your previous choice bears on your current calculations of expected moral value of your current choice. As Table 3.2 shows, if you count the headaches your previous donation will cure, then your donation to **Cure Headaches** has less value than previously (0.001 versus 5) due to the diminishing moral value of curing headaches. As such, it seems that you ought to give to **Save Life** instead. Table 3.3 gives the opposite result: if you do not count your previous donation, then you ought to give to **Cure Headaches**, just as you did before. The problem here is that you are faced with conflicting, but seemingly equally eligible, verdicts about what to do. Thus, the Bounded Approach faces its own version of the Agglomeration Problem.

Possibilities	NOT DEFECTIVE	DEFECTIVE	
Probabilities	0.01	0.99	EMV
Give to **Save Life**	10	0	0.1
Give to **Cure Headaches**	≈ 0.001	≈ 0.001	≈ 0.001

Table 3.2: **Windfall** with Act Inseparability

Possibilities	NOT DEFECTIVE	DEFECTIVE	
Probabilities	0.01	0.99	EMV
Give to **Save Life**	10	0	0.1
Give to **Cure Headaches**	≈ 5	≈ 5	≈ 5

Table 3.3: **Windfall** with Act Separability

3.3 Solving the Agglomeration Problem

I will now outline how the Bounded Approach might avoid the Agglomeration Problem. Before doing so, however, it is important to note that the Agglomeration Problem (like its previous incarnations concerning the lexicographic model of Chapter 2) has nothing to do with risk as such. Instead, the source of the problem is that our lexical priority theory is underspecified in a crucial respect. Unlike in Chapter 2, this crucial respect does not concern the question of whether moral considerations should be understood *de dicto* or *de re* (see: Section 2.5). Even after we specify whether our moral theory cares about *particular* people or people *in general*, there is still the problem of figuring out whether to 'count' the moral value of our past actions. Put more precisely, the challenge is to give an account of whether our actions are, from a moral point of view, 'separable': that is, whether our actions across time should be morally evaluated separately or together.[6]

Defenders of the Bounded Approach have three available strategies for resolving the ambiguity about action separability. The first holds that our actions are always separable, which in **Windfall** leads to the verdict that you ought to give to **Cure Headaches**. The second strategy holds that actions

---

6. Holly S. Goldman, "Dated Rightness and Moral Imperfection," *The Philosophical Review* 85, no. 4 (1976): 449–487; Frank Jackson and Robert Pargetter, "Oughts, Options, and Actualism," *The Philosophical Review* 95, no. 2 (1986): 233–255; Brian Hedden, "Options and Diachronic Tragedy," *Philosophy and Phenomenological Research* 87, no. 1 (June 2013): 1–35

are always morally inseparable, which leads to the conclusion that you ought to give to **Save Lives**, instead. As we shall see, each of these approaches is deeply problematic. As such, I will consider versions of a third strategy that aims to give a principled explanation for why and when actions are separable sometimes but not others.

3.3.1   An Instrumental Justification

Suppose that our actions are always *separable* from a moral point of view, such that the 'counter' of the bounded value function is reset each time a choice is made. As mentioned, this approach will lead you in **Windfall** to give to **Cure Headaches**, since the previous cured headaches no longer 'count'. Although this solves the Agglomeration Problem, it is not a satisfactory solution.

The main problem is this: holding that actions are always separable leads to strange cases where decomposing a decision into smaller decisions leads to conflicting verdicts. For example, in **Windfall**, suppose that you now have the option of donating via a payment plan, which allows you to give your $1000 in smaller instalments over the course of a year. We will stipulate that neither approach has an advantage over the other in terms of the number of headaches cured or any other morally-relevant factors. As such, we should think that you are equally permitted to pay all-at-once or in instalments to **Cure Headaches**. Unfortunately, if lexical priority theories hold that our actions are always separable, then they cannot treat the two options as equal. This is because the payment plan will have higher moral value than giving the money all at once, since each instalment will cure headaches whose moral value will be undiminished by the previous cured headaches. This is clearly an embarrassing result of defenders of the Bounded Approach: why should a trivial matter such as the existence of a payment plan option determine what you morally ought to do, when everything else is held equal? For this reason, it seems that they cannot hold that actions are always temporally separable.

So, let us now consider the position that actions are always *inseparable*, such that the headaches you cured in the past bear on your choice today. The first question is: what could justify this position? Some have argued for the

following answer: perhaps you want to make the world a better place and it turns out that, instrumentally speaking, curing additional headaches does not help as much to bring about what really matters.[7] For example, suppose that higher prevalence of mild headaches in society increases various risks to life – such as decisions made in the medical profession, driving, law enforcement, and so on – and that reducing the prevalence of mild headaches will have high, but diminishing, effects on overall safety. In this case, inseparability is explained by the background causal assumptions about how, say, curing headaches saves lives. This seems to be a *prima facie* plausible way of justifying inseparability of actions.

Bracketing the fact that this instrumental justification seems rather tenuous, there is a larger problem at hand: namely, if actions are always inseparable, then seemingly irrelevant actions from our past are going to determine the value of our current available actions. How implausible this is depends on whether our moral theory cares about the moral quality of our own actions or of states of affairs (where the latter defined so as to exclude your action).

Suppose that our moral theory holds that, other things equal, we should minimise the occurrence of mild headaches. Such a moral theory might hold that you should refrain from curing others' headaches if your omission leads to others curing greater numbers of headaches than you ever could, since the moral theory simply 'wishes' for there to be maximal headaches cured. By contrast, a moral theory that cares about the moral quality of your actions might hold that you should maximise *your* curing of headaches, such that you should cure headaches even if doing so leads to a lesser overall number of people being cured.[8]

If a lexical priority theory is concerned with achieving particular states of affairs and takes actions to be always morally inseparable, then it would seem that all of humanity's cured headaches of the past should bear on the current

---

7. For discussion of this instrumental justification, see: Michael Ridge, "How to Avoid Being Driven to Consequentialism: A Comment on Norcross," *Philosophy & Public Affairs* 27, no. 1 (1998): 50–58. See also: Dorsey, "Headaches, Lives and Value."

8. This is roughly similar to the agent-relative/neutral distinction, which itself not a clearly defined concept. For discussion, see: Matthew Hammerton, "Distinguishing Agent-Relativity from Agent-Neutrality," *Australasian Journal of Philosophy*, 2018, 1–12.

value of your curing additional headaches. This seems absurd, however, since there is no apparent reason why past actions of individuals long-past should bear on the value of your curing someone's headache here and now. Moreover, there is the epistemic difficulty of estimating how many headaches have ever been cured in the course of human history. As it stands, this approach seems untenable.

Focusing instead on the moral quality of your actions slightly mitigates this irrelevance issue, since then the moral value of your curing a headache depends on how many other headaches *you* have cured in the past. But this latter position is completely baffling, particularly on the instrumental justification presented above, since your curing of headaches would have negligible effects on society's ability to secure what really matters. There must be something else that is explaining action inseparability. We will next consider an approach that offers such an explanation.

3.3.2 The Discount Approach

In 'Moral Sunk Costs', Seth Lazar offers a principled justification for how our actions can be inseparable in some choice contexts, but not others.[9] In cases like **Windfall**, the crucial factor is something we have so far overlooked: whether each of your donations will affect the same people or different people. Lazar's insight is that in cases involving trade-offs between people, even if the potential welfare benefits to the people are held constant, our actions may nevertheless express differing degrees of respect towards their status as moral agents. These latter 'status-based' considerations are an additional moral dimension of our actions aside from welfare-based considerations, and they explain why our actions may be inseparable.

To illustrate, recall that in **Windfall**, you have already given to **Cure Headaches**, providing each individual with a ready-to-use aspirin. Suppose now that your second donation would give those *same* individuals one additional aspirin for some future mild headache. At the same time, since in your first choice you did not give to **Save Life**, there is an individual who has

---

9. Seth Lazar, "Moral Sunk Costs," *The Philosophical Quarterly*, 2018, 1–40.

not yet received any prospective benefit. By giving to **Cure Headaches** a *second* time, you would be promoting just as much welfare as before (let us suppose), but your repeated preferential treatment towards to the headache sufferers would arguably count as disrespecting the potential beneficiary of **Save Life**. Here's one reason why: having donated to **Cure Headaches**, you have shown respect to these individuals, who could not say that they would be treated unfairly if you were to refrain from donating to them a second time; by contrast, if you were to give to **Cure Headaches** a second time, you would effectively be asking the potential malaria sufferer to forego a cumulative 1/50 probability of avoiding malaria just so that others can enjoy a second round of aspirin. Asking her to forego this chance is arguably unfair, constituting a failure of respect towards her as a moral agent of equal moral status.[10]

In this way, your decision in **Windfall** involves both wellbeing-based considerations and status-based reasons. Lazar holds that the importance of these wellbeing-based considerations would be the same in both choices, since the donations will have the same impact on wellbeing as previously. However, he argues that due to how your actions might express respect or disrespect to the potential beneficiaries, your status-based considerations in **Windfall** are different. Specifically, having given to **Cure Headaches** already, your status-based considerations to benefit those individuals are now weaker than before. As such, since the moral value of your donation is a function of both your wellbeing-based considerations and your status-based considerations, you have less reason than before to give to **Cure Headaches** and should perhaps give to **Save Life** instead. Thus, we have a justification for why we should treat **Windfall** as involving morally inseparable actions.

Note, by contrast, that if **Windfall** involves different potential beneficiaries, your status-based considerations in favour of giving to **Cure Headaches** do not diminish. As such, you should treat your actions as being effectively separable, and give to **Cure Headaches**.

Although this approach offers a compelling justification for considering

---

10. If you are unsure about the specific probabilities in this case, repeat the choice scenario, giving to **Cure Headaches** each time, until it becomes near certain that you could have saved the individual from malaria. At some point, it will be compelling that the potential beneficiary has been wronged by your repeated neglect.

two actions to be inseparable, it has difficulty upholding the idea of lexical priority that we are aiming to defend here. This is because wellbeing-based considerations do not diminish in value; rather, they have constant value. As such, they will never tend towards an insurmountable upper ceiling of value that is less than the value of a higher consideration. For this approach to uphold lexical priority of life over headaches, an amendment is needed.

For instance, we might posit not only that there are diminishing status-based considerations to donate to the same beneficiaries of **Cure Headaches**, but also that negative status-based considerations accumulate against disrespectful actions. These negative status-based reasons would offset the value of the wellbeing-based considerations, potentially creating an upper limit of value that is below that of a higher consideration. The problem with this approach is that these negative status-based considerations will only emerge after repeated choices. In one-shot cases, there will be no such offsetting, hence it is possible that there will be some number of trivial wellbeing-based considerations that have collectively greater value than that of a higher consideration. There may be other amendments within the status-based-reason approach, but it seems that they will be revisionary or potentially *ad hoc* extensions to Lazar's original theory.

Overall, even though appealing to status-based considerations gives a principled account of when to treat actions as inseparable, it is unclear that it can justify lexical priorities. We need an account that can do both.[11]

3.3.3   An Alternative Justification

I will now sketch an alternative justification that not only gives a principled explanation for why some actions are inseparable, but also upholds the idea that some moral considerations have lexical priority over others. This justification will be based on Alex Voorhoeve's principle, *Aggregate Relevant Claims*.[12] Although it is by no means a complete proposal, it hopefully points the way towards future progress on this question of the separability of actions.

---

11. For a recent account that aims to do both, but which does not subscribe to the Bounded Approach, see: Lazar, "Limited Aggregation and Risk."

12. Voorhoeve, "How Should We Aggregate Competing Claims?"

Let a *claim* be a moral consideration related to another person's wellbeing. If someone has a claim against you, then (other things equal) you ought to promote their wellbeing. Say that two or more claims (e.g. for resources or aid) are *competing* from your point of view when they are held by different individuals, but if you satisfy some of those claims, you cannot satisfy others. How should you decide between competing claims? The principle Aggregate Relevant Claims offers the following prescription: you should disregard any claims that are, on a one-to-one comparison with all other claims, sufficiently less important from a wellbeing point of view.[13]

This approach is motivated by a principle called the *Separateness of Persons*, which states that we should take the interests of each individual separately, rather than viewing their interests as being subsumed in an aggregated mass of others' interests.[14] This principle requires that trade-offs between individuals be justifiable from the points of view of each of the individuals involved. Thus, if we respect the Separateness of Persons, we cannot say to an individual whose life we could save that their claim was defeated by someone's relatively trivial claim of wanting their temporary, mild headache cured, since a claim of having a headache cured cannot be justified to a person whose life would be saved. In Voorhoeve's terminology, these latter claims are deemed 'irrelevant' to the former, such that no amount of headaches aggregates to outweigh a life. In this way, Aggregate Relevant Claims preserves the lexical priority of life over headaches.

This does not mean, however, that the principle never permits aggregation. As its name says, in cases where competing sets of claims are 'relevant' – that is, when they are sufficiently similar in their wellbeing implications – then we ought to choose the course of action that has the highest aggregate benefits for wellbeing. This is consistent with the Separateness of Persons, since these relevant claims may be justifiably satisfied to each individual involved. In

---

13. Voorhoeve also allows that we have a personal prerogative to weight the moral importance of our own wellbeing higher than that of others, within limits of what can be mutually justified.

14. Thomas Nagel, "Egalitarianism," chap. 7 in *Equality and Partiality* (Oxford: Oxford University Press, 1991); Scanlon, *What We Owe to Each Other*, p. 238-41; Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, p. 100.

this way, Aggregate Relevant Claims is able to posit a lexical priority relation between higher and lower claims, while also allowing for optimisation regarding the satisfaction of the higher claims. As such, it is suitable candidate theory for justifying the Bounded Approach.

If we understand moral value as a measure of the importance of satisfying claims, then Aggregate Relevant Claims implies that the moral value of a claim is not separable from the other claims with which it competes. For instance, we saw that the presence of a life-saving alternative action makes it such that no amount of curing headaches will ever be as morally important as that of saving a life. Understood in terms of moral value, the presence of this alternative changes the moral value of curing headaches. In particular, the value of curing headaches never aggregates to be greater than that of saving a life. Clearly, this ranking of claims can be modelled in terms of moral value using the Bounded Approach.

Combining Aggregate Relevant Claims and the Bounded Approach yields the following result: in the absence of this higher consideration – for instance, if the choice were instead between headaches and some other trivial harm, such as hangnails – then curing headaches does not have bounded moral value. Aggregate Relevant Claims thus offers a justification for the particular and rather peculiar behaviour of the Bounded Approach's value functions. Moreover, this approach also explains why some actions are separable from other actions. However, to see this, we need to extend Voorhoeve's theory to cover competition between claims across time, not just at a time.

Suppose that you have the choice between giving now and giving later. For whatever reason, you are presently able to help cure headaches, but there is no person whom you can help save from malaria. This is not because there is no one at risk of malaria; rather, there is no way you can help them right now. As it stands, for all practical purposes, there are no claims that are competing at the same point in time. However, suppose also that you believe that if you wait a day or so, you will be able to donate to help save someone from malaria. The problem is: if you choose to cure headaches now, you will not have any money to donate to help this person avoid malaria. These claims are *competing across time.*

I propose that, in these types of cases, the value of giving to **Cure Headaches** should be bounded because those claims are irrelevant to the competing claims of the person you may soon be able to save from malaria. More generally, I suggest that our actions are inseparable insofar as they adjudicate between claims that are competing across time. To see why, consider the alternative: if claims that you cannot presently serve do not matter even though you will soon be able to serve them, you may be permitted (or even required) to squander your resources on less important moral causes.[15] After all, ambulances don't help people move furniture when there are no emergencies afoot. More generally, morality requires forethought, not just reacting to present circumstances.

But what about your past actions? Barring cases of backwards causation, you cannot serve claims of the past if they no longer exist. If those claims no longer exist, then they are not competing with any present or future claims. Alternatively, if those claims continue to exist, then they may compete with other present and future claims, depending on which claims can be jointly served.

For instance, in **Windfall**, the question is: which present and future claims are competing? The answer depends on whether the beneficiaries of each charity *continue* to have claims. On a natural reading of the case, the very same claims are being assessed again. As such, the value function at play is the same as before, except that you are now evaluating curing additional headaches (point Y in Figure 3.2) on top of those you have already contributed to curing (point X). That is, since the same claims are still competing, your actions regarding those claims are inseparable. By contrast, if your new choice in **Windfall** is between different competing claims (held by different particular claimants), then it is a new choice, and your previous decision is separable from your current one. This separability of your choices is suggested by the principle of the *Separateness of Persons*, which holds that changing the identities of the claimants changes the moral evaluation of your action, since persons are not exchangeable even if their claims are otherwise the same.

One might worry that other versions of the Agglomeration Problem are

---

15. Here I am following Lazar's terminology is talking about 'serving' claims. Lazar, "Limited Aggregation and Risk."
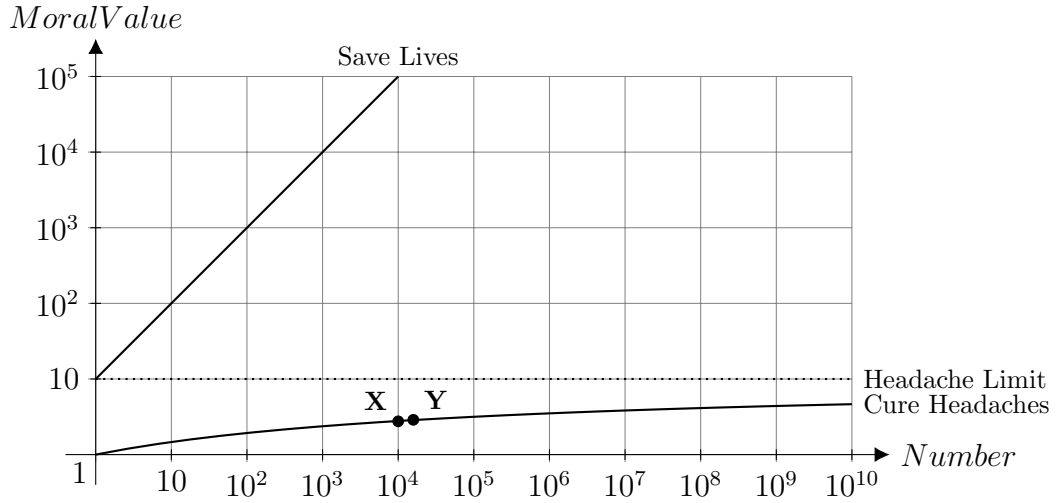
Figure 3.2: Avoiding the Agglomeration Problem by appealing to Competing Claims Across Time

lurking in the background. For example, it seems that changing the relevant claimants for each decision will reset the 'counter', such that over a sequence of actions, you should repeatedly choose to promote the interests of the headache sufferers, not the potential malaria sufferer. However, I don't think that follows. If the potential malaria sufferer – that is, the person with the 'lexically prior' claim – is the same, then repeated neglect of this claim accumulates even if we substitute the identities of the headache sufferers. Given that this person can justifiably complain about this neglect, this suggests that we should take the most important claims as anchoring the value of additional efforts to serve irrelevant claims, whoever they happen belong to. If we change the identity of this person, however, then the 'counter' resets because there is not accumulated neglect of their claims. This seems like a *prima facie* plausible approach to dealing with extensions of the Agglomeration Problem, which is largely in keeping with both Aggregate Relevant Claims and Lazar's Discount Approach.

To recap, I have argued that Aggregate Relevant Claims can be plausibly extended to give a principled approach to dealing with the Agglomeration Problem. Firstly, in a given choice, you must determine which claims you

can serve now and in the future. Second, you must assess whether these claims are competing. If they are, then your actions regarding these claims are inseparable. If these claims are not competing – for example, if giving to one charity now will not affect your ability to give to another charity in the future – then your actions are separable. In either case, you should perform the sequence of actions that maximises expected moral value. In particular, in **Windfall**, you should give to **Save Life**, since you have already satisfied a number of competing claims by having given to **Cure Headaches**. As such, your decision looks like Table 3.2.

A final question: what if you are uncertain about whether two claims are competing, where one of the claims is irrelevant to the other? In such cases, you would be uncertain about whether or not one of the claims has bounded value: if they are competing, then the irrelevant claim has bounded value, but if they are not competing, then it does not. On the face of it, this seems like another Agglomeration Problem. However, this problem may be solvable using standard decision-theoretic approaches. Namely, to make progress in these cases, it seems that you should consult the probabilities that the claims are competing, discount the possible values of the irrelevant claims accordingly, and maximise expected value.

This is obviously only a sketch of a solution to the Agglomeration Problem. Nevertheless, it seems that a Bounded Approach based on Aggregate Relevant Claims presents a promising way for lexical priority theories to avoid the Agglomeration Problem, as well as other risk-based objections. Overall, it constitutes an additional decision-theoretic approach to the lexicographic decision-theoretic approach presented in Chapter 2. I will next argue that the Bounded Approach helps to rescue Aggregate Relevant Claims from a serious theoretical challenge.

3.4   The Spectrum Problem

We have seen that the Agglomeration Problem for the Bounded Approach has nothing in particular to do with risk.[16] Rather, the problem arises because crucial details about the background moral theory are missing. I argued that Voorhoeve's Aggregate Relevant Claims principle presents a good candidate for filling in these details, thereby saving the Bounded Approach from the Agglomeration Problem. However, we will see that Voorhoeve's approach faces its own serious challenge: so-called *Spectrum Problems*.[17] To illustrate, consider the following case:

Headaches, Migraines, and Life: We know that claims to have a temporary, mild headache cured are not relevant to claims to have a life saved. However, it seems plausible that they are relevant to claims to having a *migraine* cured and, also, that claims to have a migraine cured are relevant to those regarding lives. If that is the case, then which of the following options should you choose?

> **Save Life:** Save one life.
>
> **Cure Migraines:** Cure 10^3 migraines.
>
> **Cure Headaches:** Cure 10^{10} mild, temporary headaches.

In this case, let us assume (perhaps implausibly) that 1000 migraines aggregate to outweigh a life. Suppose also that 10^{10} mild, temporary headaches aggregate to outweigh 1000 migraines. By transitivity, it follows that there is some number of headaches (in this case, 10^{10}) that outweighs a life. However, this contradicts Voorhoeve's position that such trivial claims never aggregate to morally outweigh a life. What to do?

---

16. See also: Chapter 2.

17. Larry S. Temkin, "A Continuum Argument for Intransitivity," *Philosophy and Public Affairs* 25, no. 3 (1996): 175–210; Alastair Norcross, "Comparing Harms: Headaches and Human Lives," *Philosophy and Public Affairs* 26, no. 2 (1997): 135–167; Derek Parfit, "Justifiability to Each Person," *Ratio* XVI, no. 4 (2003): 368–390; Johanna Privitera, "Aggregate Relevant Claims in Rescue Cases?," *Utilitas* 30, no. 2 (2018): 228–236.

One might reject one of the initial assumptions. For example, we might reject the idea that temporary, mild headaches are relevant to migraines. Alternatively, one might reject the idea that migraines are relevant to lives. However, this doesn't work: for harms like headaches, we can easily come up with a more gradual sequence of headache-like symptoms ranging from temporary and mild to ongoing and excruciating. It seems dogmatic to assert that somewhere along this smooth continuation of headache-like symptoms, there is a sharp cutoff in terms of wellbeing such that a claim to have a headache cured has lexical priority over a claim to have a very similar headache cured, or that a claim to have a life saved has lexical priority over a claim to have cured a migraine that makes life not worth living.[18]

To avoid such implausible conclusions, a proponent of *Aggregate Relevant Claims* might hold that in these more fine-grained sequences, it becomes *indeterminate* whether competing claims regarding slightly different types of headaches are relevant to each other. This approach might then allow you, the decision-maker, to have the option of resolving the indeterminacy either by treating two adjacently similar claims to be either relevant or irrelevant to each other.[19] By accepting that there is no complete ordering of claims in terms of relevance, a defender of Aggregate Relevant Claims might thereby avoid Spectrum Cases. However, allowing for this kind of indeterminacy is clearly a theoretical cost, since it undermines the theory's ability to guide action in cases involving similar competing claims.

Voorhoeve's own approach to this type of case is that you should adopt the following sequential decision process.[20] First, you determine whether there is unanimous agreement that any particular claim is irrelevant to any other particular claim. In this case, all agree that claims to have a headache cured are irrelevant to saving a life. At this first stage of the process, you should therefore eliminate **Cure Headaches**. The second stage then has you consider "the

---

18. See also: Sergio Tenenbaum and Diana Raffman, "Vague Projects and the Puzzle of the Self-Torturer," *Ethics* 123, no. 1 (2012): 86–112.

19. For the classic formalisation of this supervaluationist approach to indeterminacy (specifically, vagueness), see: Bas C. van Fraassen, "Singular Terms, Truth-Value Gaps, and Free Logic," *Journal of Philosophy* 63, no. 17 (1966): 481–495.

20. Voorhoeve, "How Should We Aggregate Competing Claims?," pp. 77-8.

remaining eligible alternatives" regarding which there is no unanimous agreement: in this case, you must consider the claims of the migraine sufferers versus that of the person whose life could be saved. Since they are relevant to each other, you should aggregate. This will lead you to choose **Cure Migraines**.

In a recent response to Voorhoeve's view, Johanna Privitera argues that this approach is unsatisfactory because, from the perspective of the mild headache sufferers, their claim has been unfairly defeated by those of the migraine sufferers.[21] After all, the former's claims were only defeated by that of the person whose life could be saved, not by the migraine sufferers' claims. Thus, it seems that Voorhoeve's approach does not provide a justifiable resolution to cases like **Headaches, Migraines, and Life**.

Another way of putting the complaint is that Voorhoeve's verdict depends on a seemingly arbitrary way of ordering the choices. Were you to start with a different but seemingly equally justifiable pair-wise comparison, you would come to a different but equally justifiable conclusion. For instance, suppose that you were to run the decision process in this way: starting with a comparison between **Cure Headaches** and **Cure Migraines**, you eliminate **Cure Migraines** – after all, everyone agrees that the migraine sufferers' claims are outweighed by the claims of the mass of headache sufferers. In the second stage, you then consider **Cure Headaches** and **Save Life**. Everyone involved agrees that the claims to cure headaches are irrelevant to saving life, so you should choose **Save Life**. At no point have you violated the precepts of Aggregate Relevant Claims, yet you have come to a different conclusion that seems just as justifiable as Voorhoeve's alternative.

To avoid the Spectrum Problem, Aggregate Relevant Claims does not need to adopt a sequential decision process that compares options two-at-a-time. Instead it can adopt the Bounded Approach and compare all options at once. By using well-defined values to clarify how the strength of claims are measured and compared in cases like **Headaches, Migraines, and Lives**, the Bounded Approach will impose transitivity on the ordering of competing claims. *Aggregate Relevant Claim's* anti-aggregationist commitments will be preserved by holding that when a claim is irrelevant to another claim in the choice con-

---

21. Privitera, "Aggregate Relevant Claims in Rescue Cases?"

text, it has diminishing moral value that converges to some finite limit below that of the sufficiently stronger claim. The potential beneficiaries of **Cure Headaches** can recognise that in a choice context in which their claim is irrelevant to some other claim, their claims have limited importance (as marked by the value limit: Headache Limit).

The view's aggregationist commitments will be preserved by holding that relevant competing claims do not have that kind of diminishing moral value. Figure 3.3 illustrates this approach, showing that in **Headaches, Migraines and Lives**, **Cure Headaches** has diminishing marginal moral value, but **Cure Migraines** and **Save Life** do not.
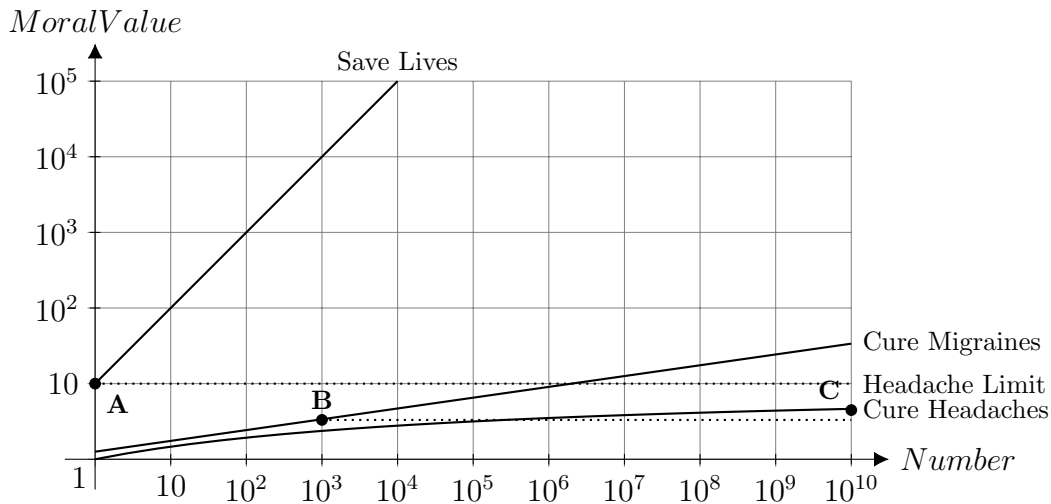
Figure 3.3: Avoiding the Spectrum Problem

Importantly, the Bounded Approach shows that the Spectrum Problem is incoherent. Once the relevant considerations have been assigned their respective values in this choice context, it will not be possible that your actions are intransitively ordered.[22] It will not be possible for a life to be outweighed by migraines, which will in turn be outweighed by mild headaches, which will in turn be outweighed by a life.

---

22. For more general explorations of this specific result, see: Kartsen Klint Jensen, "Millian Superiorities and the Repugnant Conclusion," *Utilitas* 20, no. 3 (2008): 279–300; Arrhenius and Rabinowicz, "Value Superiority."

Instead, there will be at least one option with maximal value. In this case, it is **Save Life**, since the value of one life (at point **A**) is greater than either curing 10^5 headaches (point **C**) or 10^3 migraines (point **B**). The Bounded Approach thus preserves lexical priority while also ensuring a consistent and complete evaluation of the moral claims in question. Unlike Voorhoeve's approach, it allows us to view all available actions at once, assigning them values that are appropriate to the context. Overall, this seems to be a promising approach for lexical priority theories to take to avoid the Spectrum Problem.[23]

Perhaps the main challenge for this approach is that it misrepresents the spirit, if not the verdicts, of Aggregate Relevant Claims. After all, Aggregate Relevant Claims is intended to be an alternative to standard value-based approaches to moral decision-making. It is part and parcel of upholding the Separateness of Persons that we do not view individuals as mere receptacles of value, since that would make them morally exchangeable with each other in a way that individuals are not. Thus, even if the Bounded Approach successfully models the verdicts of *Aggregate Relevant Claims*, it misrepresents *why* these verdicts are made.[24]

A related problem is that the Bounded Approach is merely an ersatz representation, in that it simply offers an otiose representation of a conclusion that Aggregate Relevant Claims has already reached. This is because the Bounded Approach seems to obscure the *mechanism* by which claims are deemed relevant or irrelevant: unanimous agreement between individual claimants as to the relevance of their respective claims to each others'. This mechanism explains why, in the Bounded Approach, particular claims cannot have greater value than particular other claims. Yet this mechanism is entirely lacking from the representation.

In response to the first objection, it is worth emphasising that the values at play in these cases are merely numerical representations of how a moral theory prioritises particular considerations over others. They are a handy index for learning, at a glance, the relative importance of moral considerations.

---

23. For a general defence of context-variant value, see: Amartya Sen, "Internal Consistency of Choice," *Econometrica* 61, no. 3 (1993): 495–521.
24. Thanks to Seth Lazar for pressing these objections.

It might seem that this response only bolsters the objection that the representation is ersatz. However, this is a mistake. We have seen that the Bounded Approach is consistent with orthodox expected value theory, and hence is able to give acceptable guidance in risky situations.

Of course, the question then is whether the decision-theoretic representation is merely a numerical representation of a preset ordering of prospects. If so, then an expected value approach that is based on the Bounded Approach would be doubly redundant, since (1) it only offers a representation of what we already knew about the ranking of outcomes involving these considerations and (2) it only offers a representation of what we already knew about the ranking of prospects involving those considerations.

However, there are other ways of determining the value of prospects. Whereas the traditional approach to expected value theory takes value to be secondary (as a representation of a prior ranking of prospects), there are other approaches take the evaluation of outcomes as primary, using that evaluation to derive a ranking of prospects.[25] For instance, in Figure 3.3, curing 10^5 headaches is roughly half as morally important as saving one life. This evaluation entails the following ranking of prospects: a $50 - 50$ chance of saving a life is roughly equal to a certainty of curing 10^5 headaches. Whether or not this particular evaluation is correct is irrelevant; the main point is that, by taking the evaluation of outcomes as primary, the Bounded Approach can represent Aggregate Relevant Claims in a way that can help *guide* us through risky situations.[26]

3.5   Conclusion

Lexical priority theories can use orthodox expected value theory to avoid the Problem of Risk. However, to do so, they must posit value functions that seem, at first pass, mysterious at best and implausible at worst. I have argued that Voorhoeve's Aggregate Relevant Claims approach to aggregation can help the

---

25. On a classic sketch of this approach, see: D. Ellsberg, "Classic and Current Notions of "Measurable Utility"," *The Economic Journal* 64, no. 255 (1954): 528–556. For a contemporary approach, see: Kenny Easwaran, "Decision Theory without Representation Theorems," *Philosophers' Imprint* 14, no. 27 (2014): 1–30.

26. For a competing view, see: Lazar, "Limited Aggregation and Risk."

Bounded Approach to avoid the Agglomeration Problems, giving a principled and satisfactory explanation for why lower-considerations diminish in value in the presence of higher ones, and for when we should treat our actions as being separable or inseparable. I then argued that Voorhoeve's approach, in turn, can use the Bounded Approach to respond to the Spectrum Problem. Although more work is required to develop this approach, it seems that lexical priority theories have at least two principled and defensible decision-theoretic models for guiding us through risky cases.

Chapter 4

Duty and Ignorance

4.1 Introduction

Holly Smith (2014) argues that deontological moral theories cannot correctly determine whether we ought to gather more information before acting.[1] For brevity, I will call this: the Problem of Ignorance. To illustrate, consider the following case:[2]

Diagnostics: You are a doctor in the emergency ward, faced with a patient who has a life-threatening condition. Your primary duty is to cure your patient. You also have a duty, albeit a less weighty one, to reduce her pain. You have the following choice:

Administer Drug: You can give the patient Drug A, which is a potential cure, or Drug B, which is merely a pain-reliever. If the patient is A-Receptive, then it will both cure her and relieve her pain. If the patient is not A-Receptive, then it will increase her pain and not cure her. In either case, Drug B will not cure the patient, but it will provide pain relief.

---

1. Smith, "The Subjective Moral Duty to Inform Oneself before Acting."

2. This case is intended to be structurally identical to Smith's lead case, which involves determining who a manager should lay-off. Here, I am instead using a medical case, similar to Jackson (1991), because its moral dimensions are arguably clearer. See: Jackson, "Decision-Theoretic Consequentialism."

Alternatively, you can first:

Test, then Administer Drug: You can run a diagnostic test that will correctly determine whether the patient is receptive to Drug A (giving an output: A-Positive) or whether she is not so (giving an output: A-Negative). Then you can **Administer Drug**.

Running the test will have no cost (moral or otherwise): it will not cause the patient undue harm and will not jeopardise her or anyone's safety. As it stands, you do not have any beliefs about whether the patient is A-Receptive.[3]

We will assume that you morally ought to run the test: accepting free and reliable information that allows you to better treat your patient is something that any adequate moral theory should require. The challenge for subjective deontological theories – those which hold that what we ought to do is sensitive to our beliefs about the facts of our situation – is to give a principled and non-problematic explanation for why they agree with this verdict. Smith argues that no such explanation is available to them.

Against this, I shall argue that deontological theories do not necessarily face the Problem of Ignorance. In general, deontologists can use a decision-theoretic approach to evaluating the importance of information, even in Smith's target cases in which we have no or mistaken beliefs about our situation.[4]

Part 4.2 presents Smith's argument against deontology. **Part 4.3** argues that deontologists should reject Smith's analysis of their position in favour of a decision-theoretic approach to information gathering. **Part 4.4** compares this approach to the relatively piecemeal approach given by Philip Swenson (2016).[5]

---

3. As I discuss below (in Part 4.3.1), this assumption about 'lacking beliefs' plays a crucial role in Smith's argument.

4. Like Smith, I will focus primarily on cases involving no beliefs. The approach I give here also applies to cases involving mistaken beliefs.

5. Philip Swenson, "Subjective Deontology and the Duty to Gather Information," *Ethics* 125, no. October (2016): 257–271.

4.2   The Problem of Ignorance

A distinctive and admirable feature of Smith and Swenson's respective discussions of the Problem of Ignorance is that both theorists construct original arguments for their positions, based on novel moral principles that are independently plausible. Notably, however, neither theorist engages with the well-established literature in decision theory on the topic of when an individual ought to gather more information.[6] Instead, they draw conclusions about whether deontology can be adequately extended to cases of incomplete information using piecemeal extensions of existing moral theory. Below, I outline their respective arguments. I argue that their piecemeal nature leads to an unduly narrow – and, in fact, mistaken – approach to determining whether deontologists face the Problem of Ignorance.

4.2.1   Smith's Argument

Smith's contention is that if you are a deontologist then you cannot, by your own lights, have a moral duty to gather more information before acting. If true, then this is a startling conclusion. Given that the critique appears to target all deontological moral theories, it would effectively refute the entire deontological tradition. How does Smith argue for this?

Smith adopts a simple but powerful argumentative strategy: she distinguishes between two jointly exhaustive types of duties that deontologists might posit, and then proceeds to argue that neither type allows deontology to avoid the Problem of Ignorance.

The first type of duty is a *free-standing* one that holds that for *all* cases of uncertainty, you should gather information. Smith rejects any approach of this type:

---

6. The locus classicus of value of information calculations is found in: I. J. Good, "On the Principle of Total Evidence," *The British Journal for the Philosophy of Science* 17, no. 4 (1967): 319–321. For an accessible introduction, see: Michael D. Resnik, *Choices: An Introduction to Decision Theory* (London: University of Minnesota Press, 1987), pp. 57-59. For novel applications in ethics, see: Jackson, "Decision-Theoretic Consequentialism"; Frank Jackson, "How Decision Theory Illuminates Assignments of Moral Responsibility," chap. 2 in *Intention in Law and Philosophy*, ed. Ngaire Naffine, Rosemary J Owens, and John Williams (Aldershot, UK: Ashgate, 2001), 19–36.

> Scrutiny of these proposed freestanding duties to seek information reveals that they have several flaws: either they provide poor advice, or they provide insufficiently detailed advice, or they provide advice for which no rationale is forthcoming.[7]

Simply put, any plausible moral theory should be able to distinguish between cases where we should gather information (and to what extent) from other cases where we should not or need not. Since a freestanding approach is (by definition, it seems) insensitive to context, it is difficult to see how it could correctly balance the various considerations of cost, reliability of information, and so on, which are relevant to determining whether we morally ought to gather more information.[8] On this basis, Smith concludes that deontologists cannot avoid the Problem of Ignorance by positing a freestanding duty to gather more information.

If this is correct, then it seems that a deontological theory must instead adopt a *derivative* duty, one that is grounded on the fact that gathering information can help you to discharge other duties you have. Unlike with freestanding duties, it is easier to see how derivative duties will be sensitive to the moral stakes of your situation: if gathering information will help you to better discharge your other duties, then you have duty to do so; if it won't, then you don't.

However, Smith argues that, due to the particular commitments of deontology, any such duty will also incorrectly determine whether to gather more information. This conclusion is said to follow from the following principles, which I shall call **Subjective Duty** and **No Deontic Value**.[9] I will briefly explain how these premises of Smith's argument help to establish her main result: a case where deontology incorrectly permits us to avoid gathering information.

---

7. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," p. 19.

8. For a cogent response to this charge, see: Tenenbaum (2017), who holds that we have a defeasible duty to gather information until we know the relevant facts. This requirement is defeasible in that it does not apply in all cases, e.g. those in which gathering more information is not accessible or too costly to access. Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model," p. 20.

9. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," pp. 32-34.

Premise 1: Subjective Duty

Smith offers the following general principle that deontologists might accept when determining whether or not we ought to gather more information:

Subjective Duty: You have a duty to gather information if and only if you believe that doing so would lead you to produce the maximum amount of deontic value.[10]

I shall later argue that deontologists should reject **Subjective Duty**. For now, however, we will see how it plays an important role in Smith's argument. Applied to **Diagnostics**, **Subjective Duty** appears to entail the following moral code:

Code C

Clause 1. When you believe that you should cure and not-harm the patient, you should:

  a. Give Drug A, if you believe doing so will cure the patient, or;

  b. Give Drug B, if you do not believe that Drug A will cure the patient.

Clause 2. When you believe that you can gather more information relevant to treating a patient, then you should do so if and only if you believe that doing so would lead you to produce the maximum amount of deontic value.

Smith concludes, however, that **Code C** incorrectly leads you to avoid gathering information once it is coupled with the following assumption: **No Deontic Value**.

---

10. This is a simplified version of Smith's principle. For the purposes of the argument given here, nothing of substance is lost by the simplification. Compare: "An agent has a subjective derivative prima facie duty to do what he believes is acquiring information if and only if he believes that doing what he believes is gathering information would leave him subsequently to produce the maximum amount of deontic value (typically through his doing what he then believes to be carrying out the various deontic duties that would then be incumbent upon him)." ibid., p. 24.

Premise 2: No Deontic Value

Smith argues that some actions can *create* a duty that would not otherwise exist. For example, making a promise can create a *pro tanto* duty to perform the promised action, where otherwise this duty would not exist. Likewise, she argues, in cases like **Diagnostics**, your action of investigating or not investigating will create a duty (by Clause 1 of **Code C**) to perform a particular treatment by leading you to believe whether the patient is or is not receptive to Drug A.

Drawing on earlier work, Smith contends that deontology must hold that there is no deontic value to creating a duty, or else we would be required to create increasingly onerous duties: rather than promising to look after someone's goldfish, we would be required to promise to look after their children; we may even be required to damage others' property in order to create the weighty obligation to pay compensation; and so on.[11] Since we are not required to make such promises or, more generally, to create more burdensome duties, Smith concludes that deontologists must accept:

No Deontic Value: There is no positive deontic value to satisfying a created duty, but only a negative value to violating such a duty.[12]

Applied to the present case, Smith argues that since gathering information can create duties (once you become aware of a fact, you can come under a duty to respond to it appropriately), there will be no additional deontic value associated with gathering it as opposed to not doing so.

It is worth noting that deontologists have numerous resources for denying **No Deontic Value**, if they wish. For instance, they might argue that gathering information does not create duties in the relevant sense: rather, gathering information involves *discovering* what your duties actually are.[13] On top of this, deontologists can point out that they can take into account personal costs

---

11. Holly M. Smith, "A Paradox of Promising," *Philosophical Review* 106, no. 2 (1997): 153–196; Smith, "The Subjective Moral Duty to Inform Oneself before Acting"

12. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," p. 29.

13. See: Peter A. Graham, "In Defense of Objectivism about Moral Obligation," *Ethics* 121, no. 1 (2010): p. 91.

Option	*A-Receptive*	*Not A-Receptive*
Drug A	0	-1
Drug B	-1	0

Table 4.1: **No Deontic Value**, applied to **Diagnostics**.

in determining whether one has an obligation to do what is otherwise best.[14] They can also deny that we should create obligations to compensate others (say, by harming them and becoming duty-bound to compensate), since we have reasons to refrain from harming others, even if our compensating them would bring them greater benefits overall.[15] More generally, there is no reason to think that deontologists hold that the more onerous a moral duty, the more valuable it is. However, it turns out that these lines of argument are unnecessary: as we shall see, **No Deontic Value** is a red-herring that does nothing to block deontology from adopting a decision-theoretic solution to the Problem of Ignorance.

As such, for the sake of argument, we will accept **No Deontic Value** and assign all outcomes in which you fulfil a created duty a value of zero.[16] Without loss of generality, Table 4.1 lists the values for the possible outcomes of your choices in **Diagnostics**.

Conclusion

Smith argues that **Subjective Duty** and **No Deontic Value** jointly entail that you will not be required to gather more information, since gathering information will not lead you to bring about the (uniquely) maximal amount of deontic value, making it morally permissible but not obligatory. As shown in Figure 4.1, whatever you choose, getting more information does not allow you to achieve more deontic value than not getting information. As such, it is

---

14. See: Lazar, "Deontological Decision Theory and Agent-Centered Options."

15. See: Seana Valentine Shiffrin, "Wrongful Life, Procreative Responsibility, and the Significance of Harm," *Legal Theory* 5, no. 2 (1999): 117–148; Thomson, *The Realm of Rights*, pp. 228-48.

16. It is worth emphasising, however, that talk of 'zero-value' is misleading. As Smith herself notes (note 28, p. 29), given that we are using interval scales, the zero point is arbitrary, since the values can be rescaled up to positive affine transformation without distorting the ordering of the outcomes.

Don't Test

Test

A-Positive

A-Negative

0

1

2

3

	A-Receptive	Not A-Receptive
Give Drug A	0	−1
Give Drug B	−1	0

	A-Receptive	Not A-Receptive
Give Drug A	0	−1
Give Drug B	−1	0

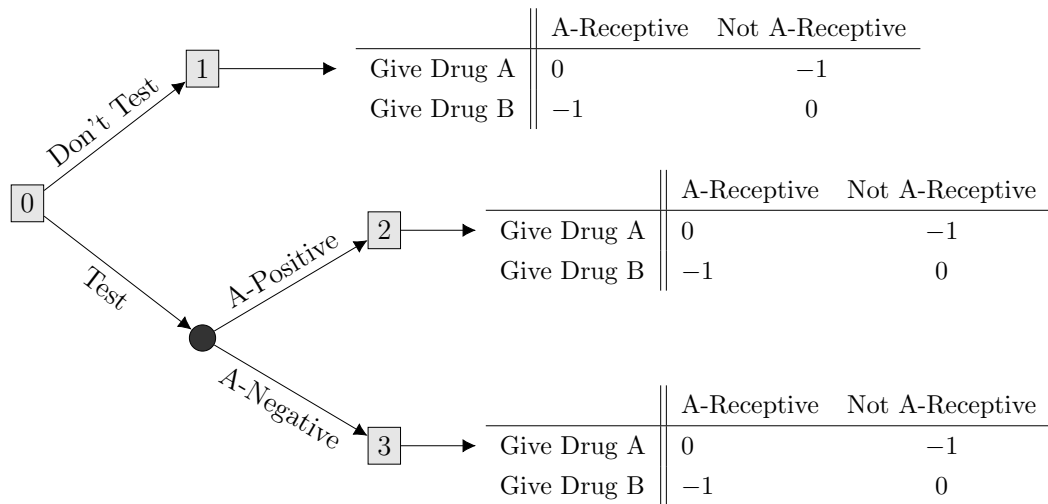|	A-Receptive	Not A-Receptive
Give Drug A	0	−1
Give Drug B	−1	0

Figure 4.1: Smith's Argument, applied to **Diagnostics**

permissible for you to choose either option.

More generally, the argument against a derivative duty to gather information can summarised as follows:[17]

Premise 1: You ought to gather more information before acting if and only if you believe that doing so will produce uniquely maximal deontic value. (**Subjective Duty**)

Premise 2: The deontic value of gathering more information is never greater than that of not gathering information. (from **No Deontic Value**)

Conclusion: Therefore, it is never the case that you ought to gather more information before acting.

Thus, having considered both freestanding and derivative approaches to the Problem, Smith concludes that there is no subjective deontological duty that can save deontology from the Problem of Ignorance.

---

17. See also: Smith, "The Subjective Moral Duty to Inform Oneself before Acting," pp. 31-32.

4.3   A Decision-Theoretic Approach

Subjective Duty determines what you ought to do on the basis of your full beliefs, rather than on your degrees of belief. I will now argue that deontologists should reject **Subjective Duty** because – even in the specific cases that Smith is targeting – you ought to gather more information because of your degrees of belief, even though you lack all-out belief about whether doing so will maximise deontic value.

4.3.1   Smith's Argument for Subjective Duty

To motivate **Subjective Duty** and perhaps to preclude a decision-theoretic approach to the Problem, Smith makes the following caveat:

> First, I will focus primarily on agents who have false beliefs or no rel-
> evant beliefs, rather than agents who are merely uncertain about their
> prospective action's character. This restriction will not distort the in-
> quiry or its outcome. If a moral theory flunks the test of appropriately
> handling the duty to gather information for agents having false beliefs
> or no relevant beliefs, then it is inadequate, even if it passes this test
> for agents laboring under uncertainty.

This restriction determines the content of **Subjective Duty**, which defines duties in terms of beliefs or lack of belief, rather than in terms of uncertainty. Recall:

Subjective Duty: You have a duty to gather information if and only if you believe that doing so would lead you to subsequently produce the maximum amount of deontic value.

As we saw, this restriction plays a crucial role in Smith's argument. In **Diagnostics**, since (by stipulation) you do not believe that running the test will lead you to produce the maximum amount of deontic value (due to the **No Deontic Value** assumption), it is not the case that you ought to do it.

Unfortunately, Smith never explains what she means by 'belief' or what it means to have 'no relevant beliefs'. More to the point, she never explains why your having no beliefs or false beliefs about a proposition entails that

you cannot be more, less or similarly confident that the proposition is true as opposed to its negation. As it stands, there is no apparent inconsistency between not believing that the patient is A-Receptive, not believing that she is not A-Receptive, while nevertheless being more, less, equally, or similarly confident that she is receptive to Drug A as opposed to not receptive to it. This suggests that Smith's target cases – those involving no or mistaken beliefs – and cases involving uncertainty are not mutually exclusive. Indeed, I shall argue that her target cases are a proper subset of cases of uncertainty. If that is correct, deontologists can hold, for familiar decision-theoretic reasons, that your degrees of confidence can make it the case that you should gather information *even if* you do not believe that doing so will maximise deontic value (thus rejecting the 'only if' clause in **Subjective Duty**).

4.3.2 A Deontic Value of Information Approach

In presenting a positive proposal of how deontology can avoid the Problem of Ignorance, I will introduce some terminology that helps to bridge deontological moral theory and decision theory.

For the moment, I will try (where possible) to avoid the term 'duty'. This is because duty-speak can be unclear about how exactly duties are relevant to decision-making in risky situations. For instance, are duties the *output* of a decision process (e.g. your duties are just the set of permissible actions)? Or are duties the *inputs* of a decision process (e.g. you consider your *pro tanto* duties in your deliberations regarding which of your available actions are morally permissible)? Or are duties some *function* from the grounds of obligation to some set of permissible actions (whereby your duty is to consider your reasons for action and to determine, from them, what you morally ought to do)?

In an attempt to side-step this ambiguity, I will adopt the choice-theoretic framework presented in Chapter 1, Section 1.4.1.[18] Recall that a *choice context*

---

18. For a general framework for modelling moral theories, one that allows for agent-relative perogatives, menu-dependence, and other structural features commonly associated with deontology, see: Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

is a set of actions, a set of possible states of the world, and a belief function defined over those states that measures how confident you are that a particular state is actually the way the world is. When an action is performed under a particular state of the world, it produces an *outcome.* This outcome encodes whether moral considerations have been upheld or violated: it may represent the nature of the action (for instance, it may represent the fact that you intentionally lied) and it may also represent the causal consequences of your action (for instance, it may represent whether or not your interlocutor believed your lie). An action's *prospect* is the set of its possible outcomes and their respective probabilities of obtaining. We will assume that these prospects can be ordered and represented by a *deontic value function*, which assigns a higher deontic value to outcomes, the more morally important they are. A *moral decision rule* is a function from a choice context to a set of morally permissible actions. It draws on some or all of the above information to determine what, if anything, is morally permissible.

Understood in this framework, the challenge for deontologists is to find an unproblematic moral decision rule that is consistent with their core tenets and which allows them to avoid the Problem of Ignorance. Smith's claim is that no such moral decision rule exists or – at the very least – coming up with such a rule would involve 'dauntingly complex theoretical issues.'[19]

For example, one reason why Smith denies that deontologists can rely on degrees of belief is that, in cases like the one she is concerned with, modelling an agent's lack of belief in terms of degrees of uncertainty may involve relying on the Principle of Indifference, which is a problematic epistemic rule.[20] Roughly speaking, the Principle of Indifference says that when we lack any evidence about our situation, we should assign equal probabilities to the various possibilities that might occur. If you do not know whether the patient is A-Receptive or not, then you should have a 0.5 credence in either possibility. It is well known, however, that the Principle of Indifference faces the problem that possibilities can be redescribed in seemingly equally eligible ways to

---

19. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," p. 33.

20. This principle was so-named by John Maynard Keynes, *A Treatise of Probability* (London; New York: Macmillan; AMS), ch. 4. For Smith's discussion, see: Smith, "The Subjective Moral Duty to Inform Oneself before Acting," p. 33.

yield different numbers of possibilities, leading to a conflicting assignment of probabilities.[21]

In response, however, note that even if the Principle of Indifference is false (which is debatable), deontologists do not need to invoke it; they can use other epistemological tools.[22] Indeed, orthodox Bayesian epistemology holds that, if we are rational, we always have precise and coherent degrees of belief even if we do not use the Principle of Indifference. The mere fact that we lack all-out belief – and that the Principle of Indifference is controversial – does not entail that we do not have well-defined degrees of belief.

Perhaps, however, Smith's target cases are not simply those in which we lack all-out belief; they are cases where we lack all-out belief because we lack evidence. It is arguable – though still a matter of significant debate – that lacking evidence prevents us from having precise degrees of belief.[23]

However, even if such cases present problems for orthodox Bayesianism, it turns out that there are well-established, alternative Bayesian approaches to dealing with them. Specifically, deontologists can represent themselves as having an indeterminate credal state that reflects their equivocal, incomplete, or otherwise ambiguous evidence.[24] Where a determinate degree of belief specifies a specific numerical value (indeed, precise up to infinite decimal places), an indeterminate degree of belief admits of a range of such values, modelled by a set of probability functions. By positing sets of probability functions, deontologists can model the cases that Smith targets, without appealing to

---

21. For a classic discussion of this issue, see: Bas van Fraassen, *Laws and Symmetries* (Oxford: Clarendon Press, 1989), ch. 12.

22. For a recent defence of the Principle of Indifference and summary of other approaches, see: Richard Pettigrew, "Accuracy, Risk, and the Principle of Indifference," *Philosophy and Phenomenological Research* 92, no. 1 (2016): 35–59.

23. For a defence of indeterminate credences, more generally, see: James M. Joyce, "A Defense of Imprecise Credences in Inference and Decision Making," *Philosophical Perspectives* 24 (2010): 281–323; Alan Hájek and Michael Smithson, "Rationality and Indeterminate Probabilities," *Synthese* 187 (2012): 33–48; Isaac Levi, "Why Indeterminate Probability is Rational," *Journal of Applied Logic* 7, no. 4 (December 2009): 364–376. For an argument in favour of the orthodox view, see: Adam Elga, "Subjective Probabilities should be Sharp," *Philosophers' Imprint* 10, no. 5 (2010): 1–11.

24. James M. Joyce, "How Probabilities Reflect Evidence," *Philosophical Perspectives* 19 (2005).

the Principle of Indifference.[25]

Adopting this approach in **Diagnostics**, we will model your epistemic situation of 'lacking beliefs' in terms of indeterminate degrees of belief. In this case, let's suppose that you are roughly equally confident that the patient is A-Receptive rather than not so, such that the probabilities range within the interval [0.4, 0.6] (or ≈ 0.5, for ease of exposition).[26]

With this prior, albeit indeterminate, degree of belief, running the test will allow you to update your degrees of belief and choose from a more informed standpoint. Let us suppose that the test is known to be 99% accurate. Given this information, being a rational agent you will update your degrees of belief to reflect the evidence. As we shall see in **Diagnostics**, this information will improve the prospects associated with giving one drug rather than the other. Following this approach, deontologists should reject **Subjective Duty** in favour of the following moral decision rule:

Choose Undefeated Prospects: An action is morally permissible if and only if its prospect is undefeated by that of any alternative action.

An 'undefeated prospect' is one that is not ranked lower than any other prospect. For current purposes, we do not need to fully define the properties that govern whether one prospect is ranked higher than another; we can simply hold that, all else equal, a treatment with a higher probability of better outcomes defeats a treatment with a lower probability of better outcomes.[27]

---

25. For defences of this approach, see: Isaac Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance* (Cambridge, Massachusetts; London, England: MIT Press, 1980); Jeffrey, *The Logic of Decision*; Mark Kaplan, *Decision Theory as Philosophy* (Cambridge: Cambridge University Press, 1996); Joyce, "How Probabilities Reflect Evidence"; Joyce, "A Defense of Imprecise Credences in Inference and Decision Making."

26. Note, however, that we can make the interval much wider: indeed, potentially maximally so – representing, perhaps, your complete and utter incomprehension of your situation. New evidence will have much less of an effect on your credences in such cases, but may still improve the prospects of your treatment. As such, it may be morally required. For discussion, see: Joyce, "A Defense of Imprecise Credences in Inference and Decision Making"; Aron Vallinder, "Imprecise Bayesianism and Global Belief Inertia," *British Journal for the Philosophy of Science* 0, no. June (2018): 1–26.

27. As mentioned in Chapter 1, a defeat relation is weaker than a weighing relation. Unlike a weighing relation, a defeat relation can allow for incommensurable considerations, strong

This can be conveniently represented in terms of one action having a strictly higher probability-weighted average (or 'expected') deontic value (EDV) than another.[28]

Choose Undefeated Prospects solves the Problem of Ignorance because it allows deontologists to use what is equivalent to a 'value of information' calculation to determine when it is morally required to gather more information. Roughly put, value of information calculations involve comparing your optimal future actions in an informed context (which results from having gathered more information) versus an uninformed one. If the informed context has an optimal action whose prospect defeats that of every action in the uninformed context, then you should choose to gather more information.[29]

For example, in **Diagnostics**, you have a choice between acting in an informed choice context (by running the test) or in an uninformed choice context (by not running the test). The rule **Choose Undefeated Prospects** determines whether or not you should run the test by calculating whether the optimal action in the informed choice context has higher expected deontic value than the optimal action in the uninformed choice context. Figure 4.2 illustrates your choice situation based on your prior and posterior degrees of belief, giving the expected deontic values of your available courses of action.

If you choose to remain uninformed, the morally optimal action of your later choice situation has an expected deontic value of between -0.4 and -0.6 (or ≈ -0.5 for short, as in the diagram above). Compare this with the context

moral dilemmas, (arguably) moral options, and even intransitivity. See: Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories," p. 432.

28. Note that this expected value representation can be weakened in several respects, notably by dropping the Completeness and Continuity axioms, allowing for incommensurable considerations and a strong version of lexical priority, respectively: see Robert J. Aumann, "Utility Theory without the Completeness Axiom," *Econometrica* 30, no. 3 (1962): 445–462; Hausner, "Multidimensional Utilities." Indeed, my argument can be run without expected value altogether, instead using a barer choice-theoretic approach based on a defeat relation, as described in: Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

29. For an accessible introduction, see: Resnik (1987). For details on how such calculations work with indeterminate probabilities (as in our current case), see: Bradley and Steele (2016). Resnik, *Choices: An Introduction to Decision Theory*; Seamus Bradley and Katie Steele, "Can Free Evidence Be Bad? Value of Information for the Imprecise Probabilist," *Philosophy of Science* 83, no. 1 (2016): 1–28.

Give Drug A   0

Node 0 → Don't Test → 1

| | $p \approx 0.5$ | $(1-p) \approx 0.5$ | |
	A-Receptive	Not A-Receptive	EDV
Give Drug A	0	−1	≈ -0.5
Give Drug B	−1	0	≈ -0.5

Test → A-Positive → 2

| | $p \approx 0.99$ | $(1-p) \approx 0.01$ | |
	A-Receptive	Not A-Receptive	EDV
Give Drug A	0	−1	≈ -0.01
Give Drug B	−1	0	≈ -0.99

A-Negative → 3

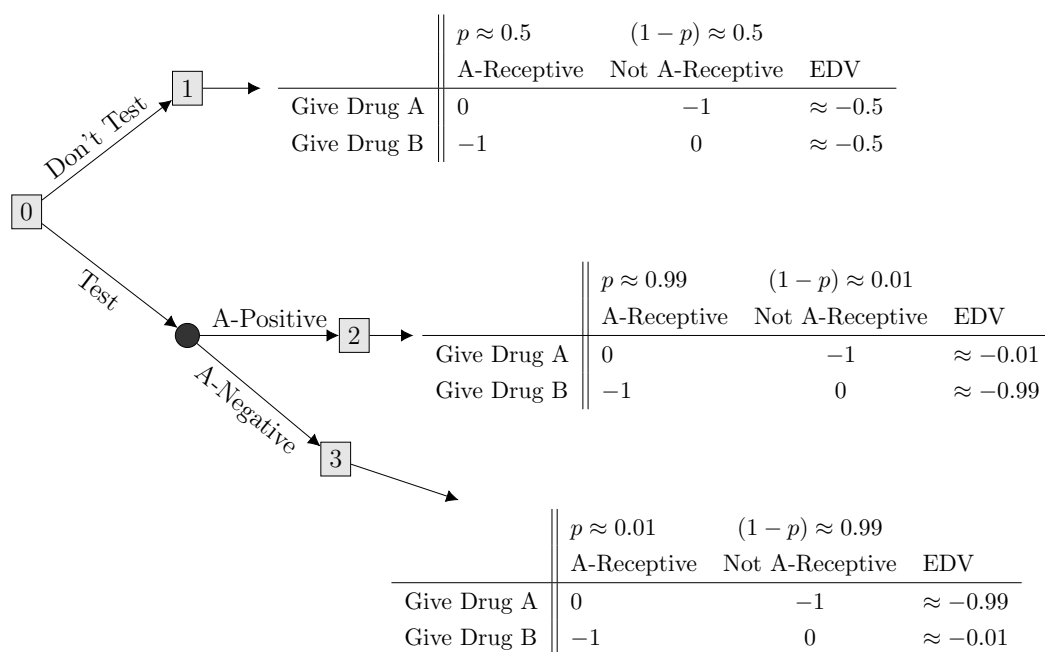| | $p \approx 0.01$ | $(1-p) \approx 0.99$ | |
	A-Receptive	Not A-Receptive	EDV
Give Drug A	0	−1	≈ -0.99
Give Drug B	−1	0	≈ -0.01

Figure 4.2: Value of Information Calculation

in which you run the diagnostic test. If you Test, then the optimal actions have an expected deontic value of approximately −0.01, which is greater than that of the optimal actions in the uninformed context.[30] According to **Choose Undefeated Prospects**, you are morally required to run the diagnostic test, even though you do not all-out believe that doing so will maximise actual (as opposed to expected) deontic value. This result shows that there is a subjective duty (or moral decision rule) that deontologists can adopt to avoid the Problem of Ignorance.[31] The question, then, is whether deontologists can avail themselves of this solution without giving up their core substantive commitments.

---

30. Strictly speaking, this will be an interval of values around the value 0.1.

31. Note also that this approach can be extended to account for cases of multiple conflicting duties, costly or not fully reliable information, or uncertainty about our future actions.

4.3.3 Objections

Perhaps anticipating this approach, Smith offers some lines of resistance to any decision-theoretic model of deontology. For instance, she mentions that deontologists are not obviously committed to the idea that deontic value is representable by a linear value function.[32] Some deontologists agree.[33] As Sergio Tenenbaum puts it:

> There is no obvious measurable value that the deontologist is trying to bring about to the greatest degree. So when a deontologist multiplies a probability by a value, it is far from clear what this product represents. If I am bound by a rule that prohibits lying because lying expresses disrespect, does raising the probability of lying always express disrespect (rather than at most incurring a risk of expressing disrespect)? And if risking disrespect is a form of disrespect, is it disrespectful in proportion to the risk?[34]

There are a few ways of understanding this concern. On one interpretation, the concern is that deontological moral considerations are not ordered in the right way to allow for a well-defined measure of deontic value. Bracketing the fact that Smith's argument presumes that deontological considerations *can* be represented by a deontic value function, there are reasons to think that deontology violates at least some of the axioms of orthodox expected value theory.[35] However, this requires further investigation, not least because – as have seen in earlier chapters – decision theorists have made many extensions and generalisations of orthodox decision theory that allow it to accommodate risk attitudes, incommensurability, and lexical priority – all while maintaining a systematic and coherent approach to risk.[36] It is therefore not enough to show

---

32. Smith, "The Subjective Moral Duty to Inform Oneself before Acting," p. 33.

33. See, for example: Aboodi, Borer, and Enoch, "Deontology, Individualism," p. 272.

34. Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model," p. 9

35. For arguments to this effect, see: Frances M. Kamm, "Supererogation and Obligation," *Journal of Philosophy* 82, no. 3 (1985): 118–138; Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, ch. 7; Lazar, "Deontological Decision Theory and Agent-Centered Options"; Lazar, "Limited Aggregation and Risk"; Tenenbaum, "Action, Deontology, and Risk: Against the Multiplicative Model."

36. John W Pratt, "Risk Aversion in the Small and in the Large," chap. Chapter 19 in *Handbook of the Fundamentals of Financial Decision Making* (1964), 317–331; Aumann, "Utility Theory without the Completeness Axiom"; Hausner, "Multidimensional Utilities."

that deontology cannot be represented by *some* particular version of expected value theory; those pressing this approach need to show that deontology cannot be represented by *any* version of expected value theory. We are yet to see an impossibility result to this effect.

A weaker interpretation of the concern attends to the question of whether deontologists should be *risk neutral*: that is, whether they should consider two prospects that have same expected deontic value to be equally choice-worthy, even if one prospect has much greater variance (or greater 'spread') in possible deontic value than the other. As it turns out, however, non-neutral attitudes to risk are consistent with orthodox decision theory.[37] Although deontologists are yet to settle such questions as whether non-neutral risk attitudes are morally appropriate, the solution given above is compatible with various answers to the question.[38]

For their part, deontologists might worry that this proposal is too consequentialist: it determines the deontic value of gathering information based on the deontic value your later choices, which in this case amount to correctly or not correctly treating the patient, based on your evidence. To anyone with deontological inclinations, it seems that your choosing to gather more information is based on *more* than just achieving the best consequences for your patient – there is something *in itself* morally important about your efforts. For instance, gathering more information can be a way of showing concern and respect for your patient. The objection goes: if the solution I've presented here does not accommodate that kind of non-consequentialist consideration,

---

See also: Edward Elliott, "A Representation Theorem for Frequently Irrational Agents," *Journal of Philosophical Logic* 46, no. 5 (2017): 467–506; Dietrich and List, "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories."

37. Pratt, "Risk Aversion in the Small and in the Large"; Kenneth J. Arrow, "The Theory of Risk Aversion," chap. 2 in *Aspects of the Theory of Risk Bearing* (Helsinki: Yrjo Jahnsonin Saatio, 1965); Broome, *Weighing Goods*, Ch. 4.

38. Given that **Choose Undefeated Prospects** is logically weaker than expected value theory, one alternative approach for deontologists might be to adopt Lara Buchak's Risk-Weighted Expected Utility Theory. See: Buchak, *Risk and Rationality*. One worry with this particular approach, however, it that is might lead deontologists back to a version of the Problem of Ignorance, whereby individuals are not only led to reject free, reliable information and to also choose a state-wise dominated course of action. See: R.A. Briggs, "Costs of abandoning the Sure-Thing Principle," *Canadian Journal of Philosophy* 45, nos. 5-6 (2015): 827–840.

then it is not an appropriate model of deontology.

As others have noted, this type of consideration can be included in a decision-theoretic model like the one above.[39] In Smith's terms, the duty to gather information would be based on both a free-standing *and* derivative duty. The free-standing duty can be represented by giving some additional positive deontic value to any course of action that includes gathering information out of concern and respect for others.[40] The derivative duty is represented by the difference in expected deontic value that you gain from gathering information. This approach is entirely consistent with the decision-theoretic model presented earlier, and does seem to offer a more accurate representation of deontological concerns.

Overall, while much work remains to be done in constructing a fully-fledged deontological decision theory, it appears that deontologists can adopt **Choose Undefeated Prospects** to avoid the Problem of Ignorance.

4.4 Comparison with Swenson's Account

Let us now compare the approach given above with that given by Philip Swenson (2016).[41] In response to Smith's argument, Swenson accepts **No Deontic Value**, but denies that it applies to actions that create duties by gathering more information. In **Diagnostics**, this would allow him to assign higher deontic value to gathering more information as opposed to not doing so.

What is the rationale for treating information gathering in this way? Swenson argues that deontologists should be concerned with best *approximating* deontic ideals, and that gathering information is a useful way of better achieving this. Thus, in **Diagnostics**, you should presumably run the test because this best approximates the more important deontic ideal of treating the patient correctly.

More generally, he posits a number of additional 'approximation principles'

---

39. Colyvan, Cox, and Steele, "Modelling the Moral Dimension of Decisions."

40. This could be a constant value or some function of other morally relevant considerations. For present purposes, we do not need to go into detail about this.

41. Swenson, "Subjective Deontology and the Duty to Gather Information."

that, he argues, deontologists would find appealing. For instance, he posits a *Subjective Approximation Principle for Promise Keeping*, a *Subjective Approximation Principle for Justice in Layoffs*, and so forth. Each of these rules roughly holds that we should perform an action that best approximates (in some undefined sense) the particular deontic ideals at play.

Thus, on Swenson's account, deontic value is not only a measure of the significance of a deontic ideal; it is also a measure of how well an action approximates some set of deontic ideals. This is given by his Deontic Value Principles:

First Deontic Value Principle: All things being equal, Option A has higher deontic value than Option B if A approximates some deontic ideal to a greater degree than B.

Second Deontic Value Principle: All things being equal, Option A has higher deontic value than Option B if A achieves some deontic ideal X and X is (in the relevant context) a more significant ideal than any deontic ideal achieved by B (where achieving a deontic ideal counts as maximally approximating it).

More generally, Swenson's strategy proposes that deontologists should respond to the Problem of Ignorance by positing additional principles on a case-by-case basis. As he puts it:

> When Subjective Deontologists are confronted with cases in which it is intuitive that the agent has a duty to gather information, they should look for principles that allow them to assign deontic value in a manner that accounts for this duty. Given that the agent has the duty, the failure to gather information will be regrettable. This reveals that the deontic ideal has not been achieved. So the very fact that the agent has the duty is evidence that the deontic values should be assigned in a way that accounts for the duty. Thus it is highly plausible that there will always be (perhaps very complex) principles of deontic value that can do the necessary work.

Swenson is optimistic that deontology will be able to construct a general and consistent set of principles to cover all cases involving incomplete information.

107

However, I am less confident about this approach. This is because, as it stands, Swenson's proposal is underspecified in a number of important respects. For instance, we do not yet know why the regret heuristic will induce a well-defined measure of deontic value.[42] Nor do we know what it means for an action to *approximate* an ideal, and why this notion would support a well-defined measure of deontic value.

To illustrate this concern, consider the deontic ideal of correctly treating your patient.[43] It seems that there are multiple ways of approximating this ideal. Suppose that you have the following options:

Send to Dr Knowledgeable: Refer your patient to a doctor who is less caring than you (e.g. worse bedside manner), but more knowledgeable than you about the most appropriate treatment plans available.

Send to Dr Caring: Refer patient to a doctor who is more caring (e.g. better bedside manner), but less knowledgeable than you about the most appropriate treatment plans available.

Send to yourself: Treat the patient yourself, giving a normal amount of care and competence.

Whatever you choose, there will be some other option available that better approximates the deontic ideal in some respect. Suppose that you begin by not referring. Then you (rightly) become concerned that you are potentially depriving your patient of more effective treatments, so you refer the patient to Dr Knowledgeable. However, Dr Knowledgeable's horrible bedside manner

---

42. For instance, in cases where there is no fact of the matter what would have occurred, had we done otherwise, it seems that the regret heuristic might lead us to assign higher deontic value to whatever we choose *not* to do. See: Caspar Hare, "Obligation and Regret When There is No Fact of the Matter About What Would Have Happened if You Had not Done What You Did," *Noûs* 45, no. 1 (2011): 190–206.

43. The following is a similar case to those found in: Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. The best explanation for such cases seems to me to be given by Luce and Raiffa (1957), who observe that intransitive choices can arise "when a subject forces choice between inherently incomparable alternatives". See: R. Duncan Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (New York: John Wiley / Sons, Ltd., 1957). See also: John Cusbert, "Acting on Essentially Comparative Goodness," *Thought: A Journal of Philosophy* 6, no. 2 (2017): 73–83.

makes you quickly redirect the patient to Dr Caring. However, as good as Dr Caring's intentions are, his treatment plan is probably not going to be the most appropriate, so you finally send the patient back to yourself. Thus, you've led yourself (and your patient) in a circle. Each option better approximates *in some respect* the deontic ideal of correctly treating your patient. However, by that same token, each option will also be regrettable in some respect. Without further information about how the approximation relation is to be understood, it is possible that the relation may sometimes be intransitive, preventing us from assigning well-defined deontic values to options.

This is merely to illustrate that Swenson's account does not explain how its most critical aspects function as he hopes, and so it is unclear why we should expect our adding more case-by-case principles will lead to an informative and internally-consistent account. Without a clearer sense that the foundations of the account are secure, deontologists should be wary of building further complex principles on top of them.

Another problem is that it is entirely unclear how Swenson's account accommodates degrees of uncertainty. Recall that in **Diagnostics**, you are 99% certain that the test is accurate. Should you run the test? Swenson's account has no answer. Indeed, it falls silent for *any* degree of uncertainty, since its approximation and deontic value principles are formulated without any reference to degrees of uncertainty. The question is: can it be extended to cover such cases, thus providing a general solution to the Problem of Ignorance for deontologists?

Following Swenson's preferred approach of positing additional (potentially complex) principles, we might attempt to extend Swenson's account to cases of uncertainty by holding that there is a close connection between uncertainty and approximation. For example, the following would be a natural extension of his account:

Third Deontic Value Principle: All else equal, Option A better approximates a deontic ideal than Option B when A is more certain to achieve the ideal than B.

Unfortunately, this cannot be the correct approach. 'Achieving' a deontic

ideal, on Swenson's proposal, "counts as maximally approximating it".[44] So, according to the Third Deontic Value Principle, Option A better approximates a deontic ideal by virtue of being more certain to better (indeed, maximally) approximate it. This is clearly circular. Hence, there must be a different sense of approximation at play: perhaps, a subjective one. Accounting for this requires an additional principle, such as:

Subjectively Approximate Maximum Approximation: All else equal, if Option A subjectively approximates maximum approximation of a deontic ideal more than Option B, then you should choose Option A rather than Option B.

By virtue of being more certain that running the test will lead you to better approximate the deontic ideal of treating your patient, you ought to do so. Note, however, that even this is not enough. What if all else is *not* equal? What are we to say about cases where the two dimensions of deontic value (approximation and importance) disagree? For instance, what if one option has a low probability of achieving maximum approximation, compared to another option where there is a high probability of achieving lesser approximation? Until we know more about the core aspects of Swenson's account, we cannot treat it as a general solution to the Problem of Ignorance.

Of course, an easy alternative approach would keep uncertainty separate from approximation. To deal with cases of uncertainty, we might posit a moral decision rule that takes into account degrees of uncertainty and deontic value. However, it is unclear how this will not just turn out to be a heavily cloaked version of the decision-theoretic solution presented earlier.

4.5 Conclusion

The Problem of Ignorance does not apply to deontological moral theories that are compatible with expected value theory. Using value of information calculations, deontologists can determine when we morally ought to gather more

---

44. Swenson, "Subjective Deontology and the Duty to Gather Information," p. 264.

information before acting. However, to determine whether deontologists can truly avail themselves of this solution, we must further investigate the compatibility of deontological commitments with the core axioms of expected value theory. Nevertheless, as it stands, there are reasons to be optimistic that deontology can guide us through an uncertain world.

Chapter 5

Authority, Obedience, and Uncertainty

> There is but one power to which I can yield a heart-felt obedience, the decision of my own understanding, the dictate of my own conscience.

> William Godwin,
> *An Enquiry Concerning Political Justice* (1793)

5.1   Introduction

Suppose that you are serving in the military, and have just been ordered by your superior officer to commandeer a civilian's van to deliver live-saving goods to a distant village.[1] You have doubts about whether this is the best course of action. Moreover, as it happens, you are certain that if you disobey, it will go undetected. What ought you to do?

Many hold that the answer depends at least partly on whether your superior officer is a *legitimate* authority. What does it take to be a legitimate authority? According to Joseph Raz's Service Conception, the pre-eminent

---

1. This case is loosely based on: Raz, *Practical Reason and Norms*, pp. 38-45; Joseph Raz, *From Normativity to Responsibility* (Oxford; New York: Oxford University Press, 2011), p. 116.

theory of legitimate authority, an authority is legitimate just in case it reliably helps us to better act in accordance with the reasons that apply to us. This idea is more precisely captured by:

The Normal Justification Thesis: An agent is a legitimate authority if and only if obeying its orders would lead you to better conform with the reasons that independently apply to you, compared to if you acted on your own assessment of those reasons.[2]

Let us assume that in the case at hand, you have most reason to save as many lives as you can, even if doing so involves violating another's property rights. However, if you can save just as many lives without violating anyone's property rights, then you should take the less rights-infringing course of action. According to the Normal Justification Thesis, if obeying your superior's orders is a more reliable or efficient means of helping you to save more lives compared to following your own judgement, then your superior officer has legitimate authority over you.

To better understand the distinctive normative power of legitimate authority, we need to know more about the kind of obedience such authorities can demand of us. This is defined by:

The Pre-Emption Thesis: when a legitimate authority issues you a command, it thereby gives you a:

---

2. This *if and only if* formulation attempts to take into account both the Independence Thesis and the Dependence Thesis. The Independence Thesis states that: "the matters regarding which the first condition is met are such that with respect to them it is better to conform to reason than to decide for oneself, unaided by authority" Joseph Raz, "The Problem of Authority: Revisiting the Service Conception," *Minnesota Law Review* 90, no. 1979 (2006): p. 1014. Arguably, however, this is already captured in the Normal Justification Thesis, since in the relevant cases capture by the Independence Thesis are those in which you have most reason to *not* defer to others' judgements and to instead decide on the basis of your own. The Dependence Thesis states that: "All authoritative directives should be based, among other factors, on reasons which apply to subjects of those directives and which bear on the circumstances covered by the directives." (Joseph Raz, *Ethics in the Public Domain* (Oxford: Clarendon Press, 1994), p. 214) This is invoked in the above formulation of the Normal Justification Thesis through the clause that the relevant reasons are those that *independently* apply to you.

Content-Independent Reason: A reason to do as it commands *because a legitimate authority commanded it*, not because of the content of the command.[3] And an:

Exclusionary Reason: A reason to *not act upon* your own assessment of your situation.[4]

Applied to the present case, the Pre-Emption Thesis holds that if your superior officer is a legitimate authority, you should obey by not acting on your own ideas of how best to deliver the goods to the village, and also by doing as commanded simply because it was directed by a legitimate authority. Obedience in this *pre-emptive sense* is taken to be the correct response to a legitimate authority. Indeed, according to Raz:

> The pre-emptive force of [legitimate] authority is part and parcel of its nature. It cannot succeed as an authority (i.e., succeed in improving our conformity with reason) if it does not preempt the background reasons.[5]

With this theory in hand, let us return to our central case: should you take the van or not? It seems that even though you have doubts about the correctness of the order, if you are sure that your superior officer has more reliable judgement than you, then you should obey.

However, what if you are *uncertain* about whether your superior officer is more reliable than you? Should you pre-emptively obey, despite your doubts? If so, how certain do you need to be that the authority is not legitimate before you should no longer do so? As we shall see, existing answers to these questions are problematic.

Before proceeding, however, it is worth pausing to remember why we should care about how theories of legitimate authority operate in cases of uncertainty. Anyone with a passing knowledge of human history will notice that the practical stakes are high: many of humanity's greatest moral atrocities were the result of obedience to illegitimate authority. Without a clear moral standard

---

3. On this notion, see: Stefan Sciaraffa, "On Content-Independent Reasons: It's Not in the Name," *Law and Philosophy* 28, no. 3 (2008): 233–260.
4. Raz, *The Morality of Freedom*.
5. Raz, "The Problem of Authority: Revisiting the Service Conception," p. 1019.

for justifiable obedience to authorities of questionable legitimacy, individuals who execute the wrongful commands of authority can almost always assert that they did not *know* that the authority is illegitimate, and that they were "just following orders".[6]

There are also important theoretical stakes at play. In particular, many believe that the Service Conception refutes Philosophical Anarchism, a view which holds that we should never surrender our will or judgement to that of another, as doing so would violate our moral obligation to be autonomous and rational decision-makers.[7] According to the Anarchist, theories of legitimate authority are mistaken because they permit – and, in some cases, require – us to submit our will in this way.[8] To authority, and particularly to political authority, the Anarchist says:

> It is yours to shackle the body and restrain our external actions; that is a restraint we understand. Account your penalties; and we will make our election of submission or suffering. But do not seek to enslave our minds. Exhibit your force in its plainest form, for that is your province; but seek not to inveigle and mislead us. Obedience and external submission is all you are entitled to claim; you can have no right to extort our deference, and command us not to see, and disapprove of, your errors.[9]

In response, the Service Conception is taken to show that there is no necessary conflict between autonomy, rationality, and obedience to legitimate authority. Firstly, pre-emptive obedience does not require us to not assess the merits of an order or legitimacy of an authority; it merely requires that we not act upon our less reliable judgement of the situation.[10] Secondly, since authorities are

---

6. On the jurisprudential difficulties of such cases, see: Mark J. Osiel, "Obeying Orders: Atrocity, Military Discipline, and the Law of War," *California Law Review* 86, no. 5 (1998): 943.

7. See, for example: Leslie Green, *The Authority of the State* (Oxford: Clarendon Press, 1988); Shapiro, *Authority*.

8. William Godwin, *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness* (Toronto: University of Toronto Press, 1946); Robert Paul Wolff, *In Defense of Anarchism* (University of California Press, 1970).

9. Godwin, *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness*, pp. 236-237.

10. Joseph Raz, "Legitimate Authority," in *Philosophical Law*, ed. Richard Bronaugh (Westport, Conn.: Greenwood Press, 1978), 3–27; Joseph Raz, "Authority and Justification," *Philosophy and Public Affairs* 14, no. 1 (1985): 3–29.

legitimate only insofar as they promote our conformity with our reasons for action, far from being in conflict with rationality and autonomy, pre-emptive obedience to legitimate authority is the *very definition* of how rational and autonomous decision-makers respond to the commands of a legitimate authority.[11] In Leslie Green's words, the Service Conception explains how we can act *contrary to the balance of reasons* without thereby acting *contrary to reason*.[12]

As I shall argue, however, cases involving uncertainty threaten to show that the Service Conception requires irrational obedience to authority, thereby vindicating the Anarchist. I propose that decision-theoretic modelling helps the Service Conception to deal with such cases, thereby supporting it as a response to the Anarchist and as a guide to evaluating the justifiability of obedience to authorities of questionable legitimacy.

Part 5.2 discusses existing attempts to deal with uncertainty regarding an authority's legitimacy. I argue that the Service Conception can – and, indeed, should – be modelled in terms of maximising expected value. **Part 5.3** presents a decision-theoretic model of the Service Conception, giving a precise account of when pre-emptive obedience is justified, and when it is not. **Part 5.4** responds to Scott Shapiro's (2004) objections to decision-theoretic approaches to obedience to authority. **Conclusion** follows.

5.2   Uncertain Legitimacy

How should we respond to authorities when we are unsure whether they are legitimate? A natural answer is to hold that, in cases of uncertainty, we should exclude our judgement to the extent that we believe the authority is legitimate. Unfortunately, this proposal conflicts with the very motivations for, and idea of, exclusionary reasons. The motivation for exclusionary reasons is that they save us from costly or unreliable deliberations. The idea of exclusionary reasons is that they have absolute priority over the reasons they exclude.[13] If, upon

---

11. Raz, "Authority and Justification," p. 29.
12. Green, *The Authority of the State*, p. 37
13. As Raz puts it: "Exclusionary reasons always prevail, when in conflict with [the reasons they exclude]." Raz, *Practical Reason and Norms*, p. 40.

receiving an order, you were to partially exclude your judgement by weighing the excluded reasons against the exclusionary reason, then you would be undertaking the costly task of combining your potentially less reliable judgement with that of the potentially legitimate authority. In cases of uncertainty, the 'service' given by legitimate authorities would therefore be undermined.

Perhaps for this reason, Raz rejects the idea of 'partial exclusion'.[14] Instead, he appeals to the idea of *knowability*, such that if we cannot know that an authority satisfies the Normal Justification Thesis by a reasonable inquiry (roughly, one that is worth the cost, given the stakes of the situation), then the authority is not legitimate. This is because, according to Raz:

> [G]enerally speaking, the only reliable way of conforming to authority is through having a reliable belief that it is an authority, and therefore should be obeyed ... When reasonable inquiry will not reveal the case for authority, that case, if it exists at all, is unknowable. It follows that people are not subject to any authority regarding those matters.[15]

This sets out a necessary condition for determining when pre-emptive obedience is justifiable. However, it leaves a lot unsaid. For example, how reliable do our beliefs need to be? What if we are unsure whether we have reliable beliefs about the legitimacy of an authority? Or, what if we are uncertain about whether a reasonable inquiry will give us reliable beliefs about the legitimacy of an authority? Perhaps in an attempt to side-step such matters, Raz qualifies the knowability constraint by assuming that that if we can form reliable beliefs about an authority's legitimacy, then we can also know that these beliefs are reliable.[16] This amounts to applying an additional knowability constraint on the knowability constraint: an authority is legitimate only if we can know (by a reasonable inquiry) that we can have knowably reliable beliefs about its legitimacy.

---

14. See Raz's discussion of whether exclusionary reasons are, as Stephen Perry argues, 'reweighting reasons'. See: Stephen Perry, "Second-Order Reasons, Uncertainty and Legal Theory," *California Law Review* 62 (1988): 913–994; Joseph Raz, "Facing Up: A Reply," *California Law Review* 62 (1988).

15. Raz, "The Problem of Authority: Revisiting the Service Conception," p. 1025. See also: Raz, *From Normativity to Responsibility*.

16. Raz, "The Problem of Authority: Revisiting the Service Conception," p. 1025.

One problem with this approach is that unless knowledge is perfectly *luminous* (that is, when you know, you *know* that you know), it seems that knowability will not be sufficient for determining whether pre-emptive obedience is justified. And, as it turns out, it seems that knowledge is *not* perfectly luminous.[17] As such, it is unlikely that applying layers of knowability will ever amount to a complete account of rational obedience under uncertainty. From a measurement theory perspective, this should be expected: knowability seems to be a binary notion – you can know something or you cannot – whereas uncertainty admits of degrees; as such, it seems that any knowability constraint will need to draw a bright line between shades of uncertain facts to distinguish those which are knowable from those which are not. In doing so, any such constraint seems liable to make incorrect or arbitrary judgements about whether particular uncertain facts should guide us. To avoid these problems, we need more than knowability.

For this reason, it is tempting to build uncertainty *into* the standard of legitimacy, in the following way:

The Normal Justification Thesis (More Likely): You should treat an authority's commands pre-emptively just in case doing so is *more likely* to lead you to better conform with the reasons that independently apply to you, compared to if you acted on your own assessment of those reasons.[18]

If successful, then this approach provides an elegant answer to the problem of what to do in cases of uncertainty (here we will assume that the relevant interpretation of 'likelihood' is informed by your evidence).[19] As we shall see, however, taking this approach leads us to make a substantive commitment about reasons for action. Specifically, it requires us to hold that the normative weight of reasons is *probability-sensitive*. To see why, suppose that reasons are

---

17. See: Timothy Williamson, *Knowledge and Its Limits* (Oxford Scholarship Online, 2003), ch. 4 On this point with respect to Raz's more general theory of reasons, see: Ruth Chang, "Raz on Reasons, Reason, and Rationality: On Raz's From Normativity to Responsibility," *Jerusalem Review of Legal Studies* 8, no. 1 (2013): p. 208.

18. For instances of this formulation, see: Raz, "Authority and Justification"; Raz, "Facing Up: A Reply."

19. c.f. Raz, *From Normativity to Responsibility*, p. 115.

not probability-sensitive, and consider the following case: [20]

Questionable Orders: You have been ordered to deliver the life-saving goods by hang glider. If you obey this order, then there is a probability of 0.01 that you will save 101 people, and a probability of 0.99 that you will save no-one (say, because the goods will almost certainly be destroyed when you drop them from such a great height). If you follow your own judgement, you will take the nearby truck, which is slightly slower, but more reliable. Using it has a probability of 0.99 that you will save 100 people, and a probability of 0.01 that you will save no-one (say, by breaking down). All other considerations are equal.

In this case, on a probability-insensitive interpretation of reasons, the Normal Justification Thesis is satisfied because you are *more* likely (probability of 0.01 vs probability of 0) to *better* conform to reason (101 lives saved vs. 100 lives saved) by pre-emptively obeying. This is problematic because, clearly, obedience is morally reckless! This suggests that the Normal Justification Thesis (More Likely) should not be interpreted in terms of probability-insensitive reasons.[21]

Now, of course, it is possible that you could get lucky and save the 101 individuals. However, the Anarchist would (rightly) chalk this up as a victory, since this kind of indifference to the consequences of obedience makes us "the ready tool of injustice, cruelty, and profligacy; and, if at any time [we] are not employed in their purposes, it is the result of accident, not of [our]

20. Jackson, "Decision-Theoretic Consequentialism"; Raz, *From Normativity to Responsibility*, pp. 120-128.

21. Some might object that this is a case where the authority is clearly mistaken and, as such, the Service Conception allows you to disregard the order (Raz, 1986, p. 62). In response, however, this is not the relevant kind of 'clear mistake' (on Raz's view, a clear mistake is one that is detectable without having to deliberate about the excluded reasons). Secondly, even if the assignment of probabilities does yield a 'clear mistake', then we can slightly alter the probabilities to yield a structurally similar case where it is not clear. Lastly, in any case, there are good reasons to think that this 'clear mistake' exception is not compatible with Raz's overall view (see: Regan 1990, pp. 20-21). Raz, *The Morality of Freedom*, p. 62; Donald H. Regan, "Reasons, Authority, and the Meaning of 'Obey': Further Thoughts on Raz and Obedience to Law," *Canadian Journal of Law and Jurisprudence* 3 (1990): pp. 20-21.

own precaution and honesty."[22] In general, any adequate theory of legitimate authority must not tell us to 'just follow orders' and recklessly turn ourselves over to authority.

There are even worse problems in the vicinity if we understand the normative weight of reasons to be probability-insensitive. Consider two competing authorities: X and Y. X is more likely (with probability 0.51) to bring about a small gain in reasons conformity compared to your own judgement, but might (0.49) lead you to catastrophic results. Y is more likely (0.51) to lead you to slightly worse conformity with reason compared to following your own judgement, but may (0.49) lead to much better conformity with reason. By the Normal Justification Thesis, X has legitimate authority over you and Y does not have legitimate authority over you. And now here's the problem: if the probabilities that the authorities will improve or worsen your conformity with reason are independent of each other, then you face the following cyclical ranking of options: you should follow Y rather than X, you should follow yourself rather than Y, but you should follow X rather than yourself.[23] Table 5.1 illustrates this result, where the states RIGHT and WRONG refer to whether the authority correctly or incorrectly orders a course of action that is better than what your own judgement would lead you to:

X-STATES Y-STATES PROBABILITY	X RIGHT Y WRONG 0.26	X RIGHT Y RIGHT 0.25	X WRONG Y WRONG 0.25	X WRONG Y RIGHT 0.24
Obey X Obey Y	small gain small loss	small gain large gain	large loss small loss	large loss large gain

Table 5.1: Comparison of Competing Authorities

Here, we can see that the probability that following Y will lead you to a better outcome than following X is 0.74 (X does better than Y only when X RIGHT, Y WRONG). As such, you are more likely to better conform to reason by obeying Y, rather than X. At the same time, you are more likely to better

22. Godwin, *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness*, p. 233.
23. This case is based on: Paul A. Samuelson, "Risk and Uncertainty: A Fallacy of Large Numbers," *Scientia* 98 (1963): p. 4

conform to reason by obeying X rather than yourself. However, you are more likely to better conform to reason by following your own judgement, rather than obeying Y. What are you to do? In such cases of overlapping authorities, Raz counsels us to "decide, to the best of our ability, which is more reliable as a guide."[24] However, as we can see, if you adopt a probability-insensitive understanding of reasons, then you have no guide at all: whatever you do, you are doing something you ought not to do.

Let us diagnose the source of the problem. The probability-insensitive interpretation of the Normal Justification Thesis gives the wrong result because it grants legitimacy and dictates obedience merely on the basis of an ordinal test: it merely assesses whether the authority improves your conformity with reason. Under conditions of certainty, this test is sufficient to guide action. However, to guide action under conditions of risk and uncertainty, we need to also take into account cardinal information: specifically, *how much* improved conformity with reason should we expect from following the authority's orders? Does this potential improvement outweigh the risk that, if the authority is mistaken, we may do much worse if we follow the order? Thus, if we build uncertainty into the standard of justification, we must ensure that we include the relevant cardinal information. This can be done by making the weight of reasons sensitive to probability.[25]

Allowing that the weight of reasons is probability-sensitive yields the following version of the Normal Justification Thesis:

The Normal Justification Thesis (Probability-Weighted): You should treat an authority's commands pre-emptively if and only if doing so would *more likely* lead you to better conform with the probability-weighted balance of reasons that apply to you, compared to if you acted on your own assessment of those reasons.

---

24. Raz, "The Problem of Authority: Revisiting the Service Conception," p. 1021.

25. To what extent does this approach depart from orthodox understandings of the Service Conception? This is unclear. However, as it turns out, allowing that the weight of reasons is probability-sensitive does fit with Raz's recent views on normativity. Raz now holds that in cases like the above, we should act in accordance with the balance of probability-weighted reasons: "Given two worthwhile ends, other things being equal the reason to pursue the one more likely to be achieved is the better or stronger reason". Raz, *From Normativity to Responsibility*, p. 116.

In **Questionable Orders**, this leads to the correct verdict that you should not pre-emptively obey your superior officer. Note, however, that there is a tension in this interpretation of the Normal Justification Thesis: namely, there is a trivial sense in which obeying the authority *is* more likely to lead you to better conform to the probability-weighted reasons that apply to you, since doing so *just is* acting on the probability-weighted balance of your reasons. To act otherwise is *guaranteed* to not lead you to better conform to the probability-weighted balance of your reasons. As such, we can set aside the 'more likely' proviso in the Normal Justification Thesis (Probability-Weighted).

At this point of the analysis, it is helpful to note the structural similarities between a probability-weighted version of the Normal Justification Thesis and orthodox expected value theory. Expected value theory holds that we should maximise probability-weighted average value. The notion of 'value' is purely formal, providing a numerical representation of the relative importance of normative considerations. In this way, values provide a useful index for determining the relative importance of the normative considerations that are relevant to our actions. *Expected* value merely represents the idea that the importance of normative considerations (or 'reasons') is sensitive to probability.[26] The probability-weighted interpretation of the Normal Justification thus closely mirrors expected value reasoning. This suggests the following formulation:

The Normal Justification Thesis (MEV): You should treat an authority's commands pre-emptively if and only if doing so uniquely maximises expected value (MEV).

If this formulation is correct – and it seems to be suggested by the foregoing line of analysis – then the Service Conception can avoid problematic cases like **Questionable Orders** by ensuring that it tracks the recommendations of ex-

---

26. More precisely, expected value theory assumes that value varies in linear proportion to probability. There are, however, ways of allowing expected value theory to account for attitudes to risk (roughly, by making value vary linearly with probability, but non-linearly with quantity). See: Arrow, "The Theory of Risk Aversion"; Pratt, "Risk Aversion in the Small and in the Large"; Broome, *Weighing Goods*.

pected value theory. Indeed, to do otherwise would lead it to other problematic cases of involving uncertainty.[27]

Some might worry, however, that this analysis is unhelpful because it simply confirms an existing objection to the Normal Justification Thesis: namely, that it lacks explanatory power because it is silent about what exactly we have most reason to do.[28] For instance, for all we know, given the Normal Justification Thesis (MEV), we have most reason to obey a brutal dictatorship! It all depends on the underlying axiology. As Scott Hershovitz notes, this apparent emptiness of the Service Conception is unsatisfactory, given that:

> [Raz] presents [the Normal Justification Thesis] as an answer to the question "How can it ever be that one person has a duty to subject one's will and judgement to those of another?" But the normal justification thesis hardly answers that question if it can be satisfied simply because an authority passes some other test for legitimacy.[29]

However, I shall argue that whatever explanatory power the Normal Justification Thesis (MEV) loses through lack of *substance*, it makes up for in *structure*, by reconciling obedience to legitimate authority with our best theory of rational decision-making under uncertainty. As we shall see, this will allow us to gain valuable insights into how legitimate authorities can guide us in cases of uncertainty, and when obedience is unjustified.

Now, before proceeding, some might worry about the robustness of this identification between the Normal Justification Thesis and a decision-theoretic rule such as Maximise Expected Value. For example, one might wonder: Does the identity hold when we adopt the long-run formulation of the Normal Justification Thesis?[30]

---

27. Violations of expected utility theory are proven to lead to problematic verdicts, for example, in sequential choice cases. See: Katie Steele, "What are the Minimal Requirements of Rational Choice? Arguments from the Sequential-Decision Setting," *Theory and Decision* 68, no. 4 (2010): 463–487. For a classic discussion, see: Donald Davidson, J. C. C. McKinsey, and Patrick Suppes, "Outlines of a Formal Theory of Value, I," *Philosophy of Science* 22, no. 2 (1955): 140–160.

28. Scott Hershovitz, "The Role of Authority," *Philosophers' Imprint* 56, no. 7 (2011): 1–19.

29. ibid., p. 5.

30. Raz, *Ethics in the Public Domain*.

The Normal Justification Thesis (Long-Run): You should treat an authority's commands pre-emptively if and only if doing so would more likely lead you over the long run to better conform with the reasons that independently apply to you, compared to if you acted on your own assessment of those reasons.

As it turns out, including the long-run proviso only strengthens the identification with maximising expected value. This is because, over the long-run, expected value maximisers become overwhelmingly more likely to gain higher average value than non-expected value maximisers.[31] This means that over the long-run, pre-emptive obedience maximises chances of higher average long-run value if and only if doing so maximises expected value.

In the following section, I show how accepting the Normal Justification Thesis (MEV) strengthens the Service Conception, allowing it to draw on other important resources in decision theory. This will yield a generalised version of the Service Conception of authority that can systematically account for all cases of uncertainty.

5.3   The Value of Legitimate Authority

Having received an order from a legitimate authority, one choice you must make is whether it is better to ignore the order or to obey it. The Service Conception implies that if the authority is legitimate, then it is prospectively better to obey the authority rather than ignore it and follow your own judgement. One source of justification is that the authority is a more reliable judge than you. Another is that pre-emptively obeying its orders is more efficient, saving you the costs of information-gathering and deliberation. In this section, we will model the 'reliability justification' with expected value theory. We will see that calibrating your judgement to that of a more reliable authority leads to higher expected value than judging for yourself. With this result in hand, it is a trivial exercise to add cost-saving considerations to the model (say, by adding

---

31. Samuelson, "Risk and Uncertainty: A Fallacy of Large Numbers"; Johanna Thoma, "Risk Aversion and the Long Run," *Ethics*, 2019, 1–31.

or subtracting values to the possible outcomes, reflecting the costs incurred or saved by deciding for oneself versus obeying the authority).

To illustrate how greater reliability allows an authority to satisfy the Normal Justification Thesis (MEV), consider the following case:

Reliable Order: You are deliberating about whether to commandeer the civilian's van or to use a military-issue truck. If the road ahead is paved, then you should use the van (the truck is slower). If the road is unpaved, you should use the truck (in such conditions, the van would be slower). You think that the road ahead is unpaved. In these cases, you take yourself to be 60% reliable about the correct course of action. You are about to receive an order from your superior officer, who you know to be right in these cases 90% of the time. Your decision situation is shown in Figure 5.1.

In this case, given that the authority's judgement is deemed to be a more reliable guide to the world than your own, choosing to pre-emptively obey has greater expected value, whatever the content of the order turns out to be. The screening-off of your judgement is justified by the fact that including your own beliefs (say, by combining your prior judgement with the authority's) would amount to 'muddying the waters' or 'introducing noise' into the analysis of the situation, leading you to a prospectively worse set of choices.[32]

This decision-theoretic approach also gives a principled way of determining when an authority should or should not be treated as having legitimate authority. Firstly, we need to be sufficiently sure that the authority is more reliable than us. This turns out to be consistent with Raz's position – which we discussed earlier in relation to the knowability constraint – that reliable beliefs about an authority's legitimacy are generally necessary for improved conformity with reason.[33] On the decision-theoretic analysis, if there is no available

---

32. This reasoning is closely related to those based on expert principles in formal epistemology, see: Richard Pettigrew, "Accuracy, Chance, and the Principal Principle," *Philosophical Review* 121, no. 2 (2012): 241–275; Richard Pettigrew and Michael G Titelbaum, "Deference Done Right," *Philosopher's Imprint* 14, no. 35 (2014): 1–19.

33. Raz, "The Problem of Authority: Revisiting the Service Conception," p. 1025.

Probabilities	0.6	0.4	
Possibilities	Unpaved	Paved	EV
Van	50	100	70
Truck	80	80	80

	0.1	0.9	
	Unpaved	Paved	EV
Van	50	100	95
Truck	80	80	80

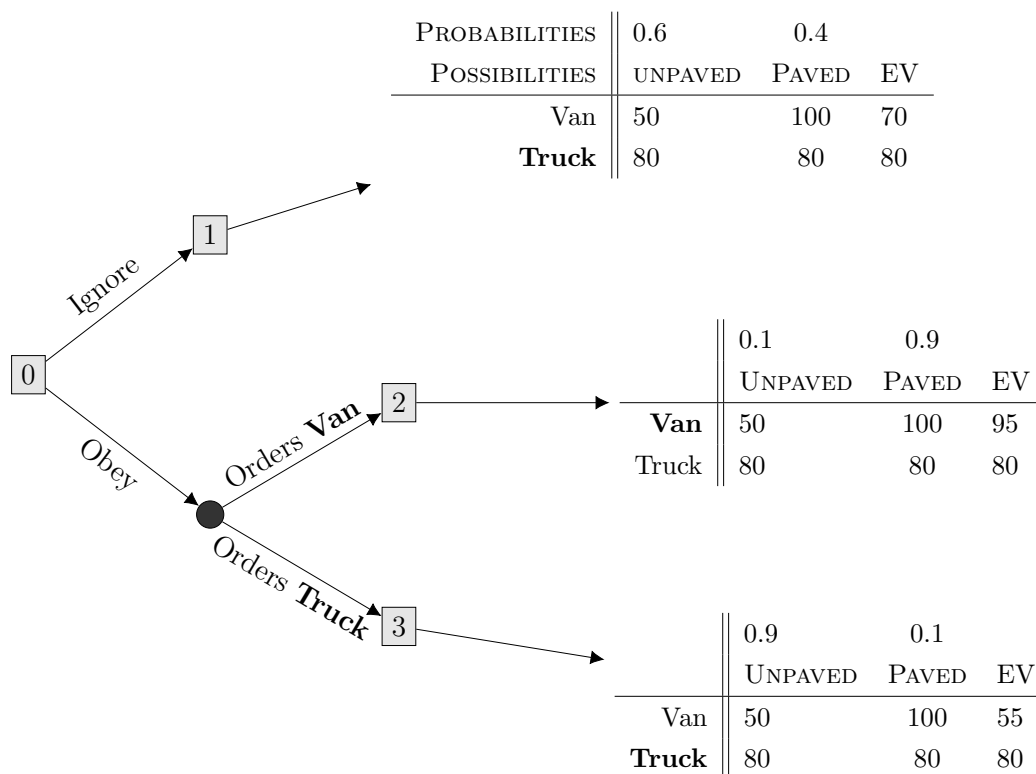|	0.9	0.1	
	Unpaved	Paved	EV
Van	50	100	55
Truck	80	80	80

Figure 5.1: Value of Legitimacy

information about the authority's legitimacy, then acting on the basis of its orders does not necessarily have determinately greater (or equal) expected value compared to ignoring it. As such, you may be justified in disobeying. For what it is worth, the Anarchist agrees with this verdict:

> Wherever I have good reason to believe that another person knows better than myself what is proper to be done, there I ought to conform to his direction. But the advantage which he possesses must be obvious, otherwise I shall not be justified in my proceeding.[34]

To illustrate, suppose now that you are uncertain about whether your authority is more reliable than you. As such, you must weigh up the prospects of following a reliable authority versus an unreliable one (see: Figure 5.2). Suppose that your superior orders you to take the van. If they are reliable, then it

---

34. Godwin, *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness*, p. 236.

is 0.9 probable that they are correct, such that the expected value of obeying is 95. However, if they not more reliable, then (due to their incompetence or malice) it will only be 0.2 probable that they are correct, such that following their order will have an expected value of 60. In this case, you do better by obeying the authority so long as you are sufficiently confident that it is more reliable (in this case, this means being at least about 60% confident about this). Any less and you would be not justified in proceeding.

PROBABILITIES	0.6	0.4	
POSSIBILITIES	UNPAVED	PAVED	EV
Van	50	100	70
Truck	80	80	80

	0.1	0.9	
	UNPAVED	PAVED	EV
Van	50	100	95
Truck	80	80	80

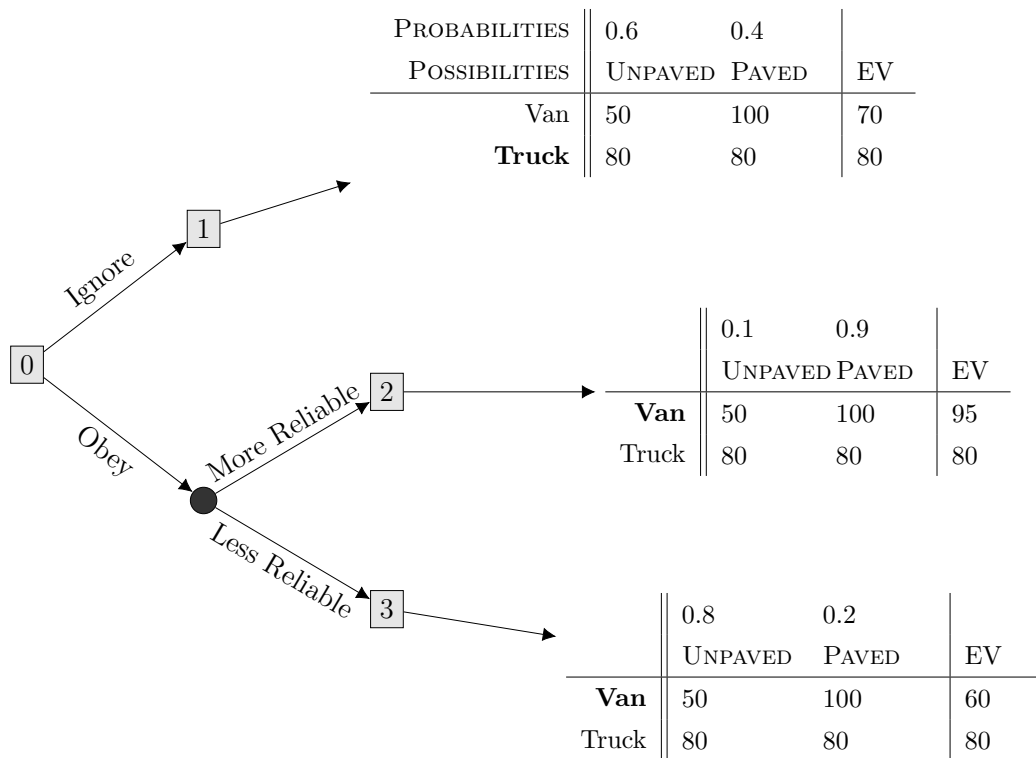|	0.8	0.2	
	UNPAVED	PAVED	EV
Van	50	100	60
Truck	80	80	80

Figure 5.2: Uncertain Reliability

A second case in which pre-emptive obedience is not better than ignoring the order is when determining the reliability of the authority is too costly to be worth the potential improvement. This, also, is consistent with Raz's position:

> How much it can be expected to improve our conformity to reason, and how important the matter is, establish what inquiry is reasonable to undertake. When reasonable inquiry will not reveal the case for authority, that case, if it exists at all, is unknowable. It follows that

people are not subject to any authority regarding those matters. This argument is used here to establish not merely that it is not rational, or not worthwhile, to carry on with the inquiry about the existence of certain reasons, but that those reasons, authoritative directives, do not exist. There is no authority over the matter, because to exist, authorities must be knowable [in the sense that the cost of acquiring their orders is not greater than the benefit].

In the previous example, if the time it takes you inquire about the authority's reliability costs, say, 15 lives, then the expected value of the inquiry is not worth the cost. This would be another case where you should decide on your own.

Admittedly, these are merely first steps towards reconciling our theories of legitimate authority and rational decision-making under risky situations. The foregoing suggests that modelling the Service Conception in decision-theoretic terms gives a systematic and precise explanation for when pre-emptive obedience is better than deciding alone, and when it is not. Moreover, it seems largely consistent with Raz's earlier stated views on authority and his recent work on reasons. Overall, this analysis sketches how the Service Conception's refutation of Anarchism extends into cases involving uncertainty, by showing when and why pre-emptive obedience is morally and rationally justifiable.

5.4   Objections

I have argued that theories of legitimate authority must face up to the challenges of uncertainty. The approach sketched above is the first attempt in the literature to systematically address these important challenges. In particular, I have suggested that the Service Conception can be modelled using decision theory to give adequate guidance in risky cases.

For their part, decision theorists might be surprised that this project has not already been developed. After all, they have long had a wealth of theoretical resources for dealing with expert testimony, conflicting evidence, and related issues. Why have theorists of legitimate authority apparently ignored this work?

It turns out that many theorists of legitimate authority are skeptical about decision-theoretic approaches to authority. For example, in an influential critique, Scott Shapiro (2004) presents a litany of objections to these approaches, which he calls the Decision Model. He argues that "not only is the Decision Model false in some type of cases, it is false in all types of cases."[35] Below, I introduce and respond to his objections. Theories of legitimate authority have no good reason to ignore the resources available in decision theory.

5.4.1 Motivational Weakness

Suppose that your superior officer, who you recognise to be a legitimate authority, orders you to take the van. However, suppose also that you are sure that were you to act on that order, you would fail due to weakness of will (you always wanted to drive the truck, and you are sure you will somehow delude yourself into believing that you should take it, instead). In this case, Shapiro argues that if you deliberate about whether to obey your superior officer, then you will be unable to benefit from his order, since your weakness of will (by stipulation) stops you from being able to follow the order. Shapiro takes this to show that "authoritative directives can serve the benefits they are meant to serve just in case we think that we have the ability to [causally] constrain our future selves."[36] Since the Decision Model does not represent us as being causally constrained by an authority's orders, it fails to explain how legitimate authorities can benefit the weak-willed. On an instrumental justification such as the Service Conception, legitimate authorities benefit their subjects. Thus, it seems that the Decision Model leads us to the absurd conclusion that there are no legitimate authorities when it comes to the weak-willed.

In response, note firstly that this objection moves the goal posts for decision-theoretic approaches. Our original challenge was to explain how obedience to authority is compatible with being rational and morally conscientious. Since the weak-willed are arguably practically *irrational*, it is no surprise that they may not benefit from a legitimate authority. In the example above, the author-

---

35. Shapiro, *Authority*, p. 420.
36. ibid., pp. 429-430.

ity is legitimate because doing as it says has greater expected value compared to acting on your own. If you do not do as it says, you are acting irrationally, but the authority is still legitimate. As such, there is no problematic 'false negative' result when it comes to the weak-willed. Shapiro's criticism therefore does nothing undermine the Decision Model.

Here is perhaps a more conciliatory response. The expected value approach determines what you ought to do, given a description of the decision problem. Shapiro's contention can be read as asserting *how* decision problems should be described for particular kinds of agents: specifically, he argues that the weak-willed should describe decision problems so that they do not include particular future actions (namely, disobedience) as being available for choice. His contention thus aims at a higher-order decision problem: how should we describe decision problems, given our behavioural tendencies to respond to particular kinds of decision problems? Note that Shapiro's approach can be justified by adopting an expected value approach to second-order decisions, whereby we assign probabilities to our acting in an optimal fashion, given the choices available to us.[37] If we are weak-willed and the expected value of treating future choices as constrained is optimal, then we should adopt Shapiro's approach to modelling the decision problem. However, if we are not weak-willed, we can treat disobedience as an available, if irrational, option. Both approaches are consistent with a decision-theoretic model.

5.4.2 Guiding Significantly Uncertain Agents

Shapiro argues that the following kinds of cases show that the Decision Model is false:

Radical Cluelessness: You have maximally indeterminate credences regarding what to do. Luckily, an authority who you take to be reliable has just told you what to do. You update your credences, but you are still clueless.

---

37. One might worry here that we cannot rationally assign probabilities to our own choices. For arguments to the contrary, see: Alan Hájek, "Deliberation Welcomes Prediction," *Episteme* 13, no. 4 (2016): 507–528.

Shapiro holds that this is a counterexample to the Decision Model, since it does not give the correct verdict about what you ought to do: he submits that you should consider yourself bound to follow the authority's order.[38] More generally, he states that: "[Isaac] Levi has shown that it is usually not possible for a rational agent to harness the informational value of theoretical authorities if the agent were free not to follow the advice, but nevertheless treats the advice as reliable evidence."[39]

There are a number of problems with this line of argument. Firstly, Shapiro's argument appropriates Levi's argument against a specific theory of inductive logic called Objectivist Necessitarianism. The Decision Model is not committed to this theory, or indeed any particular theory of inductive logic. Indeed, even Levi muses that "[p]erhaps no one has ever been a strict objectivist necessitarian."[40] Secondly, Levi's argument does not support the idea that agents should not be free to follow advice from theoretical authorities. It merely supports the idea that any adequate theory of formal epistemology should include the notion of credal commitment, which is a rule governing how we should epistemically respond to particular kinds of evidence. This notion of credal commitment is orthogonal to Shapiro's view that benefiting from authority requires us to act as though we could not possibly disobey. Indeed, if anything, the idea of credal commitment is better suited to the Decision Model.[41]

Shapiro further argues that the Decision Model cannot "validate acceptance of expert advice when the agent's epistemic state is highly indeterminate [in the sense of assigning a low, but precise, credence to a range of possibilities]."[42] He has in mind the following kind of scenario:

Mild Cluelessness: You have mildly indeterminate credences regarding what to do. Luckily, an authority you take to be reliable has just told you what

38. Shapiro, *Authority*, pp. 421-3.

39. ibid., p. 421.

40. Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*.

41. Shapiro, *Authority*, p. 421; Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*.

42. Shapiro, *Authority*, p. 423.

to do. You update your credences, but you still do not come to accept that you ought to do as directed.

In his discussion of this case, Shapiro erroneously assumes that accepting a decision-theoretic approach requires endorsing an acceptance-based epistemology, whereby propositions are 'accepted' just in case they are assigned a sufficiently high probability (or likelihood). On the decision-theoretic approach presented earlier, so long as the authority's order improves your later choice situations, then the authority is legitimate and you should obey, even if you do not necessarily 'accept' or 'believe' the proposition implied by their order. However, as we saw earlier, if you are not sufficiently certain about the authority's reliability, then you should not obey the authority – rather, you should decide for yourself.

5.4.3 Deciding to Decide to Obey

Shapiro also contends that no decision-theoretic approach can account for cases where authorities save us from having to undertake costly and risky deliberation. Consider the following scenario:[43]

Deciding to Obey: Your commanding officer – whose judgement you consider to be as reliable as your own – has just issued you an order. So you think to yourself: should I just obey automatically (saving time and effort) or deliberate for myself?

Shapiro argues that by treating obedience as optional, you have already lost the benefits of authority in this case. Even worse, it seems that you face a (potentially infinite) regress of deciding how to decide. You must deliberate about whether deliberation is most appropriate, which itself is a mode of decision-making that must be justified by further deliberation, and so on.

In the face of this objection, we must ask: what is the alternative course of action? In general, given the risks associated with obeying an illegitimate authority, we must always assess whether obedience to authority is more reliable or efficient than following our own judgement. In this case, you have

---

43. Shapiro, *Authority*, pp. 423-425.

assessed the authority to be no more reliable than yourself. However, that is not enough to determine its legitimacy. You must also assess whether obedience is more efficient. The authority will be legitimate on this efficiency score only if assessing its legitimacy is worth the potential savings in time and cost. If this assessment is more costly than the potential efficiency gains, then the authority is not legitimate. Pre-emptive obedience is "optimal from an ex ante perspective" only once we can trust that the authority is legitimate.[44] To act other would be reckless and – most likely – irrational.

As for the regress, this is a general problem for theories of practical rationality, for which there are solutions. For instance, we can hold that you should deliberate in a way that is *robustly goal-conducive*, in a sense defined by Hanti Lin (2014).[45] This approach can take into account time and effort costs in deliberations about how to deliberation, while also allowing that even after hearing the order, you can choose to disobey (though, as before, doing so may be irrational). In general, there is no need to treat disobedience as an 'unavailable' option. Even after choosing to obey an authority, it is always open to us to disobey, even if doing so is contrary to reason.

5.4.4 Coordination Problems

Shapiro further argues that "if authorities are able to solve coordination problems, then the Decision Model cannot be correct."[46] He presents the following case:

> The Decision Model claims that it is rational for a player, call him X, to decide to comply with such a directive when, and only when, X can establish that it provides good evidence about other players' behaviour. But the directive provides good evidence about others' behaviour only when it would be rational for others to follow it. However, if these players are rational, the question of whether it is rational for them to follow the directive is the same as whether it is rational for X to follow the directive. Hence, X can establish that the directive constitutes good evidence only if X can first establish that it is rational for him to follow

44. ibid., p. 423.
45. Lin, "On the Regress Problem of Deciding How to Decide"
46. Shapiro, *Authority*, p. 424.

the directive. X has now travelled in a circle. If X wants to establish the rationality of his following the directive, it seems that he must already know that it is rational for him to follow the directive. But since he is trying to establish the rationality of following the directive, he cannot assume the proposition for the purpose of proving it. So, if X does not already believe that it is rational for him to follow the directive, he will never come to that conclusion.[47]

This argument is unsound because it ignores the fact that evidence about the authority and its ability to secure coordination can be sufficient to guide action, even if we do not know directly what is rational for others to do. To illustrate, consider the following case:

Coordination: You are driving to a location to hand over the supplies to the local medical staff. However, during your call with them, the phone line went down, preventing you from coordinating on a location. Either they will meet you at the top of the hill (TOP) or at the bottom of the hill (BOTTOM) – you are not sure which. Luckily, you have just received notification from an authority to meet at the bottom of the hill. Your evidence suggests that this authority has a very good track record in securing coordination. Figure 5.3 illustrates the decision situation.

In this case, nothing prevents you from coordinating appropriately with the medical staff on the basis of the authority's directive. Since you consider the authority to be legitimate, you take its order as giving you evidence about what will occur, given that you choose one location or the other. Thus, we have a case where our decision-theoretic approach to authority secures coordination, falsifying Shapiro's claim.

How does a decision-theoretic approach secure coordination in cases where there is a Prisoners' Dilemma at hand? Roughly, these are cases where there are greater social benefits to be gained if a sufficient number of parties cooperates; however, the problem is that it is rational for a sufficient number of parties to *not* cooperate, leading to a socially suboptimal outcome overall. If it is rational for these parties to not cooperate, then it follows *a fortiori*

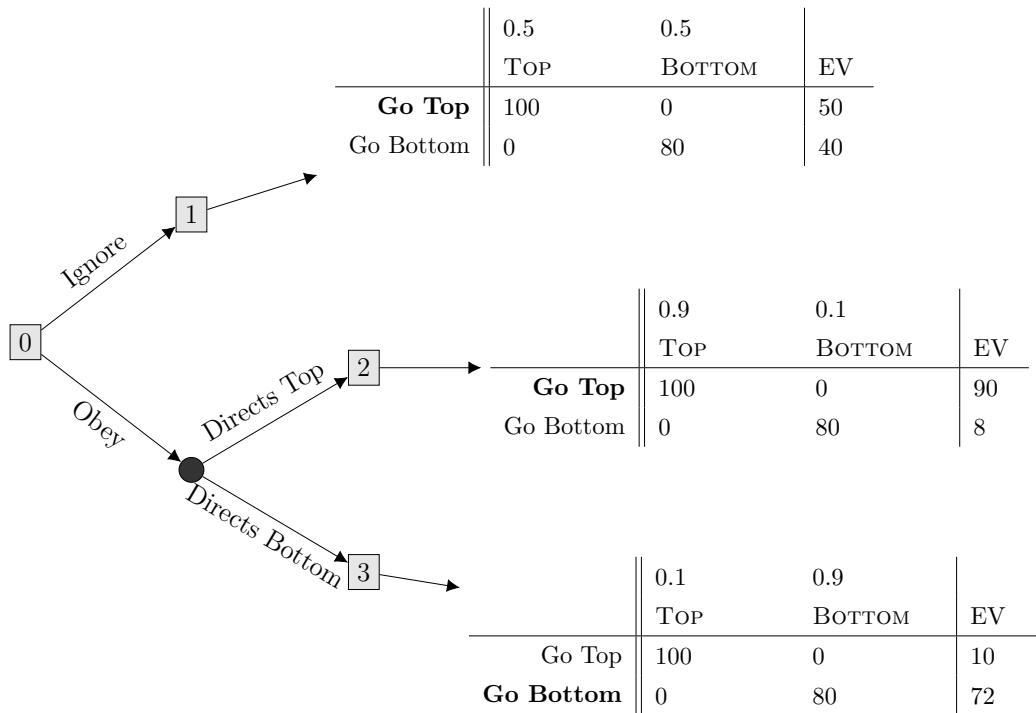47. Shapiro, *Authority*, pp. 426-427.

Figure 5.3: How Legitimate Authority Solves Coordination Problems

that, from their perspective, obeying an authority who merely directs them to cooperate does not maximise expected value. Thus, it seems that if we take a decision-theoretic approach, then legitimate authorities cannot resolve Prisoner's Dilemmas, which is one of the main functions of political authority.

These cases reveal that to solve Prisoners' Dilemmas, a legitimate authority must have a certain amount of *de facto* authority. That is, they must have an ability to change the 'payoffs' of cooperation or non-cooperation. For instance, authorities may need to have the power to punish non-cooperators, or to give additional incentives to make cooperation rational. In the decision-theoretic model given above, this can be easily represented by assigning additional value or disvalue to cooperation versus non-cooperation. Thus, far from being unable to explain how legitimate authorities resolve Prisoners' Dilemmas, decision theory illuminates exactly how they do so.

5.4.5 Practical Versus Theoretical Authority

One final objection to the decision-theoretic approach given here is that it seems to confuse practical authority with theoretical authority. Legitimate *practical* authority, properly understood, is about how agents can give other agents reasons for action by issuing commands. By contrast, theoretical authorities merely give reasons for belief.[48] The objection is that rendering the Service Conception with expected value theory reduces legitimate practical authority to legitimate theoretical authority.

However, this does not follow. If the distinction between practical and theoretical authorities rests on the kinds of speech-acts that they employ – commanding versus advising – then this is consistent with the formal approach given above. Alternatively, if the distinction tracks differing social or institutional roles, a more nuanced decision-theoretic approach could take into account the risks of enshrining, through obedience, an authority who is not duly empowered by the relevant power-conferring social or institutional norms.[49] Just as it is easy to underestimate the flexibility of the Service Conception, it is easy to underestimate the flexibility of the decision-theoretic model of it provided here.

As it stands, none of the examples given above appear to be cases where legitimate authorities are merely theoretical authorities. To see why, note that theoretical authorities typically give us first-order evidence or second-order evidence.[50] If they give first-order evidence, we should revise our beliefs on the basis of that evidence and our prior beliefs about what is best to do. This is not what occurs in the decision-theoretic account given above: the authority's order screens-off our understanding of the situation. One might then argue that this sounds more like the authority is giving second-order evidence that undermines our evidence.[51] For example, if our eyes tell us that a room is red, that first-order sensory evidence may be undermined by someone who gives

---

48. Raz, "Legitimate Authority"; Green, *The Authority of the State*, ch. 1.
49. Andrei Marmor, "An Institutional Conception of Authority," *Philosophy & Public Affairs* 39, no. 3 (2011): 238–261; Hershovitz, "The Role of Authority."
50. Of course, they could give still higher higher-order evidence, but those cases are even less plausible than those discussed here.
51. Thanks here to Sergio Tenenbaum for discussion.

us reason to believe that our seeing red is a drug-induced hallucination. In that case, an appropriate response might be to suspend our judgement about whether the room is red. Clearly, however, this is not a plausible interpretation of the cases given above: the authority's order does not give us reason to suspend our belief in our assessment of the situation. Rather, it gives us reason to do as it says because it is a more reliable or cost-effective guide to the situation than following our own judgement. The expected value approach presented above simply provides a way of measuring how legitimate authority changes our reasons for action in the face of uncertainty.

Of course, there is much more to be said on this point. It may very well be that, in cases of uncertainty, the divide between practical and theoretical authority is not as great as many previously thought. Then again, it may well turn out that the approach given here offers too thin a conception of the nature of legitimate authority. If so, then those who seek a thicker notion of legitimate authority will have to provide an alternative account of how legitimate authority is consistent with rational decision-making in risky cases.

5.5 Conclusion

The Service Conception of authority can – and, indeed, should – be understood in a way that conforms with our best formal theories of rational decision. This approach allows us to precisely determine the conditions under which pre-emptive obedience to authority is rational, given our evidence. Once further developed and defended, this account may complete Raz's answer to the Philosophical Anarchist, showing that even in cases of uncertainty, obedience to authority can be entirely consistent with the decisions of our own understanding, and the dictates of our conscience.

Conclusion

The world is a risky place.[52] Given this fact, what does morality require of us? Many have argued that morality requires us to compromise, to make trade-offs that seem morally unpalatable. In their eyes, anyone who resists compromise – by say, prioritising an innocent individual's life over *any* amount of trivial benefits for others – is a *Moral Absolutist*. These critics argue that absolutist moralities cannot guide us through a risky world because they seem to give verdicts that are either inconsistent or implausible. As such, these critics conclude that moral compromise is inevitable: lives *can* be outweighed by trivial benefits. I called this line of critique the Problem of Risk.

In this dissertation, I argued that these critics are mistaken: using an appropriate decision-theoretic approach, almost all absolutist moral theories give adequate verdicts about what to do in a risky world. I argued as follows:

Chapter 1: Prohibition and Probability introduced the debate over the Problem of Risk. I presented a formal feature, called Option Absolutism, that commits a moral theory to the Problem of Risk, and I argued that almost all broadly 'absolutist' theories can reject Option Absolutism in favour of Relational Absolutism, which is not committed to the Problem of Risk.

Chapter 2: Moral Priorities Under Risk gave an example of how one class of so-called absolutist theories – lexical priority theories – can adopt a Relational Absolutist structure using lexicographic decision theory.

---

52. Hat tip to John Cusbert.

Chapter 3: Priorities and Uncertainties defended lexical priority theories using orthodox expected value theory. In doing so, it showed how some particular objections to moral absolutism actually rely on other issues in moral theory, such as whether individual or sequences of actions are the proper objects of moral evaluation.

Chapter 4: Duty and Ignorance demonstrated how value of information calculations from decision theory can allow deontologists to correctly determine when to gather more information before acting.

Chapter 5: Authority, Obedience, and Uncertainty argued that the seemingly absolutist Service Conception of authority can be extended, using an expected value approach, to cases where we are uncertain about an authority's legitimacy.

Despite the positive results found in the chapters above, more work is required to determine that the formal decision models presented are the *most* suitable ones available for modelling the moral theories in question. Nevertheless, these chapters demonstrate a number of overlooked points in the discussion of the Problem of Risk. Firstly, they show the versatility of decision-theoretic modelling as a tool for representing moral theories. Secondly, they help us to identify various gaps in our existing moral theorising, including the rationales for positing particular kinds of priority relations between various moral considerations.[53] Thirdly, and most importantly, they show that morality can guide us through a risky world, without compromise.

---

53. See, for example, Chapter 2, Section 2.6

Bibliography

Aboodi, Ron, Adi Borer, and David Enoch. "Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith." *Journal of Philosophy* 105, no. 2 (2008): 259–272.

Altham, J. E. J. "Ethics of Risk." *Proceedings of the Aristotelian Society, New Series* 83 (1983): 15–29.

Anscombe, F. J., and R. J. Aumann. "A Definition of Subjective Probability." *The Annals of Mathematical Statistics* 34, no. 1 (1963): 199–205.

Anscombe, G. E. M. *Ethics, Religion and Politics.* Oxford: Basil Blackwell, 1981.

———. "Modern Moral Philosophy." In *Virtue ethics*, edited by Roger Crisp and Michael A Slote, 26–44. Oxford: Oxford University Press, 1997.

Arrhenius, Gustaf, and Wlodek Rabinowicz. "Value Superiority." In *The Oxford Handbook of Value Theory*, 225–243. Oxford University Press, 2015.

Arrow, Kenneth J. "The Theory of Risk Aversion." Chap. 2 in *Aspects of the Theory of Risk Bearing.* Helsinki: Yrjo Jahnsonin Saatio, 1965.

Aumann, Robert J. "Utility Theory without the Completeness Axiom." *Econometrica* 30, no. 3 (1962): 445–462.

Bales, R Eugene. "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8, no. 3 (1971): 257–265.

Bartha, Paul. "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities." *Synthese* 154, no. 1 (January 2007): 5–52.

Bjorndahl, Adam, Alex John London, and Kevin J. S. Zollman. "Kantian Decision Making Under Uncertainty: Dignity, Price, and Consistency." *Philosopher's Imprint* 17, no. 7 (2017): 1–22.

Bradley, Seamus, and Katie Steele. "Can Free Evidence Be Bad? Value of Information for the Imprecise Probabilist." *Philosophy of Science* 83, no. 1 (2016): 1–28.

Briggs, R.A. "Costs of abandoning the Sure-Thing Principle." *Canadian Journal of Philosophy* 45, nos. 5-6 (2015): 827–840.

———. "Normative Theories of Rational Choice: Expected Utility Theory." *The Stanford Encyclopedia of Philosophy*, 2017.

Broome, John. "No Argument against the Continuity of Value: Reply to Dorsey." *Utilitas* 22, no. 04 (November 2010): 494–496.

———. *Weighing Goods*. Cambridge, Massachusetts: Blackwell, 1995.

Buchak, Lara. *Risk and Rationality*. Oxford: Oxford University Press, 2013.

Carlson, Erik. "Aggregating Harms — Should We Kill to Avoid Headaches?" *Theoria* 66, no. 3 (2000): 246–255.

———. "Organic Unities, Non-Trade-Off, and the Additivity of Intrinsic Value." *Journal of Ethics* 5 (2001): 335–360.

Chang, Ruth. "Raz on Reasons, Reason, and Rationality: On Raz's From Normativity to Responsibility." *Jerusalem Review of Legal Studies* 8, no. 1 (2013): 199–219.

Colyvan, Mark, Damian Cox, and Katie Steele. "Modelling the Moral Dimension of Decisions." *Noûs* 44, no. 3 (September 2010): 503–529.

Cusbert, John. "Acting on Essentially Comparative Goodness." *Thought: A Journal of Philosophy* 6, no. 2 (2017): 73–83.

Davidson, Donald, J. C. C. McKinsey, and Patrick Suppes. "Outlines of a Formal Theory of Value, I." *Philosophy of Science* 22, no. 2 (1955): 140–160.

Dietrich, Franz, and Christian List. "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories." *The Philosophical Review* 126, no. 4 (2017): 421–479.

Dorsey, Dale. "Headaches, Lives and Value." *Utilitas* 21, no. 01 (2009): 36–58.

Easwaran, Kenny. "Decision Theory without Representation Theorems." *Philosophers' Imprint* 14, no. 27 (2014): 1–30.

Elga, Adam. "Subjective Probabilities should be Sharp." *Philosophers' Imprint* 10, no. 5 (2010): 1–11.

Elliott, Edward. "A Representation Theorem for Frequently Irrational Agents." *Journal of Philosophical Logic* 46, no. 5 (2017): 467–506.

Ellsberg, D. "Classic and Current Notions of "Measurable Utility"." *The Economic Journal* 64, no. 255 (1954): 528–556.

Fraassen, Bas C. van. "Singular Terms, Truth-Value Gaps, and Free Logic." *Journal of Philosophy* 63, no. 17 (1966): 481–495.

Fraassen, Bas van. *Laws and Symmetries.* Oxford: Clarendon Press, 1989.

Gewirth, Alan. "Are There Any Absolute Rights?" *The Philosophical Quarterly* 31, no. 122 (1981): 1–16.

Godwin, William. *Enquiry Concerning Political Justice and Its Influence on Morals and Happiness.* Toronto: University of Toronto Press, 1946.

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

Good, I. J. "On the Principle of Total Evidence." *The British Journal for the Philosophy of Science* 17, no. 4 (1967): 319–321.

Graham, Peter A. "In Defense of Objectivism about Moral Obligation." *Ethics* 121, no. 1 (2010): 88–115.

Green, Leslie. *The Authority of the State.* Oxford: Clarendon Press, 1988.

Hájek, Alan. "Deliberation Welcomes Prediction." *Episteme* 13, no. 4 (2016): 507–528.

———. "Waging War on Pascal's Wager." *The Philosophical Review* 112, no. 1 (2003): 27–56.

Hájek, Alan, and Christopher Hitchcock, eds. *The Oxford Handbook of Probability and Philosophy.* Oxford: Oxford University Press, 2016.

Hájek, Alan, and Michael Smithson. "Rationality and Indeterminate Probabilities." *Synthese* 187 (2012): 33–48.

Hammerton, Matthew. "Distinguishing Agent-Relativity from Agent-Neutrality." *Australasian Journal of Philosophy*, 2018, 1–12.

Hansson, Sven Ove. "Ethical Criteria of Risk Acceptance." *Erkenntnis* 59, no. 3 (2003): 291–309.

———. *The Ethics of Risk: Ethical Analysis in an Uncertain World.* Palgrave Macmillan, 2013.

Hare, Caspar. "Obligation and Regret When There is No Fact of the Matter About What Would Have Happened if You Had not Done What You Did." *Noûs* 45, no. 1 (2011): 190–206.

———. "Should We Wish Well to All?" *The Philosophical Review* 125, no. 4 (2016): 451–472.

———. *The Limits of Kindness.* Oxford: Oxford University Press, 2013.

———. "Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?" *Ethics* 117, no. 3 (2007).

Hausner, Melvin. "Multidimensional Utilities." Chap. 12 in *Decision Processes*, edited by R. M. Thrall, C. H. Coombs, and R. L. Davis. New York: John Wiley / Sons, Inc., 1954.

Hawley, Patrick. "Moral Absolutism Defended." *Journal of Philosophy* 105, no. 5 (2008): 273–275.

Hayenhjelm, Madeleine, and Jonathan Wolff. "The Moral Problem of Risk Impositions: A Survey of the Literature." *European Journal of Philosophy* 20, no. 51 (2011): E26–E51.

Hedden, Brian. "Options and Diachronic Tragedy." *Philosophy and Phenomenological Research* 87, no. 1 (June 2013): 1–35.

Hershovitz, Scott. "The Role of Authority." *Philosophers' Imprint* 56, no. 7 (2011): 1–19.

Horty, John. *Reasons as Defaults*. Oxford University Press, 2012.

Houy, Nicolas, and Koichi Tadenuma. "Lexicographic compositions of multiple criteria for decision making." *Journal of Economic Theory* 144, no. 4 (2009): 1770–1782.

Huemer, Michael. "Lexical Priority and the Problem of Risk." *Pacific Philosophical Quarterly* 91 (2010): 332–351.

Isaacs, Yoaav. "Duty and Knowledge." *Philosophical Perspectives* 28, no. 1 (2014): 95–110.

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

———. "How Decision Theory Illuminates Assignments of Moral Responsibility." Chap. 2 in *Intention in Law and Philosophy*, edited by Ngaire Naffine, Rosemary J Owens, and John Williams, 19–36. Aldershot, UK: Ashgate, 2001.

Jackson, Frank, and Robert Pargetter. "Oughts, Options, and Actualism." *The Philosophical Review* 95, no. 2 (1986): 233–255.

Jackson, Frank, and Michael Smith. "Absolutist Moral Theories and Uncertainty." *The Journal of Philosophy* 103, no. 6 (2006): 267–283.

Jeffrey, Richard C. *The Logic of Decision.* Second. Chicago and London: The University of Chicago Press, 1983.

Jensen, Kartsen Klint. "Millian Superiorities and the Repugnant Conclusion." *Utilitas* 20, no. 3 (2008): 279–300.

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

———. "How Probabilities Reflect Evidence." *Philosophical Perspectives* 19 (2005).

———. *The Foundations of Causal Decision Theory.* xii, 268 p. Cambridge; New York: Cambridge University Press, 1999.

Kagan, Shelly. *The Limits of Morality.* Oxford: Clarendon Press, 1991.

Kahneman, Daniel. *Thinking, Fast and Slow.* New York: Farrar, Straus / Giroux, 2011.

Kamm, Frances M. "Supererogation and Obligation." *Journal of Philosophy* 82, no. 3 (1985): 118–138.

Kant, Immanuel. *Groundwork for the Metaphysics of Morals.* Edited by Allen W. Wood. New Haven; London: Yale University Press, 2002.

———. *The Metaphysics of Morals.* Edited by Mary J Gregor (trans.) Cambridge University Press, 1996.

Kaplan, Mark. *Decision Theory as Philosophy.* Cambridge: Cambridge University Press, 1996.

Keynes, John Maynard. *A Treatise of Probability*. London; New York: Macmillan; AMS.

Kyburg, H.E. *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press, 1961.

Lazar, Seth. "Anton's Game: Deontological Decision Theory for an Iterated Decision." *Utilitas* 29, no. 1 (2017): 88–109.

———. "Deontological Decision Theory and Agent-Centered Options." *Ethics* 127, no. 3 (2017): 579–609.

———. "Deontological Decision Theory and the Grounds of Subjective Permissibility." 2018.

———. "In Dubious Battle: Uncertainty and the Ethics of Killing." *Philosophical Studies*, 2017, 1–27.

———. "Limited Aggregation and Risk." 2018.

———. "Moral Sunk Costs." *The Philosophical Quarterly*, 2018, 1–40.

———. "Risky Killing: How Risks Worsen Violations of Objective Rights." *Journal of Moral Philosophy*, 2017, 1–30.

Lazar, Seth, and Chad Lee-Stronach. "Axiological Absolutism and Risk." *Noûs*, 2017, 1–17.

Lee-Stronach, Chad. "Moral Priorities Under Risk." *Canadian Journal of Philosophy*, 2017, 1–19.

Leitgeb, Hannes. "The Stability Theory of Belief." *Philosophical Review* 123, no. 2 (2014): 131–171.

Levi, Isaac. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, Massachusetts; London, England: MIT Press, 1980.

———. "Why Indeterminate Probability is Rational." *Journal of Applied Logic* 7, no. 4 (December 2009): 364–376.

Lin, Hanti. "On the Regress Problem of Deciding How to Decide." *Synthese* 191, no. 4 (2014): 661–670.

Lin, Hanti, and Kevin T. Kelly. "A geo-logical solution to the lottery paradox, with applications to conditional logic." *Synthese* 186 (2012): 531–575.

———. "Propositional Reasoning that Tracks Probabilistic Reasoning." *Journal of Philosophical Logic* 41, no. 6 (2012): 957–981.

Lord, Errol, and Barry Maguire, eds. "The Implementation Problem for Deontology." Chap. 14 in *Weighing Reasons*, 338–354. Oxford: Oxford University Press, 2016.

Luce, R. Duncan, and Howard Raiffa. *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley / Sons, Ltd., 1957.

MacAskill, William. *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. Guardian Faber Publishing, 2015.

———. "Normative uncertainty as a voting problem." *Mind* 125, no. 500 (2016): 967–1004.

Marmor, Andrei. "An Institutional Conception of Authority." *Philosophy & Public Affairs* 39, no. 3 (2011): 238–261.

McCarthy, David. "Rights, Explanation, and Risks." *Ethics* 107, no. 2 (1997): 205–225.

McKerlie, Dennis. "Rights and Risk." *Canadian Journal of Philosophy* 16, no. 2 (1986): 239–251.

Nagel, Thomas. "Egalitarianism." Chap. 7 in *Equality and Partiality*. Oxford: Oxford University Press, 1991.

———. *Mortal Questions*. Cambridge University Press, 1979.

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

Nover, Harris, and Alan Hájek. "Vexing Expectations." *Mind* 113, no. 450 (2004): 237–249.

Nozick, Robert. *Anarchy, State, and Utopia.* New York: Basic, 1974.

———. "Moral Complications and Moral Structures." *Natural Law Forum* 13 (1968): 1–50.

Oddie, Graham, and Peter Milne. "Act and Value: Expectation and the Representability of Moral Theories." *Theoria*, 1991.

O'Neill, Onora. "Instituting Principles: Between Duty and Action." *Southern Journal of Philosophy* 36, no. S1 (1998): 79–96.

Osiel, Mark J. "Obeying Orders: Atrocity, Military Discipline, and the Law of War." *California Law Review* 86, no. 5 (1998): 943.

Parfit, Derek. "Justifiability to Each Person." *Ratio* XVI, no. 4 (2003): 368–390.

Perry, Stephen. "Risk, Harm, Interests, and Rights." In *Risk: Philosophical Perspectives*, edited by Tim Lewens. Routledge: Oxford University Press, 2007.

———. "Second-Order Reasons, Uncertainty and Legal Theory." *California Law Review* 62 (1988): 913–994.

Pettigrew, Richard. "Accuracy, Chance, and the Principal Principle." *Philosophical Review* 121, no. 2 (2012): 241–275.

———. "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research* 92, no. 1 (2016): 35–59.

Pettigrew, Richard, and Michael G Titelbaum. "Deference Done Right." *Philosopher's Imprint* 14, no. 35 (2014): 1–19.

Portmore, Douglas W. *Commonsense Consequentialism: Wherein Morality Meets Rationality.* xx, 266 p. Oxford: Oxford University Press, 2011.

Pratt, John W. "Risk Aversion in the Small and in the Large." Chap. Chapter 19 in *Handbook of the Fundamentals of Financial Decision Making*, 317–331. 1964.

Privitera, Johanna. "Aggregate Relevant Claims in Rescue Cases?" *Utilitas* 30, no. 2 (2018): 228–236.

Pummer, Theron. "Whether and Where to Give." *Philosophy and Public Affairs* 44, no. 1 (2016): 77–95.

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

———. *A Theory of Justice: Revised Edition.* Cambridge, MA: Belknap Press, 1999.

———. *Justice as Fairness: A Restatement.* Edited by Erin Kelly. Cambridge, Massachusetts and London, England: Harvard University Press, 2001.

———. "Some Reasons for the Maximin Criterion." *The American Economic Review* 64, no. 2 (1974): 141–146.

Raz, Joseph. "Authority and Justification." *Philosophy and Public Affairs* 14, no. 1 (1985): 3–29.

———. *Ethics in the Public Domain.* Oxford: Clarendon Press, 1994.

———. "Facing Up: A Reply." *California Law Review* 62 (1988).

———. *From Normativity to Responsibility.* Oxford; New York: Oxford University Press, 2011.

———. "Legitimate Authority." In *Philosophical Law*, edited by Richard Bronaugh, 3–27. Westport, Conn.: Greenwood Press, 1978.

———. *Practical Reason and Norms.* Oxford: Oxford University Press, 1999.

Raz, Joseph. *The Morality of Freedom.* Oxford: Clarendon Press, 1986.

————. "The Problem of Authority: Revisiting the Service Conception." *Minnesota Law Review* 90, no. 1979 (2006): 1003–1044.

Regan, Donald H. "Reasons, Authority, and the Meaning of 'Obey': Further Thoughts on Raz and Obedience to Law." *Canadian Journal of Law and Jurisprudence* 3 (1990): 3–28.

Resnik, Michael D. *Choices: An Introduction to Decision Theory.* London: University of Minnesota Press, 1987.

Ridge, Michael. "How to Avoid Being Driven to Consequentialism: A Comment on Norcross." *Philosophy & Public Affairs* 27, no. 1 (1998): 50–58.

Ross, Jacob, and Mark Schroeder. "Belief, Credence, and Pragmatic Encroachment." *Philosophy and Phenomenological . . .* 88, no. 2 (2012): 259–288.

Samuelson, Paul A. "Risk and Uncertainty: A Fallacy of Large Numbers." *Scientia* 98 (1963): 108–113.

Savage, Leonard J. *The Foundations of Statistics.* 2nd rev. New York: Dover, 1972.

Scanlon, T.M. *Moral Dimensions.* Cambridge Massachusetts, and London, England: Belknap Press of Harvard University Press, 2008.

————. *What We Owe to Each Other.* Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press, 1998.

Sciaraffa, Stefan. "On Content-Independent Reasons: It's Not in the Name." *Law and Philosophy* 28, no. 3 (2008): 233–260.

Sen, Amartya. "Internal Consistency of Choice." *Econometrica* 61, no. 3 (1993): 495–521.

Sepielli, Andrew. "What to Do When You Don't Know What to Do." In *Oxford Studies in Metaethics.* 2009.

Shapiro, Scott J. *Authority*, edited by Jules L. Coleman, Kenneth Einar Himma, and Scott J. Shapiro, 382–439. Oxford University Press, 2004.

Shiffrin, Seana Valentine. "Wrongful Life, Procreative Responsibility, and the Significance of Harm." *Legal Theory* 5, no. 2 (1999): 117–148.

Singer, Peter. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. Text Publishing Company, 2015.

Smith, Holly M. "A Paradox of Promising." *Philosophical Review* 106, no. 2 (1997): 153–196.

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

———. "The Subjective Moral Duty to Inform Oneself before Acting." *Ethics* 125, no. 1 (2014): 11–38.

Steele, Katie. "What are the Minimal Requirements of Rational Choice? Arguments from the Sequential-Decision Setting." *Theory and Decision* 68, no. 4 (2010): 463–487.

Swenson, Philip. "Subjective Deontology and the Duty to Gather Information." *Ethics* 125, no. October (2016): 257–271.

Taurek, John M. "Should the Numbers Count?" *Philosophy and Public Affairs* 6, no. 4 (1977): 293–316.

Temkin, Larry S. "A Continuum Argument for Intransitivity." *Philosophy and Public Affairs* 25, no. 3 (1996): 175–210.

———. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford; New York: Oxford University Press, 2012.

Tenenbaum, Sergio. "Action, Deontology, and Risk: Against the Multiplicative Model." *Ethics* 127 (2017): 1–36.

Tenenbaum, Sergio, and Diana Raffman. "Vague Projects and the Puzzle of the Self-Torturer." *Ethics* 123, no. 1 (2012): 86–112.

Tessman, Lisa. *Moral Failure: On the Impossible Demands of Morality.* Oxford: Oxford University Press, 2015.

Thoma, Johanna. "Risk Aversion and the Long Run." *Ethics*, 2019, 1–31.

Thomson, Judith Jarvis. *Rights, Restitution, and Risk: Essays in Moral Theory.* Cambridge, Mass.: Harvard University Press, 1986.

———. *The Realm of Rights.* Harvard University Press, 1990.

Tversky, Amos, and Daniel Kahneman. "Judgement Under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (1974): 1124–1131.

Vallinder, Aron. "Imprecise Bayesianism and Global Belief Inertia." *British Journal for the Philosophy of Science* 0, no. June (2018): 1–26.

Von Neumann, John, and Oskar Morgenstern. *Theory of Games and Economic Behavior.* 3rd. 1947. Princeton: Princeton University Press, 1990.

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

Wilford, John Noble. *The Mapmakers.* Revised Edition. New York: Vintage Books, 2000.

Williamson, Timothy. *Knowledge and Its Limits.* Oxford Scholarship Online, 2003.

Wolff, Robert Paul. *In Defense of Anarchism.* University of California Press, 1970.